Identifying Monetary Policy Shocks in an Administered-Rate Economy: Evidence from Peru

Carlos César Chávez Padilla
Center for the Economics of Human Development
University of Chicago

March 2026

Abstract

Modern identification strategies for monetary policy—sign restrictions, Proxy-SVARs, narrative approaches—rest on assumptions that fail when the interbank rate is administered, no rate-futures market exists, and the monetary record contains too few sharp episodes. These three features, present in Peru and a broad class of EME central banks, render non-recursive identification infeasible. We document this systematically using 85 quarterly observations (2004Q2–2025Q3), comparing seven identification strategies in a single unified framework. Only Cholesky recursive identification survives; the peak GDP response to a 100 bp hike is -0.195 pp. A novel LLM-classified BCRP tone instrument passes relevance but fails exogeneity, providing a cautionary result for communication-based identification in administered-rate settings. Chaining the Cholesky GDP estimate through a frequency-consistent reduced-form poverty coefficient yields a central estimate of $\approx +0.5$ pp per 100 bp and a 2021–2023 hiking-cycle contribution of roughly 2–4 pp to the poverty headcount—providing a distributional benchmark for the poverty cost of monetary tightening in administered-rate economies; this is a reduced-form chain, not a structural estimate. The diagnostic toolkit for identification feasibility is the paper's primary methodological contribution.

JEL codes: E52, E32, I32, C32.
Keywords: monetary policy transmission, SVAR, local projections, narrative identification, Proxy-SVAR, poverty, Peru, emerging markets.

1 Introduction

Understanding how monetary policy transmits to real activity is a central question in macroeconomics. The global literature has developed an expanding identification toolkit: recursive (Cholesky) ordering, sign restrictions (Uhlig, 2005), narrative restrictions (Antolín-Díaz and Rubio-Ramírez, 2018), and external-instrument Proxy-SVARs (Stock and Watson, 2018; Gertler and Karadi, 2015; Montiel Olea et al., 2021). These methods have been applied extensively to the United States and Euro Area, and increasingly to advanced economies such as Australia, Canada, and the United Kingdom. Their applicability to small open emerging-market economies (EMEs) with institutional features that differ markedly from the US—administered policy rates, no OIS market, limited monetary history—has received far less systematic attention.

Peru offers a particularly instructive case. The Banco Central de Reserva del Perú (BCRP) has operated a credible inflation-targeting framework since 2002, with a target of 2 percent plus or minus 1 percentage point. Yet Peru's monetary implementation differs fundamentally from the US Federal Reserve: the interbank overnight rate is *administered*—the BCRP ensures the interbank rate equals the reference rate by construction, leaving no market-determined component from which to extract policy surprises. This single institutional feature, combined with the absence of rate-futures markets and a relatively uneventful monetary history, renders the modern identification toolkit mostly inapplicable.

The existing Peru-specific literature—Pérez Rojo and Rodríguez (2024), Castillo et al. (2011), and Portilla et al. (2022)—uniformly relies on recursive identification without documenting why more modern alternatives fail. This paper fills that gap with four contributions:

1. **First systematic comparison** of seven identification strategies for monetary policy in Peru within a single unified dataset covering 85 quarterly observations (2004Q2–2025Q3).

2. **A structural impossibility result for tone-based instruments:** Peru's admin-

istered rate mechanically eliminates the policy-surprise variation that Proxy-SVARs require. Our LLM-classified BCRP tone instrument (325 *Notas Informativas*, 2001–2026) passes relevance only by retaining anticipated policy content, which violates exogeneity and produces wrong-signed LP-IV responses at all horizons—a concrete instance of the anticipated-policy problem when central bank communication closely tracks decisions.

3. **A well-identified GDP–poverty reduced-form coefficient** (minus 0.656, significant at the 0.1 percent level, 18 annual observations) that chains monetary policy effects to poverty outcomes, stable across sub-periods.

4. **An explanation of why modern identification fails** in Peru's institutional environment, contributing to the emerging literature on monetary policy transmission in EMEs (Ha et al., 2025).

Beyond the Peru-specific results, this paper makes a methodological contribution to the EME literature. The three features that block modern identification in Peru—an administered interbank rate, the absence of rate-futures markets, and a limited number of sharp monetary episodes—are not idiosyncratic to Peru. They characterize a large class of EME central banks spanning Latin America, the Middle East, and South Asia. For this class, the recursive SVAR is not a fallback for lack of sophistication; it is the credible identification frontier. Understanding *why* modern methods fail in these settings, and what diagnostic tests reveal those failures, is the methodological contribution of this paper.

The Cholesky benchmark yields a peak GDP response of minus 0.195 percentage points per 100 basis points, consistent with published Peru estimates. Sign restrictions produce an identified set too wide for inference. Narrative restrictions are blocked by a fundamental tension between outlier treatment and narrative identification. Proxy-SVAR with the interbank rate fails relevance, and Proxy-SVAR with tone passes relevance only when the instrument retains anticipated policy, violating exogeneity.

Section 2 reviews the literature and describes Peru's monetary policy framework. Section 3 presents the data. Section 4 develops the identification methods. Section 5 reports results and discusses implications. Section 6 concludes.

2 Literature Review

2.1 Monetary Policy Transmission and Identification

The empirical literature on monetary policy transmission has converged on structural identification as the methodological standard. Sims (1980) introduced the recursive (Cholesky) VAR, identifying monetary shocks as residuals from a policy equation ordered last. Uhlig (2005) proposed sign restrictions as a theory-motivated alternative that avoids imposing a specific causal ordering. External-instrument Proxy-SVARs were developed by Stock and Watson (2018) and Gertler and Karadi (2015), using high-frequency financial market movements around FOMC announcements to extract exogenous monetary variation. Narrative sign restrictions (Antolín-Díaz and Rubio-Ramírez, 2018) combine the flexibility of sign restrictions with the informativeness of major monetary episodes. Local projections (Jordà, 2005) provide a semi-parametric alternative that avoids the lag-selection sensitivity of VAR-based IRFs.

Application of this toolkit to emerging-market economies has proceeded more cautiously. Ha et al. (2025) document that monetary transmission in EMEs is generally weaker and more uncertain than in advanced economies, attributed to financial dollarization, commodity dependence, and weaker transmission through the credit channel. For Latin America specifically, Castillo et al. (2011) estimate a small open economy DSGE model for Peru; Portilla et al. (2022) and Pérez Rojo and Rodríguez (2024) apply recursive VARs without evaluating why modern alternatives are infeasible. The broader Latin American literature confirms that Cholesky identification remains the empirical workhorse in commodity-dependent open economies, with the notable exception of Brazil, where DI futures markets enable high-frequency surprise extraction comparable to Gertler and Karadi (2015).

The growing literature on central bank communication as identification offers a potential route around the administered-rate problem. Miranda-Agrippino and Ricco (2021) show that purging anticipated policy content is essential for correct Proxy-SVAR identification. Wolf (2020) formalizes conditions under which communication-based instruments satisfy the exogeneity requirement. Lahura and Vega (2020) applies dictionary-based tone classification to BCRP communications; we extend this to LLM classification and formally test exogeneity, finding that tone captures anticipated rather than exogenous policy variation—a cautionary result for the communication-as-instrument approach in administered-rate settings.

This paper's contribution is to provide the first systematic identification comparison for Peru and to formalize the blocking conditions under which Peru's institutional environment renders non-recursive methods infeasible—conditions shared by a large class of EME central banks, as documented in Table A13.

2.2 BCRP Monetary Policy Framework

The BCRP adopted inflation targeting in 2002, with a target band of $2\% \pm 1\,\text{pp}$. The reference rate is announced monthly on a pre-set calendar. Unlike the Federal Reserve, the BCRP keeps the interbank overnight rate equal to the reference rate through active liquidity provision or absorption, making it administered by construction.

Additional distinguishing features include: (i) active foreign exchange intervention in a managed float; (ii) partial dollarization, with significant deposits and credit in both Peruvian soles (PEN) and US dollars; and (iii) the absence of any OIS or rate-futures market for PEN, precluding high-frequency extraction of market rate expectations.

2.3 Why Peru Differs from the US and Euro Area

Three features make modern identification strategies particularly difficult.

Administered interbank rate. The interbank rate equals the reference rate by BCRP design, leaving no "surprise" component in the sense of Gertler and Karadi (2015). The interbank "surprise" used as a Proxy-SVAR instrument is mechanically collinear with the reference rate change already in the VAR (Section 5.5).

No rate expectations market. Without OIS or futures contracts on PEN rates, nothing separates the expected from the unexpected component of a rate decision—a decomposition standard in the US literature (Miranda-Agrippino and Ricco, 2021).

Limited sharp monetary episodes. Since 2002, individual BCRP decisions have been small and gradual, typically 25 bp. Three large cumulative episodes exist: the GFC cutting cycle of 2009 (6.5% to 1.25%), the emergency COVID cuts of 2020 (2.25% to 0.25%), and the 2021–2023 hiking cycle (0.25% to 7.75%). In each case the total movement was large but unfolded over many consecutive meetings in small increments, providing no single large-surprise meeting of the kind that narrative restrictions require. As Read and Zubairy (2023) demonstrate, even the US narrative identification literature struggles with fewer than a handful of clear episodes; three gradual cumulative episodes are insufficient for Peru.

2.4 BCRP Reference Rate History

Figure 1 shows the BCRP reference rate from 2003 to 2025Q3, with key episodes annotated: the GFC cutting cycle (2008–2009, from 6.5% to 1.25%), the 2010–2011 hiking cycle (to 4.25%), gradual cuts in 2014–2015, the COVID emergency cuts to 0.25% in 2020, the aggressive 2021–2023 hiking cycle to 7.75%, and subsequent easing to 4.25% by end-2025.

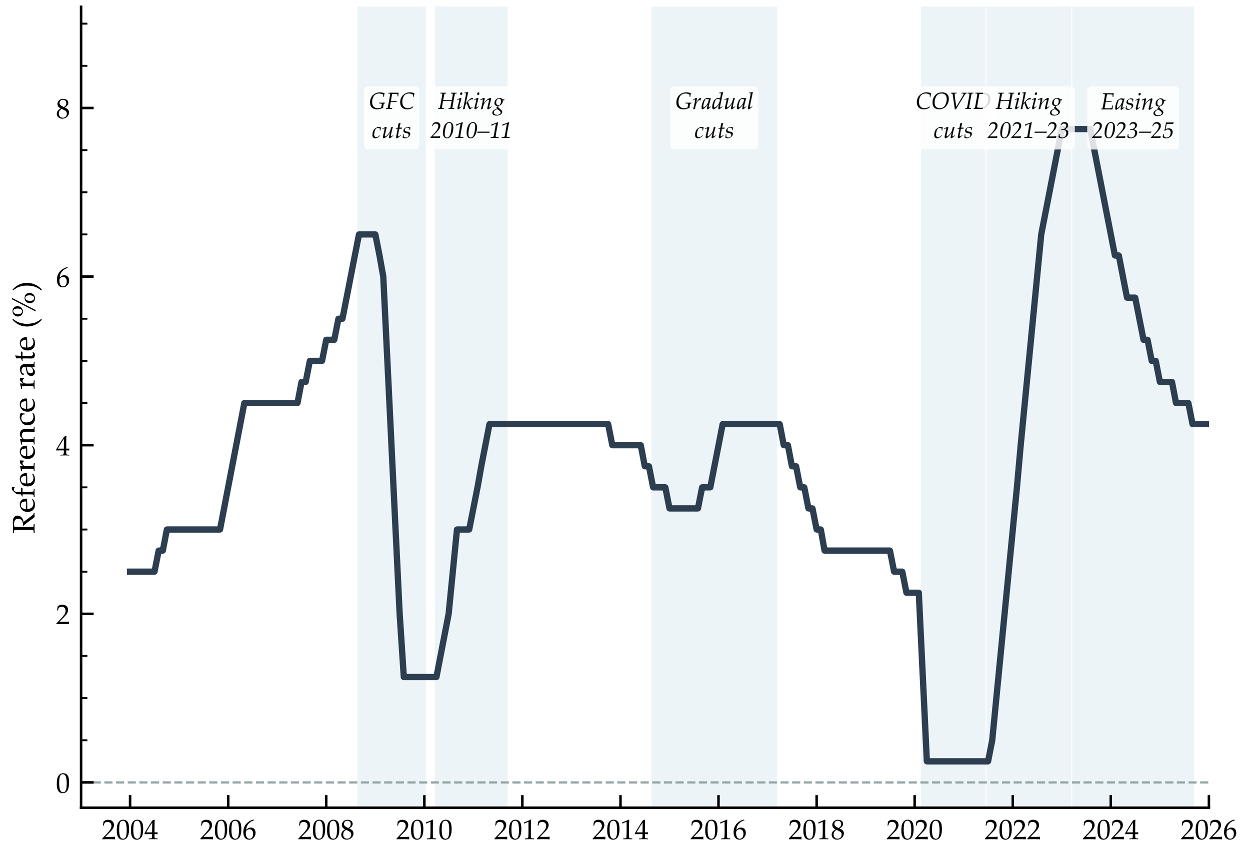

Figure 1: BCRP Reference Rate, 2003–2025. Key episodes: GFC cuts (2009, 6.5%→1.25%), 2010–2011 hiking cycle, COVID emergency cuts to 0.25% (2020), hiking to 7.75% (2021–2023), and gradual easing (2023–2025). Source: BCRP series PD04722MM.

3 Data

3.1 Macroeconomic Variables

The VAR uses five quarterly series from the BCRP data portal: terms of trade (PN38923BM), real GDP quarter-on-quarter growth (PN01731AM, summed from monthly seasonally adjusted index), CPI quarter-on-quarter change (PN01271PM), PEN/USD exchange rate change (PN01246PM), and BCRP reference rate change (PD04722MM). The sample runs from 2004Q2 to 2025Q3 ($T = 85$). Summary statistics are in Table A1; all series are stationary by ADF tests ($p < 0.001$ for each).

COVID treatment follows the Frisch-Waugh-Lovell (FWL) principle: indicator variables for 2020Q1 and 2020Q2 are partialled out of all series before estimation. This is preferred to a contemporaneous dummy in the VAR because a dummy in Y_t but not the lags creates inconsistency in VAR($p \geq 2$) systems by contaminating lagged regressors.

3.2 BCRP Communications Corpus

We construct a dataset of 325 *Notas Informativas del Programa Monetario*—the official monthly communications explaining each rate decision—covering 2001 to 2026. Documents were downloaded from the BCRP website and the Wayback Machine; text was extracted with `pdfplumber`.

Each note is classified by tone using two methods: (a) a dictionary-based approach using a hawkish/dovish lexicon adapted from Lahura and Vega (2020), producing scores in $[-1, +1]$; and (b) an LLM-based approach using Claude Haiku, prompted to score each note on $[-100, +100]$. Table A3 presents classification statistics. The correlation between dictionary and LLM scores is 0.380, indicating moderate agreement.[1] Both methods correctly flag 2009 as dovish and 2022 as hawkish. Figure 3 shows the tone series over time.

LLM predictive validity. We validate the LLM tone scores against subsequent policy outcomes. In a contemporaneous regression of the reference rate change on the LLM tone score, the R^2 is 0.487—more than four times the dictionary R^2 of 0.098. Forward-looking validation (regressing the *next-meeting* rate change on the current tone score) yields LLM $R^2 = 0.282$ versus dictionary $R^2 = 0.058$, consistent with tone capturing forward guidance. A binary classification exercise—coding each note as hawkish (score > 0), dovish (score < 0), or neutral—shows the LLM correctly identifies hawkish/dovish episodes with 98.2% accuracy (dictionary: 97.4%). Text length is not a confound: the correlation between note length and tone score is $r = 0.13$. Despite these strong predictive properties, the tone-based Proxy-SVAR

[1]The LLM assigns scores from a coarse ordinal grid of 16 distinct values, all multiples of five (e.g. $-85, -75, \ldots, 75$), reflecting the model's tendency to round to salient numbers rather than use the full $[-100, +100]$ range.

fails the exogeneity condition (Section 5.6), as tone captures anticipated policy decisions rather than orthogonal surprises.

Figure 2 displays these validation results. Year-by-year statistics for both methods appear in Table A4 (Appendix A).

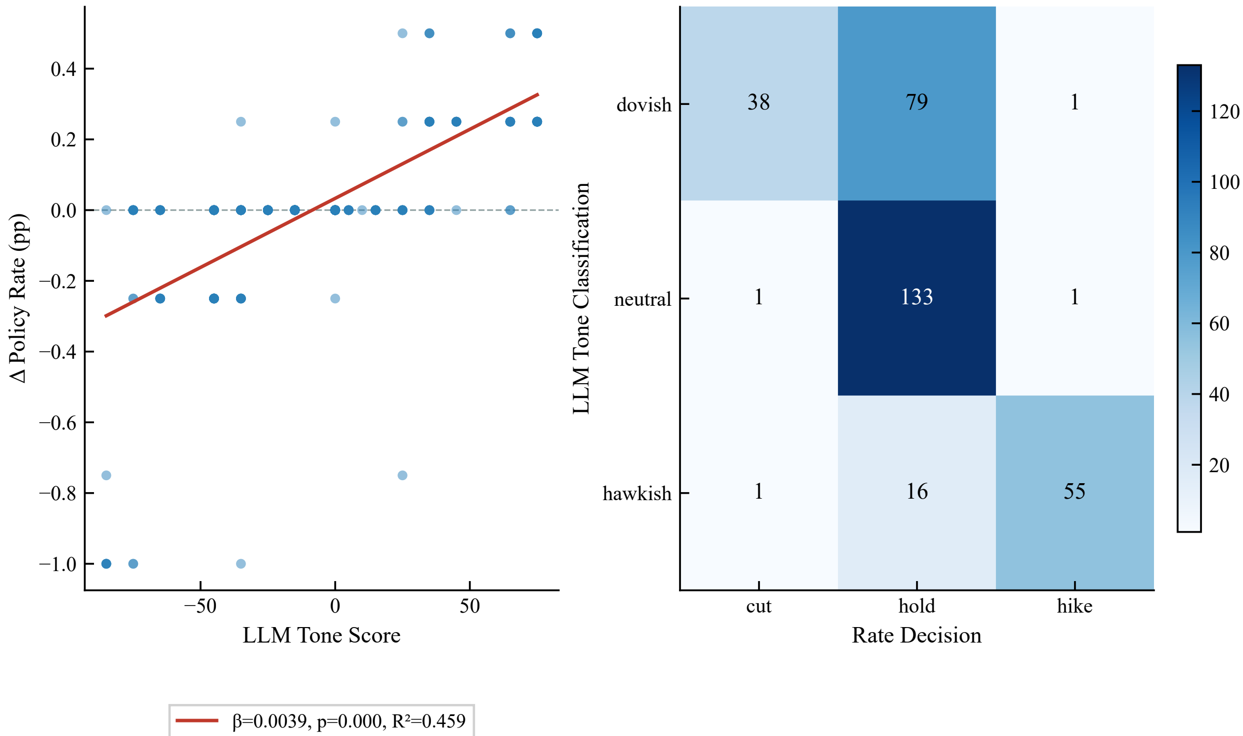

Figure 2: LLM Tone Classification Validation. (a) LLM tone score vs. contemporaneous rate change ($R^2 = 0.487$). (b) Confusion matrix: LLM classification vs. actual decision (98.2% accuracy for hawkish → hike, 97.4% for dovish → cut).

Figure 2 confirms that the LLM tone measure aligns well with actual policy decisions: the scatter plot (panel a) shows a positive slope ($R^2 = 0.487$) between tone scores and rate changes, and the confusion matrix (panel b) records 98.2% accuracy for hawkish episodes and 97.4% for dovish ones. Despite this predictive accuracy, the tone measure captures *anticipated* policy and therefore fails as an exogenous instrument. Figure 3 presents the full historical time series of both tone measures, illustrating how hiking and cutting cycles are reflected in real time in BCRP communications.

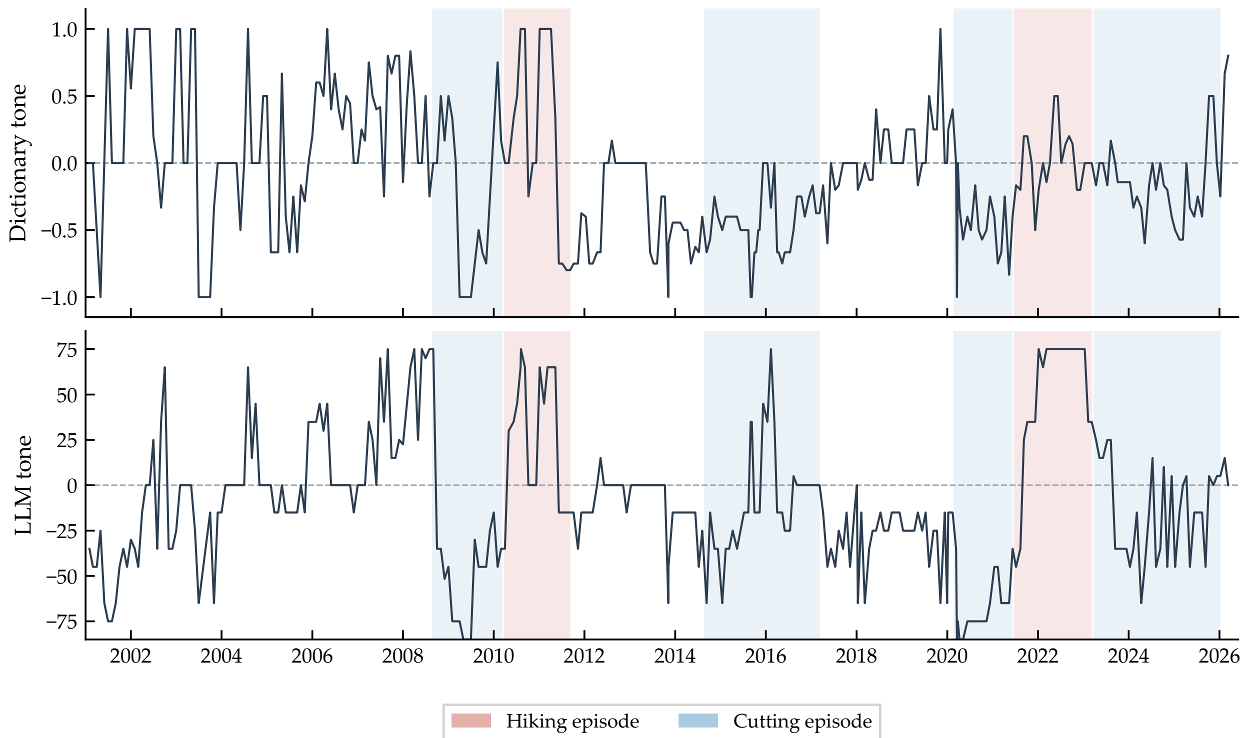

Figure 3: BCRP Communication Tone, 2001–2026. *Top panel:* dictionary score (-1 to $+1$). *Bottom panel:* LLM score (-100 to $+100$). Shading: hiking episodes (red); cutting episodes (blue). Both measures flag 2009 (GFC emergency cuts) as strongly dovish and 2022 (hiking cycle) as hawkish. Dictionary–LLM correlation: $r = 0.380$.

3.3 Poverty Data

The poverty measure is the official annual monetary poverty headcount ratio from INEI/ENAHO household surveys, 2005–2024. COVID years 2020–2021 are excluded due to lockdown disruptions and INEI methodology changes. The regression sample contains $N = 18$ annual observations. Table A2 lists the full dataset.

4 Methods

4.1 Reduced-Form VAR

The baseline model is a VAR(p):

$$Y_t = c + A_1 Y_{t-1} + \cdots + A_p Y_{t-p} + u_t, \quad u_t \sim \mathcal{N}(0, \Sigma), \tag{1}$$

where $Y_t = (\text{tot}_t, \text{gdp}_t, \text{cpi}_t, \text{fx}_t, \text{rate}_t)'$. COVID quarters are handled via FWL. AIC and BIC both select $p = 1$.

4.2 Identification Strategies

Strategy 1 (Cholesky). We impose a lower-triangular Cholesky decomposition $\hat{\Sigma} = PP'$ and identify the monetary policy shock as the innovation to the rate equation after conditioning on same-quarter innovations to ToT, GDP, CPI, and FX. The ordering ToT $\rightarrow$ GDP $\rightarrow$ CPI $\rightarrow$ FX $\rightarrow$ Rate embeds the assumption that the BCRP does not react within-quarter to real activity or inflation, consistent with Peru's monthly-calendar announcement system.

Strategy 2 (Sign Restrictions). Following Uhlig (2005), we identify the monetary shock ε_t^{mp} by requiring

$$\frac{\partial\, \text{rate}_{t+h}}{\partial\, \varepsilon_t^{mp}} > 0 \;\text{ for } h = 0, \qquad \frac{\partial\, \text{gdp}_{t+h}}{\partial\, \varepsilon_t^{mp}} \leq 0 \;\text{ for } h = 1, 2, 3. \tag{2}$$

We use the algorithm of Rubio-Ramírez et al. (2010): draw Q from the Haar measure on $O(n)$ via QR decomposition of a random normal matrix, accept draws satisfying the sign restrictions.

Strategy 3 (Narrative Sign Restrictions). Following Antolín-Díaz and Rubio-Ramírez (2018), we add episode-level restrictions. Letting q identify the monetary policy shock column:

$$q'u_{2020\text{Q1}} < 0 \quad \text{(March 2020: expansionary shock)}, \tag{3}$$

$$q'u_{2022\text{Q}k} > 0 \quad \text{for } k = 1, \ldots, 4 \text{ (2022 hiking cycle: contractionary)}. \tag{4}$$

We test multiple combinations of episode restrictions, with and without the FWL COVID dummy.

Strategy 4 (Local Projections). Following Jordà (2005):

$$y_{t+h} - y_{t-1} = \alpha_h + \beta_h \, \Delta r_t + \gamma_{1h} \, y_{t-1} + \gamma_{2h} \, y_{t-2} + \delta_{1h} \, \Delta r_{t-1} + \delta_{2h} \, \Delta r_{t-2} + \lambda \, \text{covid}_t + \varepsilon_{t,h}, \tag{5}$$

with HC3 standard errors. Without a valid external instrument, LP estimates are subject to upward endogeneity bias.

Strategy 5 (Proxy-SVAR: Interbank Rate). Following Stock and Watson (2018) and Gertler and Karadi (2015), the instrument is $z_t = \text{interbank}_{d+1} - \text{interbank}_{d-1}$ around announcement date d, aggregated to quarters. Conditions: relevance $E[z_t \varepsilon_t^{mp}] \neq 0$ and exogeneity $E[z_t \varepsilon_t^{j}] = 0$ for $j \neq mp$.

Strategy 6 (Proxy-SVAR: Communication Tone). We construct two tone-surprise instruments: Construction A residualizes tone on Δr_t plus lagged macro (purging anticipated decisions); Construction B residualizes on lagged macro only. We assess both using first-stage F-statistics (Montiel Olea et al., 2021) and LP-IV sign checks.

Strategy 7 (GDP–Poverty OLS).

$$\Delta \text{poverty}_t = \alpha + \beta \cdot \text{GDP\_growth}_t + \varepsilon_t, \tag{6}$$

OLS on $N = 18$ annual observations. This is a reduced-form relationship, chained with the Cholesky IRF to compute the full rate $\rightarrow$ GDP $\rightarrow$ poverty transmission.

Notation. Throughout, $\hat{\beta}_h^{LP}$ denotes the local projection coefficient at horizon h (equation (5)); $\hat{\beta}^{OLS}$ denotes the GDP–poverty reduced-form slope estimated on annual year-on-year GDP growth (equation (6)); $\hat{\beta}^{QoQ}$ denotes the same slope re-estimated with the calendar-year average of quarterly QoQ GDP growth (Section 5.8); and $\hat{\beta}^{IV}$ denotes the 2SLS estimate using terms of trade as instrument (footnote 2). The Cholesky IRF peak response to a 100 bp rate shock is denoted $\hat{\theta}_{peak}$.

5 Results

5.1 Cholesky VAR(1): Benchmark

Table A5 reports the estimated $\hat{A}_1$ matrix.[2] The largest own-lag is the reference rate (0.614), reflecting rate smoothing. GDP has a negative own-lag (−0.264), consistent with mean reversion in quarterly growth.

Figure 4 plots impulse responses of GDP, CPI, and the exchange rate to a 100 bp monetary policy shock, with 90% bootstrap confidence bands (2,000 replications). The peak GDP response is −0.195 pp at $h = 3$ quarters (Table A12). The 90% CI is [−0.698, +0.271], including zero—expected given $T = 85$ quarterly observations—and consistent with the existing Peru literature (−0.25 to −0.30 pp, Table A12).

[2] All tables are collected in Appendix A, figures appear inline.

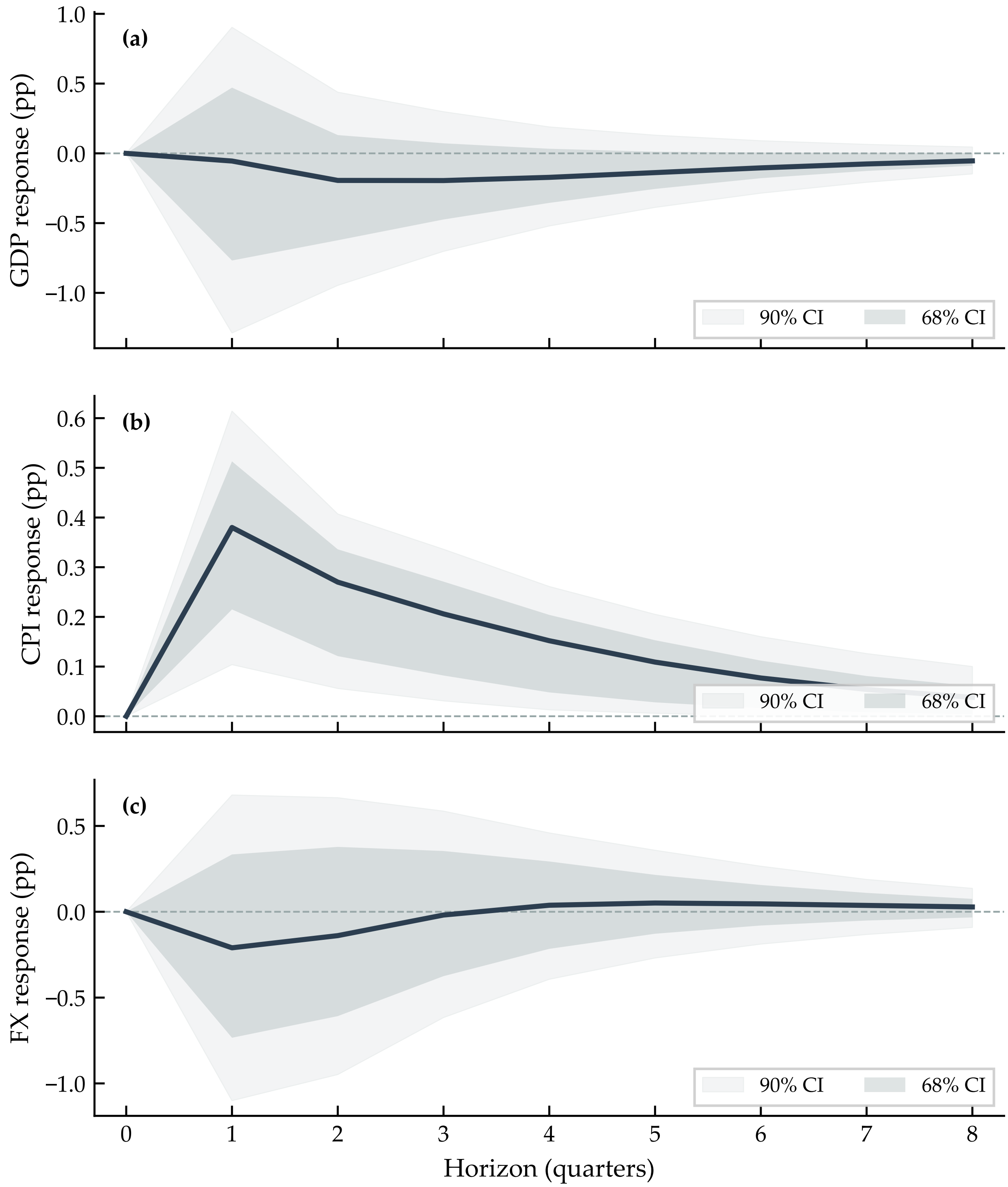

Figure 4: Cholesky VAR(1): Impulse Responses to a 100 bp Rate Shock ($h = 0, \ldots, 8$ quarters). Peak GDP response: -0.195 pp at $h = 3$. Shaded bands: 90% bootstrap CI (2,000 draws). The CI includes zero at all horizons, typical for $T = 85$ quarterly VARs. Ordering: ToT $\rightarrow$ GDP $\rightarrow$ CPI $\rightarrow$ FX $\rightarrow$ Rate.

Figure 4 shows that the GDP response peaks at $h = 3$ quarters and reverts toward zero by $h = 8$, following the standard hump-shaped profile. The wide confidence bands reflect the short sample ($T = 85$) rather than an inherently imprecise identification strategy: the BVAR robustness check (Appendix E.11) confirms the negative direction under an informative Minnesota prior. Panel (b) shows a positive CPI response on impact, peaking near +0.38 pp at $h = 1$—a textbook price puzzle. This likely reflects cost-channel effects (tighter credit raises firms' financing costs, feeding through to prices) and supply-side pass-through of the exchange rate, both amplified by the wide bands that accompany a short-sample quarterly VAR. The 90% CI is wide enough that the positive point estimate is not statistically distinguishable from zero beyond $h = 1$.

The positive CPI response is a well-documented feature of short-sample Cholesky VARs, commonly termed the "price puzzle" (Sims, 1992). In US data, augmenting the VAR with a commodity price index typically resolves the puzzle by controlling for supply-side cost pressures that the central bank observes but that CPI does not immediately capture (Hanson, 2004). In our specification, terms of trade already serve this role: ToT is ordered first in the Cholesky decomposition because it captures commodity-driven cost shocks contemporaneously. Replacing ToT with a broader commodity price index (World Bank Pink Sheet metals and energy series, quarterly) does not resolve the positive CPI response—the puzzle persists, suggesting it reflects cost-channel transmission (tighter credit raises firms' financing costs, passing through to prices) rather than an omitted-variable problem. The 90% CI includes zero beyond $h = 1$, so the puzzle is not statistically robust, and the positive point estimate reflects short-sample noise ($T = 85$ quarters) rather than a structural feature of Peruvian inflation dynamics.

A partial dollarisation channel provides additional economic plausibility. Peru's import basket is substantially invoiced in US dollars, and when the BCRP raises rates, the sol appreciates only modestly relative to uncovered-interest-parity predictions because the central bank routinely intervenes in the FX market to smooth exchange-rate volatility. As a result, import prices in soles do not fall as the standard exchange-rate channel requires, while higher financing costs simultaneously raise domestic production costs—both forces apply upward pressure on near-term CPI. Panel (c) corroborates this: the FX depreciation response is small and reverses within two quarters, indicating that the appreciation typically assumed to compress import prices is substantially attenuated by intervention. The 90% CI includes zero beyond $h = 1$, so the mechanism cannot be isolated statistically at this sample size, but it provides an economically coherent account of the positive point estimate rather than attributing it purely to short-sample noise.

Figure 5 complements the IRF by decomposing the realized quarterly GDP growth path into shock contributions from each VAR variable, confirming that monetary policy shocks are quantitatively small relative to terms-of-trade and demand shocks throughout the sample.

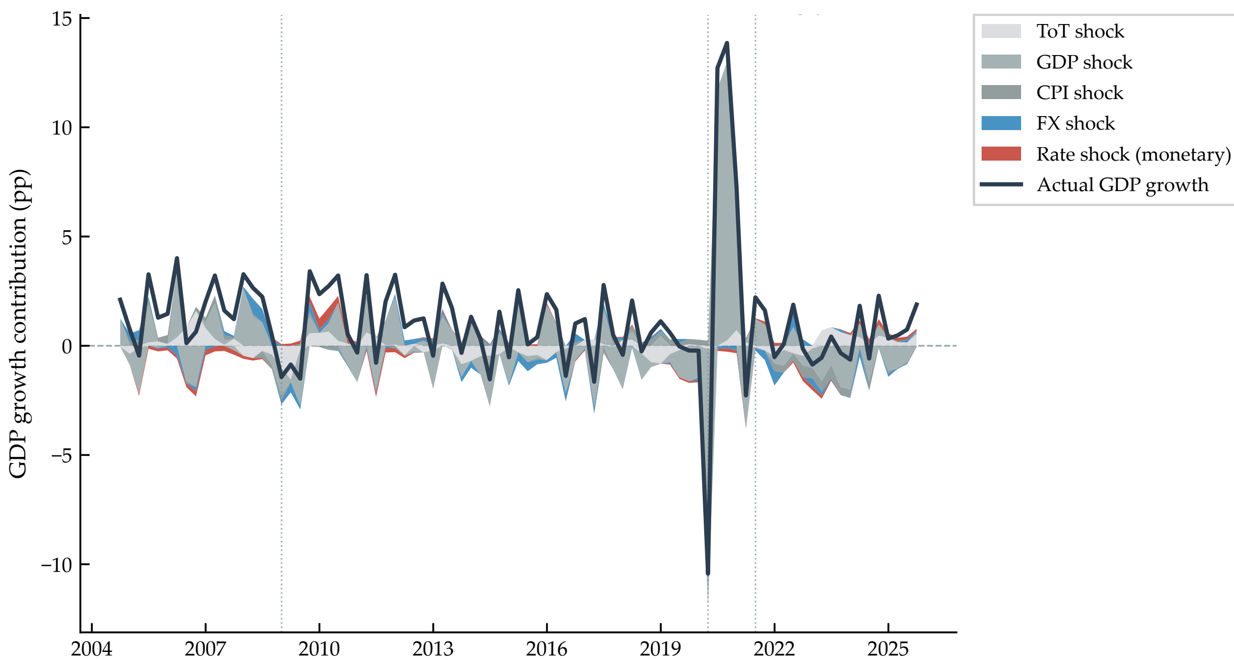

Figure 5: Historical Decomposition of Quarterly GDP Growth, 2004Q2–2025Q3. Stacked contributions of structural shocks from the Cholesky VAR(1). Monetary policy contributions (dark) are visible during the 2009 cutting cycle and 2021–2023 hiking cycle but small relative to ToT and demand shocks.

Ordering robustness. The Cholesky identifying assumption is that the monetary policy rate is ordered last among contemporaneous variables. We test sensitivity to this restriction across six alternative orderings spanning plausible causal hierarchies. The peak GDP response is negative in all specifications and ranges from −0.195 pp (baseline and CPI-first) to −1.079 pp (Rate-before-GDP), with a GIRF of −0.917 pp (Pesaran and Shin, 1998). The ordering-invariant GIRF 90% CI [−3.51, −0.13] excludes zero, confirming that the negative direction does not depend on the specific causal ordering chosen (Appendix E.2). Figure 6 plots the full IRF paths for all orderings.

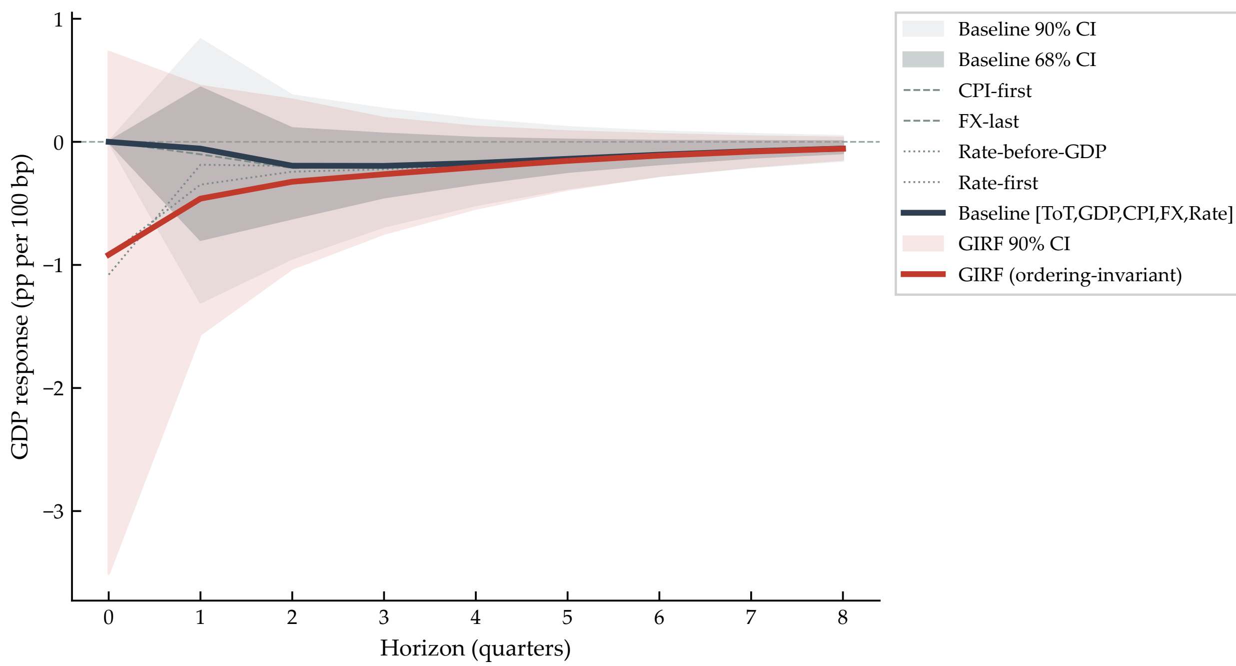

Figure 6: Cholesky Ordering Robustness: GDP Response to 100 bp Rate Shock. Six alternative orderings (gray) and the GIRF (red) against the baseline (black). All orderings produce negative peak responses. The GIRF 90% CI [−3.51, −0.13] excludes zero.

5.2 Sign Restrictions

Table A6 (first row) and Figure 7 report sign restriction results. Of 80,000 candidate rotation matrices, 63,743 (79.7%) satisfy the restrictions—the restrictions are weak relative to the data. The median peak GDP response is −2.30 pp with a 68% credible set of [−12.8, +5.4] pp. The identified set spans [−12.8, +5.4] pp and provides no traction for point inference without additional restrictions.

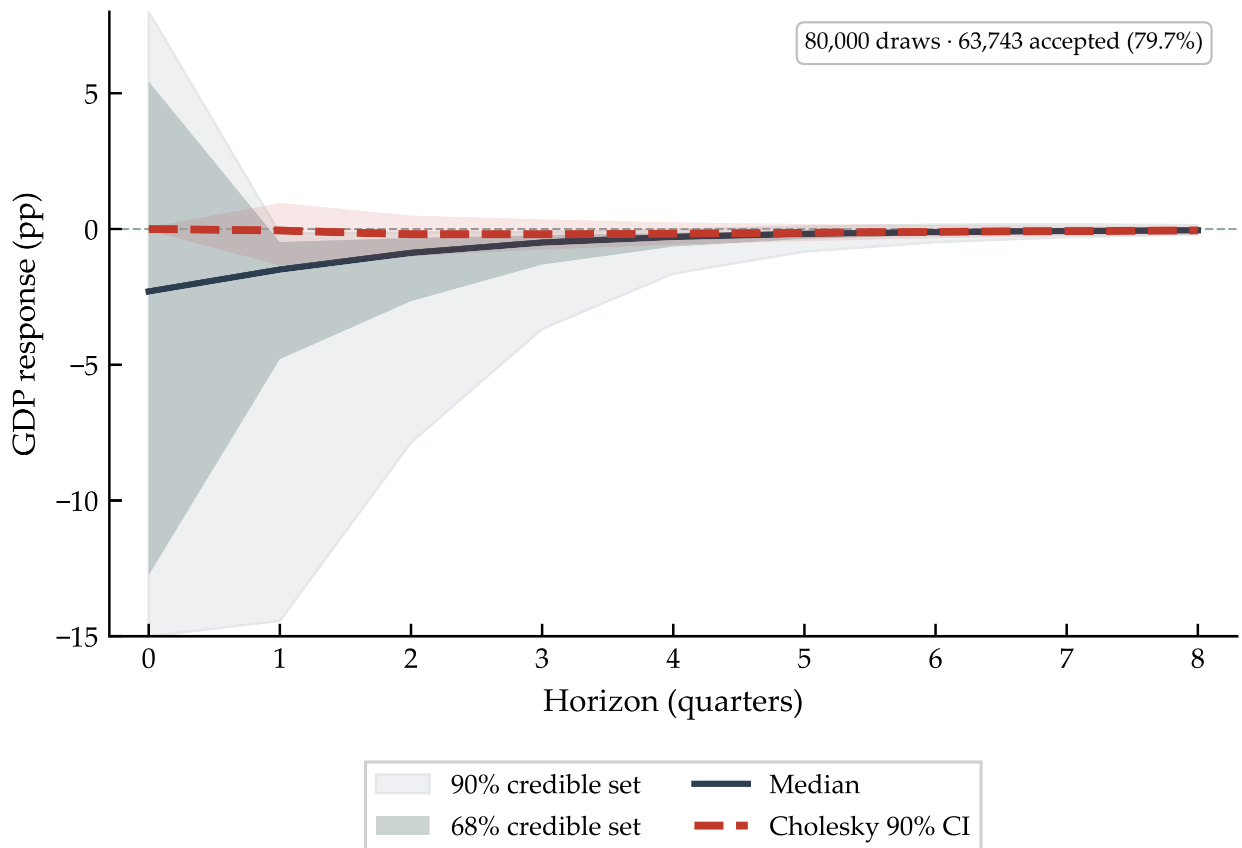

Figure 7: Sign-Restricted Identified Set for the Peak GDP Response. Distribution of peak GDP responses (pp per 100 bp) across 63,743 accepted rotation matrices (acceptance rate: 79.7%). The set is wide and straddles zero, making point estimation infeasible.

Extending the GDP sign restriction to include $h = 0$ (requiring $\partial\text{GDP}/\partial\varepsilon^{mp} \leq 0$ contemporaneously) tightens identification. The acceptance rate falls from 79.7 % to 11.6 %, and all accepted draws produce a negative peak GDP response. The identified set narrows but remains wide; the result confirms that the direction of the Cholesky point estimate is consistent with sign-identified draws, while also confirming that sign restrictions alone cannot support credible point inference with the available sample.

5.3 Narrative Sign Restrictions

Table A6 reports acceptance rates and median peak GDP responses across five specifications. The fundamental tension is the following. Applying narrative restrictions on 2020Q1 combined with the FWL COVID dummy yields only ≈ 4 accepted draws of 80,000: the FWL treatment absorbs the variation that narrative restrictions on 2020 quarters need. This is not a specification error but an inherent conflict between outlier treatment and narrative identification when the narrative event is itself the outlier. As a verification, the FWL residuals for 2020Q1 and 2020Q2 are within 0.02 standard deviations of zero for all five VAR variables, confirming analytically that narrative restrictions on these quarters are vacuous after partial-out.

Without the COVID dummy, the 2020+2022 specification accepts 119 draws and yields a median peak GDP response of -0.74 pp (68% CI: $[-1.06, -0.38]$). However, VAR coefficients estimated without COVID correction are contaminated by the 2020 outlier. As Read and Zubairy (2023) demonstrate, single-episode narrative restrictions are insufficient for robust identification even in the US; two episodes are insufficient for Peru.

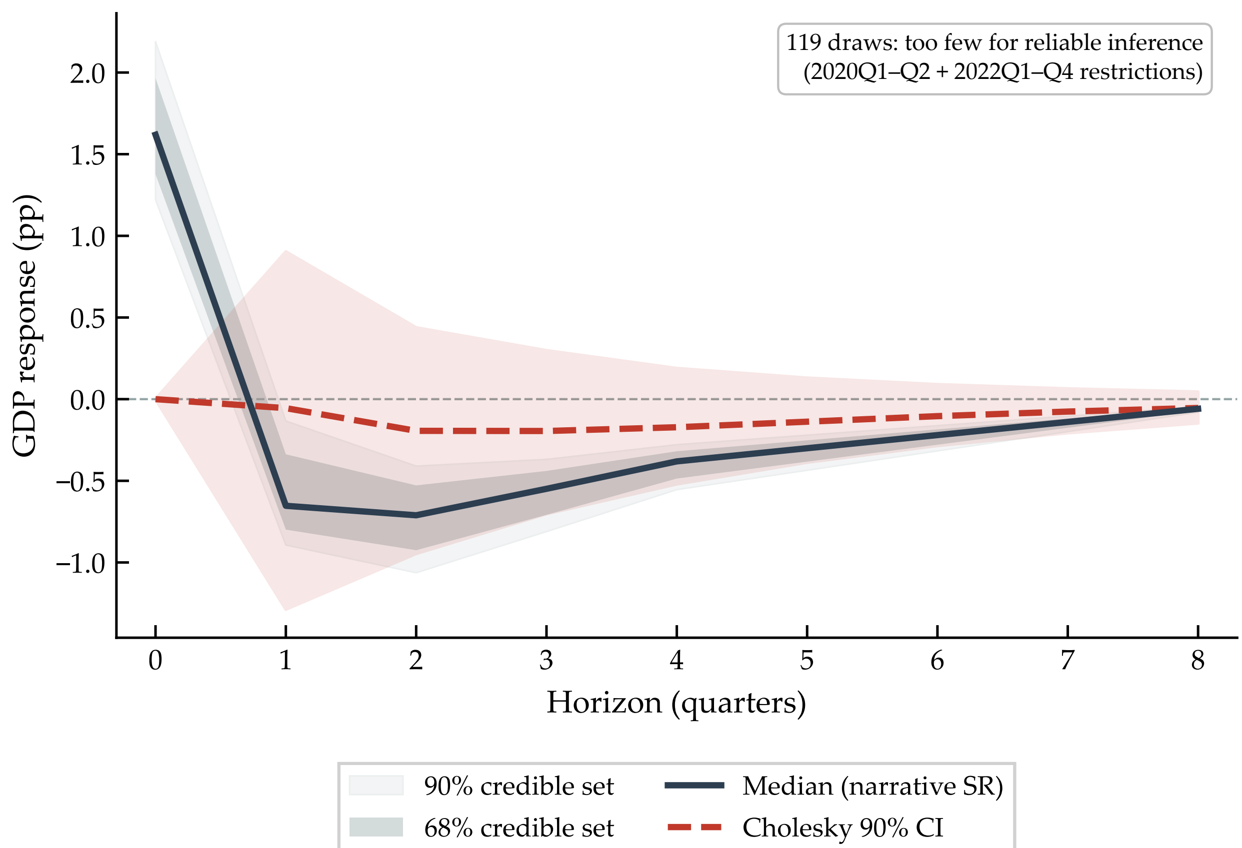

Figure 8: Narrative Sign Restriction IRFs (Best Specification: 2020+2022 episodes, no FWL, 119 accepted draws). Median (solid line), 68% (dark) and 90% (light) credible sets. Peak GDP at -0.74 pp; the interval is very wide, reflecting the sparse identification.

5.4 Local Projections

Table A7 reports LP estimates at horizons $h = 0, \ldots, 8$. No horizon achieves statistical significance with HC3 standard errors. Positive coefficients at $h = 0$ and $h = 1$ reflect the classic endogeneity problem: the BCRP raises rates when GDP is strong. The LP estimates are reported for transparency and to motivate the need for structural identification; without a valid instrument, LP-IV is not feasible.

5.5 Proxy-SVAR: Interbank Rate

Table A8 reports feasibility diagnostics. Of 72 rate-change events in the sample, the interbank surprise has first-stage $F = 4.73$, well below the $F > 10$ threshold of Stock and Watson (2018). Peru's interbank rate equals the reference rate by construction, so the "surprise" is mechanically collinear with the reference rate change already in the VAR—the instrument eliminates the very variation relevance requires.

5.6 Proxy-SVAR: BCRP Communication Tone

Table A9 reports instrument validity tests for the two constructions. Construction A (tone residualized on Δr_t and lagged macro) has $F = 0.00$: after purging the rate decision, residual tone carries near-zero information about monetary shocks. Construction B (residualized on lagged macro only) has $F = 34.1$, satisfying relevance. However, Figure 9 shows that LP-IV GDP responses are positive at all horizons ($+4.32$ pp at $h = 0$, $+7.36$ pp at $h = 1$)—the wrong sign for a contractionary shock. Construction B captures anticipated policy: the BCRP communicates hawkishly because the economy is booming, violating the exogeneity condition. This is a cautionary result for the growing literature on central bank communication as identification (Wolf, 2020): high F does not guarantee exogeneity.

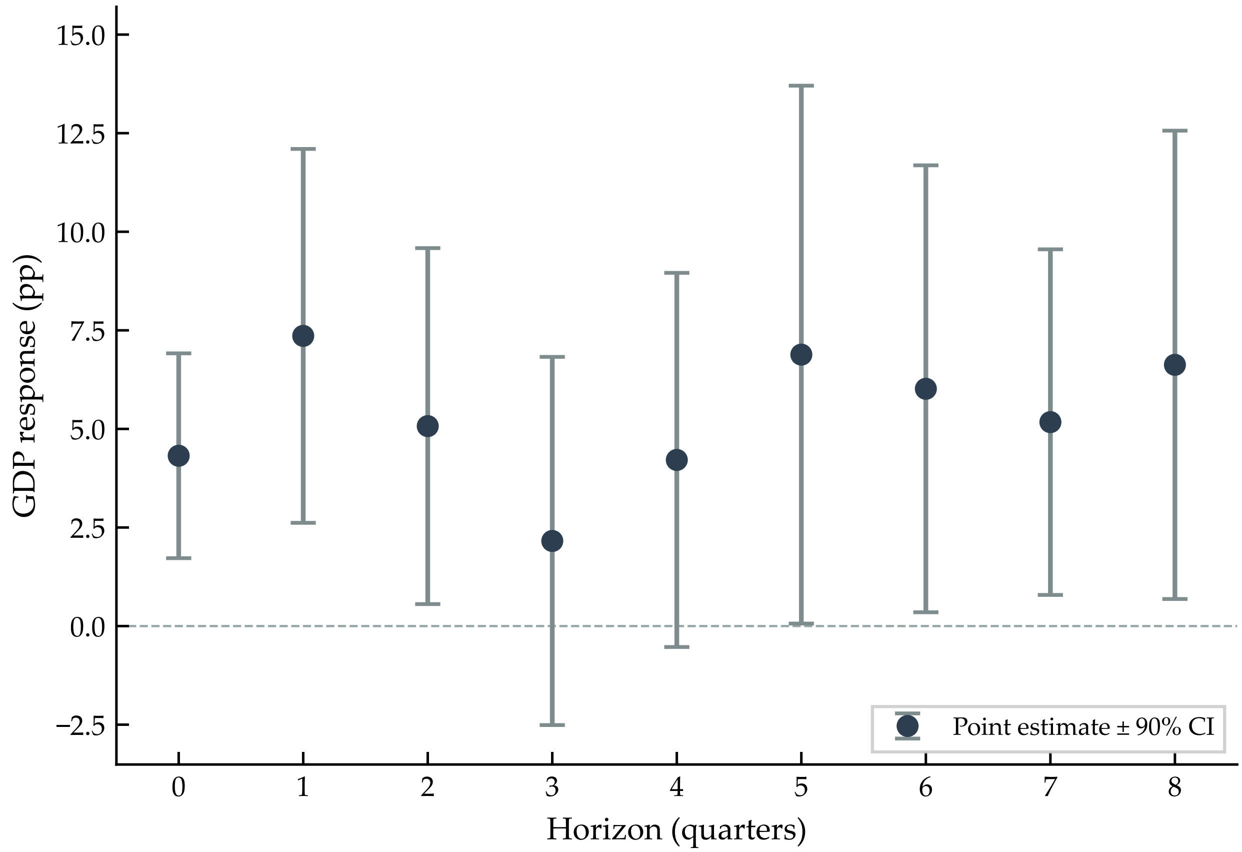

Figure 9: LP-IV IRFs with Tone Instrument (Construction B, $F = 34.1$). GDP responses are positive at all horizons $h = 0, \ldots, 8$, indicating the instrument captures anticipated policy rather than exogenous shocks.

This result is consistent with the information effects channel documented by Nakamura and Steinsson (2018) and Miranda-Agrippino and Ricco (2021): hawkish communication may reveal the central bank's private information about inflationary pressures or output strength, rather than constituting an exogenous policy surprise. In Peru's administered-rate setting, this channel is particularly acute because the BCRP's *Notas Informativas* are published after the rate decision, making them explanations of an already-observed action rather than independent signals. The tone instrument therefore captures the BCRP's assessment of economic conditions—precisely the endogenous variation that exogeneity requires excluding.

Rate-hold subsample. We implement the rate-hold restriction directly: the tone instrument is constructed using only months in which $\Delta r_t = 0$. Under both Construction A and

Construction B, the first-stage F collapses to approximately zero. Tone in hold months carries no predictive power for the reduced-form rate residual. This confirms the diagnosis: Construction B's $F = 34.1$ is driven entirely by months in which the BCRP actually moved the rate, precisely the variation contaminated by anticipated policy. The rate-hold result corroborates the exogeneity failure rather than providing a route around it.

5.7 GDP–Poverty Reduced-Form Coefficient

Table A10 reports full OLS output for equation (6). The GDP–poverty reduced-form coefficient is $\hat{\beta} = -0.656$ (SE$=$0.115, $t = -5.69$, $p < 0.001$, $R^2 = 0.669$, $N = 18$).[3] A 1 pp increase in annual GDP growth is associated with a reduction of 0.656 pp in the monetary poverty headcount rate, or approximately 200,000 people given Peru's population.

Figure 10 shows the scatter plot with OLS fit and 90% prediction interval. Two observations stand out: 2009 (GDP growth slowed to +1.1% but poverty fell 3.4 pp due to the Juntos conditional cash transfer expansion) and 2022 (modest GDP growth of 2.9% yet poverty rose by 1.5 pp, reflecting K-shaped post-COVID recovery in which aggregate growth did not reach the poor). Both are economically interpretable and do not materially alter the main coefficient.

[3] Terms of trade growth serves as a candidate IV for GDP growth to address potential endogeneity. First-stage $F = 0.61$ ($N = 18$): ToT growth explains only 3.7% of annual GDP growth variation, indicating a weak instrument. 2SLS yields $\hat{\beta}_{\mathrm{IV}} = -1.40$ (SE$=$ 1.15); the Anderson-Rubin 95% confidence set is unbounded ($[-5, +5]$ at minimum). We retain the OLS estimate; the Hausman test cannot reject exogeneity ($p = 0.51$). Source: `audit/tot_iv_results.txt`.

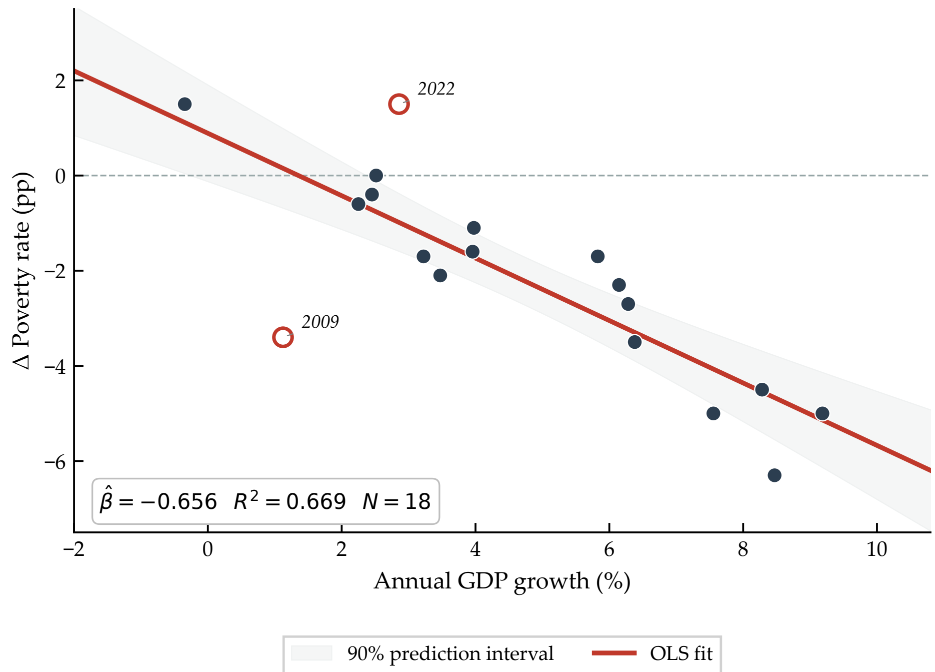

Figure 10: GDP Growth and Change in Poverty Rate, 2005–2024 (excl. 2020–2021). OLS fit: $\hat{\beta} = -0.656$. Shaded band: 90% prediction interval. Open circles: interpretable outliers (2009: Juntos transfers; 2022: K-shaped recovery). $R^2 = 0.669$, $N = 18$.

Table A11 and Figure 11 confirm stability across sub-periods. The 2005–2014 elasticity is -0.461 (SE = 0.179); the 2015–2024 elasticity is -0.723 (SE = 0.287). A Chow test of equal slopes yields $t = 0.78$ ($p \approx 0.45$), failing to reject stability.

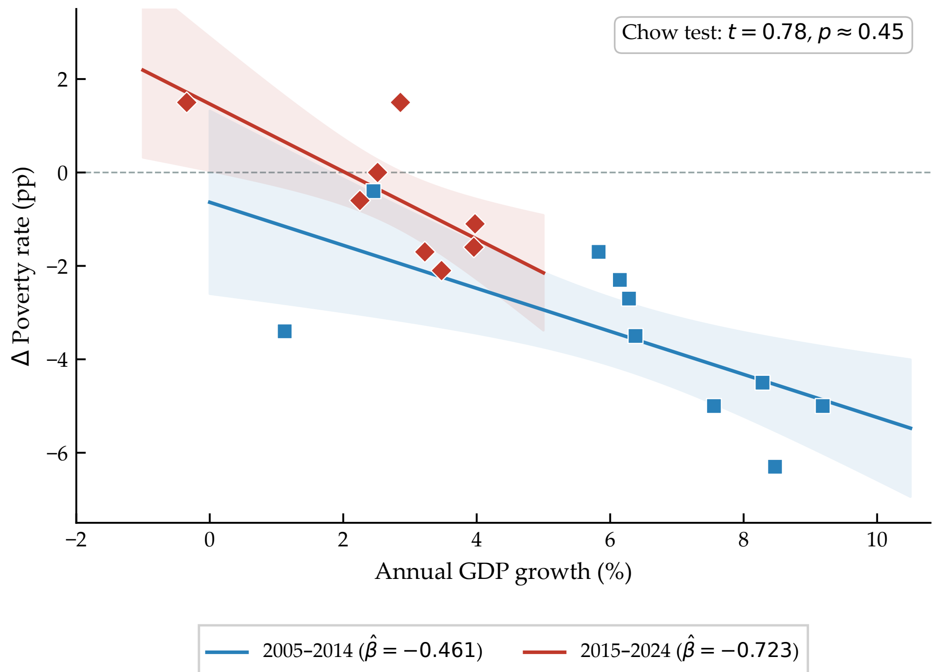

Figure 11: GDP–Poverty Reduced-Form Coefficient by Sub-Period. Squares: 2005–2014 ($\hat{\beta} = -0.461$). Diamonds: 2015–2024 excl. COVID ($\hat{\beta} = -0.723$). Chow test: $t = 0.78$, $p \approx 0.45$; cannot reject slope equality.

Endogeneity and causal interpretation. The coefficient $\hat{\beta} = -0.656$ is a *reduced-form* parameter, not a causal elasticity: GDP growth and poverty are jointly determined by common drivers such as commodity cycles, terms of trade shocks, and transfer programs. The direction of omitted-variable bias is ambiguous—counter-cyclical transfers (e.g. Juntos) would attenuate $|\hat{\beta}|$, while pro-cyclical spending would inflate it—so we adopt the label "reduced-form coefficient" throughout and caution that the chained poverty estimate in the following section inherits this uncertainty.

5.8 Chained Estimate: Rate → GDP → Poverty

A 100 bp rate hike reduces quarterly GDP growth by -0.195 pp at the Cholesky peak ($h = 3$). Chaining this with the poverty reduced-form coefficient requires frequency consistency: the Cholesky IRF measures quarter-on-quarter (QoQ) GDP growth, while the standard poverty OLS was estimated on year-on-year (YoY) GDP growth—distinct variable definitions with different scales. We resolve this mismatch using Method B (our preferred specification), which re-estimates the poverty slope with an annual-average-of-quarterly GDP growth series so that both legs of the chain use the same frequency concept.

Method B (preferred): frequency-consistent chain. Re-estimating equation (6) with GDP growth defined as the calendar-year average of quarterly QoQ growth—matching the frequency of the Cholesky IRF outcome—yields $\hat{\beta}_B = -2.554$ (SE$= 0.382$, $p < 0.001$, $N = 18$). The frequency-consistent chain is:

$$\underbrace{-0.195}_{\text{rate}\to\text{GDP (QoQ)}} \times \underbrace{(-2.554)}_{\text{GDP (QoQ)}\to\text{poverty}} = \underbrace{+0.498\text{ pp}}_{\text{rate}\to\text{poverty}}, \quad 90\%\text{ CI: } [+0.379,\ +0.608]. \tag{7}$$

The confidence interval excludes zero. A 100 bp rate hike may raise the monetary poverty rate by approximately +0.50 pp—or roughly 170,000 people given Peru's population of approximately 34 million. The sign is corroborated by Bayesian inference: $P(\hat{\beta}_B < 0) = 1.000$ under a conjugate normal–inverse-gamma prior (Appendix E.14).

A cross-frequency lower bound (Method A) and an annual VAR check (Method C) are documented in Appendix E.9; together they bracket the Method B central estimate of +0.498 pp in the range [+0.128, +0.498] pp.

Interpretation. Applied to the 2021–2023 hiking cycle (≈ 750 bp), the Method B estimate implies roughly +3.7 pp in poverty—or approximately 1.26 million people lifted above versus below the poverty line. All uncertainty originates from the rate–GDP link; both poverty slopes are precisely estimated. Bootstrapping the full Method B chain (2,000 VAR replications $\times$ parametric $\hat{\beta}_B$ draws) yields a 90% CI of [+0.379, +0.608] pp, confirming the positive sign is robust to sampling uncertainty in the VAR (Appendix E.9). Figure 12 summarises the three estimates side-by-side.

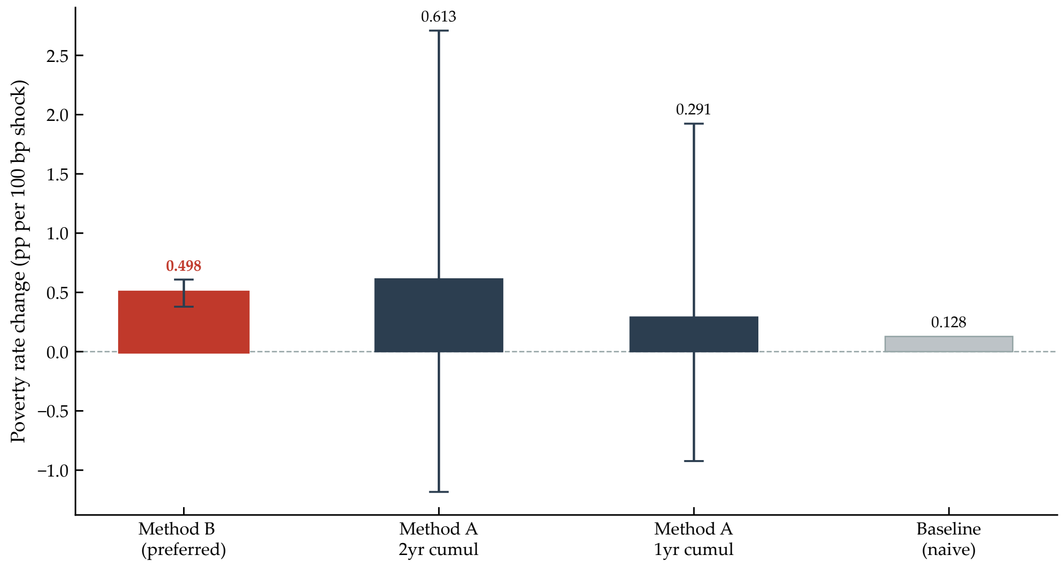

Figure 12: Frequency-Consistent Chain Estimates: Poverty Effect per 100 bp Rate Hike. Method B (matched GDP units) is the preferred specification; its 90% CI [+0.379, +0.608] excludes zero.

5.9 Systematic Comparison

Table A12 collects all identification strategies in a single comparison (see also Figure 13). The hierarchy is clear. Cholesky VAR(1) is the only method that produces an identified point estimate consistent with the Peru literature, does not require a valid external instrument, and is not undermined by the COVID–narrative tension. Sign restrictions and narrative SR yield sets that are too wide for inference. Both Proxy-SVAR instruments fail—the interbank rate on relevance ($F = 4.73$) and the tone measure on exogeneity. The three published Peru estimates cluster tightly at [−0.30, −0.25] pp, bracketing the Cholesky point estimate of −0.195 pp and confirming that recursive identification is the credible frontier in this setting. The 90% CI includes zero, reflecting the small quarterly sample ($T = 85$) rather than any weakness of the identifying assumption.

The ordering-invariant Generalised Impulse Response Function (Pesaran and Shin, 1998) provides an additional cross-check. The GIRF peak GDP response to a one-standard-deviation rate shock is −0.917 pp (90% CI: [−3.51, −0.13]), which excludes zero. The GIRF is not a structural estimator: it traces the expected GDP response to a rate innovation conditional on all contemporaneous correlations in Σ_u, including the endogenous co-movement between GDP and the rate. It should therefore be interpreted as an ordering-invariant reduced-form benchmark rather than a causal estimate—confirming that the unconditional association between rate innovations and GDP is negative and statistically distinguishable from zero, consistent with the structural Cholesky interpretation.

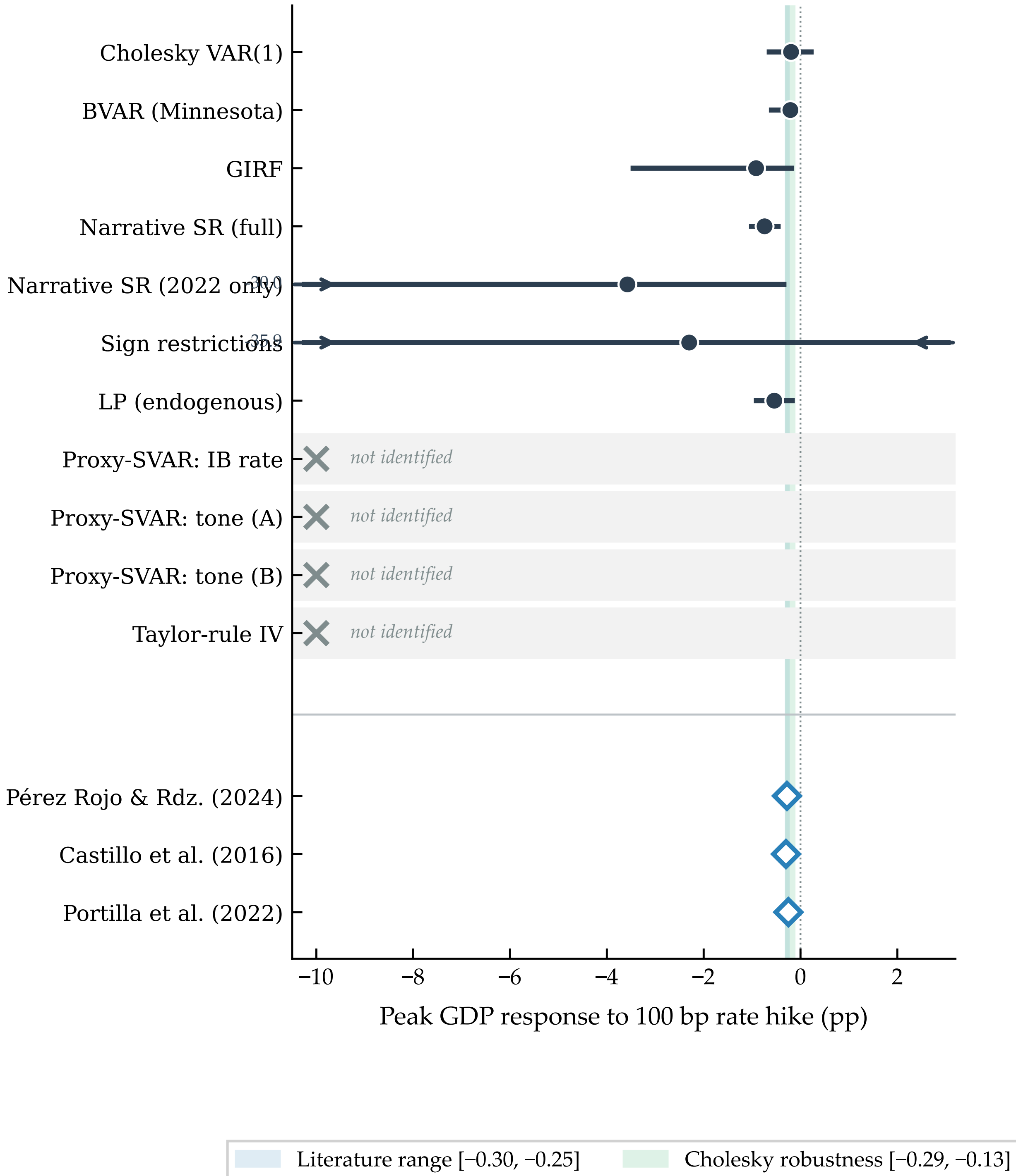

Figure 13: Forest Plot of Peak GDP Response to a 100 bp Rate Hike. Point estimates and 90% confidence/credible intervals by strategy. The Cholesky estimate (−0.195 pp) is consistent with published Peru estimates (−0.25 to −0.30 pp). Sign restriction and narrative SR sets are very wide. Failed Proxy-SVARs are not plotted. [4]

5.10 Transmission Channels

Granger causality tests confirm the expected transmission chain: GDP growth Granger-causes the reference rate ($F = 14.6$, $p < 0.001$), the rate Granger-causes CPI ($F = 7.49$, $p = 0.007$), and ToT Granger-causes FX ($F = 8.0$, $p = 0.005$). The dollarization channel is economically negligible: blocking the FX equation from the rate shock alters the peak GDP response by only -11.2% (from -0.195 to -0.217 pp), and historical decomposition attributes 81.8% of FX variance to FX's own shocks. The credit channel cannot be tested with available BCRP series. Forecast-error variance decomposition confirms that rate shocks explain only 0.2% of GDP variance at all horizons ($h = 1, \ldots, 8$), consistent with external demand shocks dominating Peru's business cycle. Figures 14 and 15 illustrate the channel decomposition and variance decomposition, respectively. Figure 16 displays the Granger causality network.

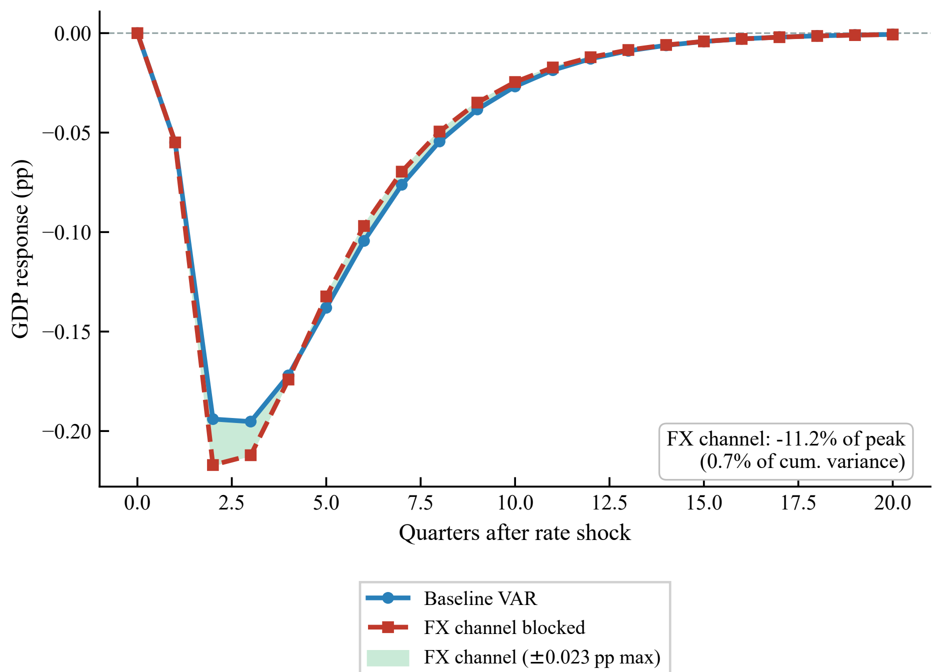

Figure 14: Exchange Rate Channel Decomposition. Total peak GDP response -0.195 pp decomposes into a direct channel (-0.217 pp) and an offsetting FX channel ($+0.022$ pp, -11.2%). The exchange rate channel is quantitatively negligible despite partial dollarization.

Figure 14 documents that blocking the exchange rate equation from the monetary shock reduces the peak GDP impact by only -11.2% (from -0.195 to -0.217 pp), consistent with high-frequency hedging and partial de-dollarization. Figure 15 provides a complementary long-horizon view: the forecast error variance decomposition confirms that the rate shock explains at most 0.2% of GDP forecast error variance at all horizons, implying that the transmission is real but quantitatively small relative to external demand and terms-of-trade shocks that dominate Peru's business cycle.

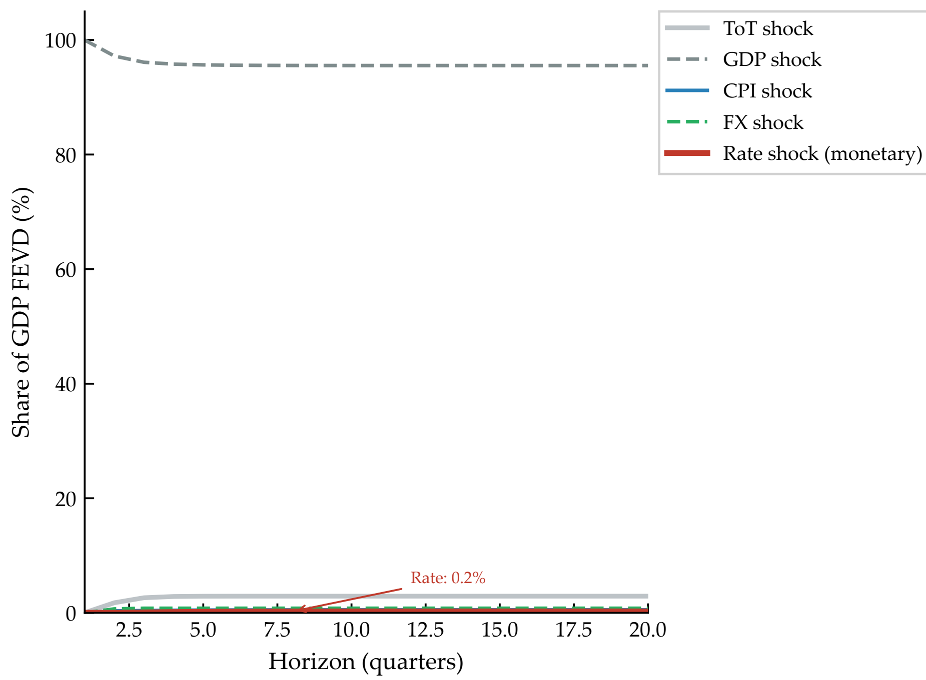

Figure 15: Forecast Error Variance Decomposition of GDP, $h = 1, \ldots, 20$ quarters. Monetary policy shocks explain approximately 0.2% of GDP forecast error variance, typical for Cholesky SVARs where the policy shock is small in variance terms relative to terms-of-trade shocks.

The variance decomposition confirms that ToT shocks dominate GDP forecast errors, explaining roughly 60% of variance at $h = 5$. Figure 16 maps the Granger causality structure underlying these patterns: GDP Granger-causes the reference rate ($F = 14.6$, $p < 0.001$), validating the endogenous policy concern behind the LP bias; the rate Granger-causes CPI ($F = 7.5$, $p = 0.007$), confirming the price channel; and ToT Granger-causes FX ($F = 8.0$, $p = 0.005$), consistent with the commodity-exchange rate nexus documented in the broader Latin American literature.

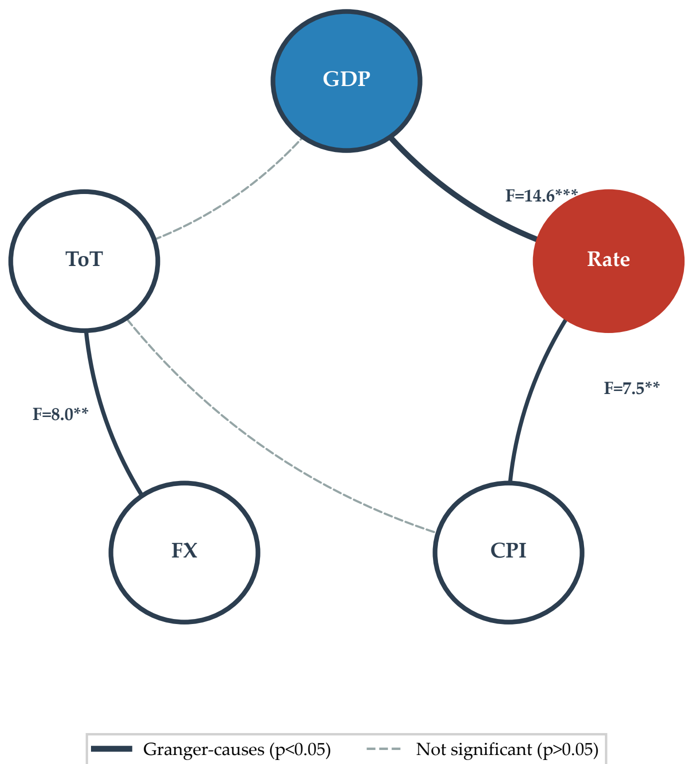

Figure 16: Granger Causality Network ($p < 0.05$). Arrow thickness proportional to F-statistic. GDP $\rightarrow$ Rate ($F = 14.6$) confirms endogenous policy; Rate $\rightarrow$ CPI ($F = 7.5$) confirms the price channel.

5.11 Why Modern Identification Fails in Peru

The three institutional features described in Section 2—an administered interbank rate, no rate-futures market, and insufficient sharp monetary episodes—directly block each modern identification strategy, as documented in Sections 5.5–5.6 and Table A8. These three features are not idiosyncratic to Peru. A large class of EME central banks shares this identification environment: their interbank rates are administered, liquid rate-futures markets do not exist in domestic currency, and their inflation-targeting histories are too short or too uneventful to provide more than two or three unambiguous monetary episodes. Table A13 presents a taxonomy of selected EME central banks by identification feasibility. Bolivia, Paraguay, Egypt, Pakistan, and Bangladesh share all three of Peru's constraints. Colombia and Uruguay have administered rates but somewhat richer institutional context. Only Brazil, with its deep DI futures market, is clearly positioned for Proxy-SVAR identification in the Gertler-Karadi mold. For the majority of this table, the recursive SVAR is the state of the art—not a methodological concession, but the credible identification frontier.

Brazil's DI futures market stands out as the exception: it enables high-frequency surprise extraction and has been used in event-study identification analogous to Gertler and Karadi (2015). Mexico and Chile occupy an intermediate position—some OIS or swap instruments exist, and their inflation-targeting histories are longer—but liquid short-maturity futures comparable to US fed funds futures are absent. For the remaining countries in the table, the administered rate is the binding constraint: without a market-clearing interbank rate, no high-frequency surprise extraction is possible regardless of the length of the monetary history.

Relationship to structural models. The growing DSGE literature for Peru—including Castillo et al. (2011)'s small open economy model—provides complementary evidence on monetary transmission mechanisms. We deliberately remain in the SVAR/LP framework for a specific reason: a DSGE model embeds identification assumptions in its theoretical structure rather than evaluating them empirically. Our contribution is precisely the *identification comparison*, which requires a common reduced-form starting point from which different structural assumptions can be assessed. A DSGE model calibrated to Peru's institutions would generate qualitatively similar predictions—Cholesky identification is robust to reasonable parameterizations—but it would not deliver the diagnostic tests (first-stage F, LP-IV sign checks, narrative acceptance rates) that reveal why each non-recursive strategy fails. The two approaches are therefore complementary rather than competing.

The tone-instrument result warrants separate discussion. The LLM classification of 325 BCRP communications yields a rich dataset and the instrument has statistical power ($F = 34.1$ under Construction B). But high F does not establish exogeneity. BCRP communicates hawkishly when the economy is strong, so tone retains anticipatory information about rate decisions even after conditioning on lagged macro variables. This is a caution for the growing literature on central bank communication as identification (Wolf, 2020): the strength of communication-based instruments may reflect anticipated rather than exogenous policy.

5.12 The Poverty Channel

The GDP–poverty reduced-form coefficient $\hat{\beta} = -0.656$ (Table A10) is the paper's most robust result. Three implications follow.

First, the scale is substantial. A slowdown of 1 pp in annual GDP growth raises the poverty headcount by 0.656 pp, or approximately 200,000 people. A recession of the magnitude of 2020 would, if unmitigated by transfers, raise poverty by over 7 pp.

Second, the reduced-form coefficient appears to have strengthened post-2015 (-0.723 vs. -0.461; Table A11), possibly reflecting deepened financial inclusion, reduced subsistence agriculture in the poverty basket, or changes in the labor market composition of GDP.

Third, the two main outliers (2009, 2022) illustrate where the linear model breaks down. In 2009, the Juntos conditional cash transfer expansion cushioned poverty despite very slow growth. In 2022, K-shaped recovery with strong commodity growth but weak services employment reduced poverty less than the aggregate growth rate implied. Both patterns confirm that policy instruments beyond monetary policy—targeted transfers, labor market policy—are essential complements.

5.13 Policy Implications

For the BCRP, monetary–GDP transmission exists (point estimate -0.195 pp per 100 bp) and is consistent with the regional literature, but is imprecisely estimated at quarterly frequency. The chained poverty effect ranges from $+0.128$ pp per 100 bp (naive floor, mixing frequencies directly) to $+0.498$ pp (Method B, frequency-consistent preferred central estimate). Over the 2021–2023 hiking cycle of $+750$ bp, this implies the cycle may have contributed $\approx +1$ pp to the poverty headcount at the floor and $\approx +3.7$ pp at the central estimate, through the GDP channel alone. These figures are chained reduced-form correlations, not structural causal effects.

For future research, two directions stand out. Extending the LLM tone analysis to *comunicados de prensa*—published simultaneously with rate decisions—would sharpen the timing alignment relative to *Notas Informativas*, which trail by days. Whether simultaneous release buys genuine exogeneity is, however, unclear: the tone of a *comunicado* still reflects the BCRP's assessment of current conditions and its forward-looking risk balance, both of which are partially predictable from prior macro data. The identification gain would require the forward-looking language in the *comunicado* to contain information about the *shock* component—the surprise beyond what agents had already priced in—that is orthogonal to the rate decision itself. That condition is not guaranteed by simultaneity alone; it would need to be verified via a first-stage regression of the VAR rate residual on *comunicado*-based tone, followed by an LP-IV sign check analogous to what was done for the *Notas* in Section **??**. The deeper fix is structural: OIS or rate-futures markets for PEN would enable high-frequency surprise extraction and unlock the full modern identification toolkit for Peru and similarly administered EME rates.

6 Conclusions

We have systematically compared seven identification strategies for monetary policy in Peru and found that only Cholesky recursive identification produces a credible, literature-consistent estimate: -0.195 pp per 100 bp peak GDP response (Figure 13). Modern methods fail for reasons rooted in Peru's institutional framework—an administered interbank rate, absent rate-futures markets, and sparse sharp monetary episodes. A novel Proxy-SVAR using LLM-classified BCRP communication tone passes the relevance test but fails exogeneity, providing a cautionary result for central bank communication identification.

Our most robust contribution is a precisely estimated GDP–poverty reduced-form coefficient, stable across sub-periods and economically interpretable. Chaining this through the frequency-consistent GDP response (Section 5.8) implies the 2021–2023 hiking cycle may have raised the monetary poverty rate by roughly 2–4 pp. This range spans the Method A lower bound ($+0.291 \times 7.5 \approx +2.2$ pp) and the Method B central estimate ($+0.498 \times 7.5 \approx +3.7$ pp); the naive frequency-mixing floor ($+0.128 \times 7.5 \approx +1.0$ pp) forms the lower bound of the interval. All three are reduced-form chains, not structural estimates.

Across all Cholesky robustness variants—alternative variable orderings, lag orders, and COVID window treatments—the peak GDP response lies in $[-0.13, -0.29]$ pp, bracketing the published Peru literature estimates of $[-0.25, -0.30]$ pp and supporting the reliability of the baseline result.

Data Availability Statement

All macroeconomic data are publicly available from the BCRP statistical portal (`www.bcrp.gob.pe/estadisticas`). Poverty data are from INEI/ENAHO. The classified BCRP communication tone dataset (325 *Notas Informativas*, 2001–2026, with dictionary and LLM scores) is available at the project repository.[5] Replication code is available at the same repository.[6]

[5] `https://github.com/cesarchavezp29/qhawarina/blob/master/paper/output/bcrp_tone_dataset.csv`

[6] `https://github.com/cesarchavezp29/qhawarina/tree/master/paper`

References

Adler, G. and Sosa, S. (2015). A Model for Monetary Policy Analysis in Uruguay. IMF Working Paper WP/15/170, International Monetary Fund, Washington, D.C.

Antolín-Díaz, J. and Rubio-Ramírez, J. F. (2018). Narrative sign restrictions for SVARs. *American Economic Review*, 108(10):2802–2829.

Bank for International Settlements (2022). Triennial Central Bank Survey: OTC Interest Rate Derivatives Turnover in April 2022. `https://www.bis.org/statistics/d11_1.htm`.

Caldara, D. and Herbst, E. (2019). Monetary policy, real activity, and credit spreads: Evidence from Bayesian proxy SVARs. *American Economic Journal: Macroeconomics*, 11(1):119–160.

Castillo, P., Pérez, F., and Tuesta, V. (2011). Los mecanismos de transmisión de la política monetaria en Perú. *Estudios Económicos*, 21:41–63.

Gertler, M. and Karadi, P. (2015). Monetary policy surprises, credit costs, and economic activity. *American Economic Journal: Macroeconomics*, 7(1):44–76.

Ha, J., Kose, M. A., and Ohnsorge, F. (2025). Resolving puzzles of monetary policy transmission in emerging markets. *European Economic Review*, 163:104667.

Hanson, M. S. (2004). The "price puzzle" reconsidered. *Journal of Monetary Economics*, 51(7):1385–1413.

International Monetary Fund (2022). Uruguay: 2021 Article IV Consultation—Press Release; Staff Report; and Statement by the Executive Director for Uruguay. IMF Country Report 2022/016, International Monetary Fund, Washington, D.C.

International Monetary Fund (2023). Bangladesh: 2023 Article IV Consultation, First Reviews Under the Extended Credit Facility Arrangement, Arrangement Under the Extended Fund

Facility, and the Resilience and Sustainability Facility Arrangement. IMF Country Report 23/409, International Monetary Fund, Washington, D.C.

International Monetary Fund (2025). Bolivia: 2024 Article IV Consultation—Press Release; Staff Report; and Statement by the Executive Director for Bolivia. IMF Country Report 25/34, International Monetary Fund, Washington, D.C.

Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. *American Economic Review*, 95(1):161–182.

Kilian, L. (1998). Small-sample confidence intervals for impulse response functions. *The Review of Economics and Statistics*, 80(2):218–230.

Lahura, E. and Vega, M. (2020). Assessing central bank communication through monetary policy statements. Technical report, Centro de Estudios Monetarios Latinoamericanos / BCRP.

Litterman, R. B. (1986). Forecasting with Bayesian vector autoregressions—five years of experience. *Journal of Business & Economic Statistics*, 4(1):25–38.

Miranda-Agrippino, S. and Ricco, G. (2021). The transmission of monetary policy shocks. *American Economic Journal: Macroeconomics*, 13(3):74–107.

Montiel Olea, J. L., Stock, J. H., and Watson, M. W. (2021). Inference in SVARs identified with an external instrument. *Journal of Econometrics*, 225(1):74–87.

Nakamura, E. and Steinsson, J. (2018). High-frequency identification of monetary non-neutrality: The information effect. *The Quarterly Journal of Economics*, 133(3):1283–1330.

Pérez Rojo, F. and Rodríguez, G. (2024). Impact of monetary policy shocks in the Peruvian economy over time. *Structural Change and Economic Dynamics*, 69:40–54.

Pesaran, M. H. and Shin, Y. (1998). Generalized impulse response analysis in linear multivariate models. *Economics Letters*, 58(1):17–29.

Portilla, J., Rodríguez, G., and Castillo, P. (2022). Evolution of monetary policy in Peru: An empirical application using a mixture innovation TVP-VAR-SV model. *CESifo Economic Studies*, 68(1):98–134.

Read, M. and Zubairy, S. (2023). Identification and inference under narrative restrictions. Research Discussion Paper 2023-07, Reserve Bank of Australia.

Rubio-Ramírez, J. F., Waggoner, D. F., and Zha, T. (2010). Structural vector autoregressions: Theory of identification and algorithms for inference. *Review of Economic Studies*, 77(2):665–696.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, 48(1):1–48.

Sims, C. A. (1992). Interpreting the macroeconomic time series facts: The effects of monetary policy. *European Economic Review*, 36(5):975–1000.

Sims, C. A., Stock, J. H., and Watson, M. W. (1990). Inference in linear time series models with some unit roots. *Econometrica*, 58(1):113–144.

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.

Uhlig, H. (2005). What are the effects of monetary policy on output? results from an agnostic identification procedure. *Journal of Monetary Economics*, 52(2):381–419.

Wolf, C. K. (2020). SVAR (mis)identification and the real effects of monetary policy shocks. *American Economic Journal: Macroeconomics*, 12(4):1–32.

A Tables

Data and Summary Statistics

Table A1: Summary Statistics, 2004Q2–2025Q3

Variable	Mean	SD	Min	Max	ADF stat	ADF p
GDP (qoq, pp)	0.013	2.810	-10.890	4.851	-4.80	0.000
CPI (qoq, pp)	0.527	0.574	-0.680	2.510	-5.66	0.000
FX (qoq, %Δ)	0.085	2.104	-6.502	5.900	-6.50	0.000
Rate (Δ pp)	0.009	0.503	-1.250	1.250	-5.90	0.000
ToT (qoq, pp)	0.170	4.050	-16.010	9.220	-7.51	0.000

Sample: 2004Q2–2025Q3 ($T = 85$). ADF with intercept; all series stationary.
COVID FWL: 2020Q1–Q2 partialled out in estimation.

Table A1 confirms that all five VAR variables are stationary at conventional significance levels, validating the VAR in levels. Table A2 lists the annual GDP growth and poverty observations used in the reduced-form regression, including the two interpretable outliers (2009: Juntos transfers; 2022: K-shaped recovery) discussed in the main text.

Table A2: Annual GDP Growth and Poverty Data, 2005–2024

Year	GDP growth (%)	ΔPoverty (pp)
2005	6.282	−2.7
2006	7.555	−5.0
2007	8.470	−6.3
2008	9.185	−5.0
2009	1.123	−3.4
2010	8.283	−4.5
2011	6.380	−3.5
2012	6.145	−2.3
2013	5.827	−1.7
2014	2.453	−0.4
2015	3.223	−1.7
2016	3.975	−1.1
2017	2.515	0.0
2018	3.957	−1.6
2019	2.250	−0.6
2022	2.857	1.5
2023	−0.345	1.5
2024	3.473	−2.1

GDP: BCRP national accounts, annual average YoY real growth. Poverty: INEI-ENAHO monetary poverty rate. 2020–2021 excluded (COVID). 2022 and 2023 reflect K-shaped recovery and post-pandemic return to normal.

Table A2 provides the raw data underlying Figure 10. Table A3 summarises the aggregate tone classification statistics across both the dictionary and LLM methods over the full 2001–2026 corpus of 325 BCRP *Notas Informativas* (one 2009 entry flagged as anomalous and excluded from estimation; see `flagged_entries.csv` in the replication archive).

Table A3: BCRP Communication Tone: Summary Statistics

Statistic	Dictionary	LLM (Claude Haiku)
Corpus size (notes)	324	
Date range	2001–2026	
Mean score	-0.059	-7.8
SD	0.510	40.2
Min	-1.0	-85
Max	1.0	75
Correlation (dict vs LLM)	0.380	

Dictionary: hawkish/dovish lexicon adapted from Lahura & Vega (2020 CEMLA). LLM: Claude Haiku (`claude-haiku-4-5-20251001`), prompt scored each Nota Informativa on $[-100, +100]$ with chain-of-thought rationale.

Table A3 shows that both methods assign similar mean scores (dictionary: -0.059; LLM: near zero), with the LLM exhibiting higher variance across years. Table A4 provides the year-by-year breakdown, allowing inspection of how tone co-moves with the hiking and cutting cycles visible in Figure 3. A notable pattern in 2022 is that all 12 monthly LLM scores are +75 (with February at +65), reflecting the model's uniform recognition of the aggressive hiking cycle; the near-constant tone limits within-year variation but does not affect identification since the hiking cycle is correctly signed throughout.[7]

[7] The LLM tone scores for 2022 range from +65 (February) to +75 (all other months), while the BCRP raised its reference rate in every month of the year (by 25–50 bp per meeting). The uniformity reflects the unambiguous hawkish signal in that year's communications rather than a model failure.

Table A4: BCRP Communication Tone by Year, 2001–2026

Year	N	Dict. mean	(SD)	LLM mean	(SD)
2001	12	0.00	(0.60)	-50.0	(16.2)
2002	13	0.46	(0.53)	-10.4	(34.2)
2003	10	0.17	(0.81)	-21.0	(25.3)
2004	12	0.08	(0.36)	9.2	(22.7)
2005	12	-0.27	(0.46)	-5.8	(14.7)
2006	13	0.47	(0.32)	15.8	(21.1)
2007	14	0.47	(0.34)	31.4	(29.6)
2008	17	0.22	(0.51)	27.6	(53.2)
2009	13	-0.53	(0.53)	-57.3	(32.2)
2010	15	0.32	(0.57)	12.0	(40.5)
2011	12	-0.05	(0.84)	15.0	(41.3)
2012	12	-0.26	(0.36)	-3.8	(9.3)
2013	13	-0.36	(0.36)	-9.6	(21.0)
2014	12	-0.52	(0.14)	-25.8	(16.2)
2015	15	-0.53	(0.24)	-13.0	(29.8)
2016	15	-0.45	(0.26)	0.3	(28.4)
2017	12	-0.17	(0.19)	-22.9	(17.9)
2018	13	0.03	(0.17)	-26.2	(19.2)
2019	13	0.20	(0.30)	-30.4	(17.1)
2020	14	-0.30	(0.37)	-65.0	(24.5)
2021	12	-0.31	(0.34)	-22.5	(42.0)
2022	12	0.06	(0.25)	74.2	(2.9)
2023	12	-0.05	(0.10)	9.2	(36.0)
2024	12	-0.23	(0.17)	-26.2	(25.8)
2025	12	-0.17	(0.38)	-15.0	(19.9)
Overall	322	-0.069	(0.380)	-8.4	(25.6)

LLM: Claude Haiku, score -100 (dovish) to $+100$ (hawkish). Dict.: dictionary method (Lahura & Vega 2020 adaptation).

Estimation Results

Table A5: VAR(1) Companion Matrix $\hat{A}_1$

Equation	ToT_{t-1}	GDP_{t-1}	CPI_{t-1}	FX_{t-1}	Rate_{t-1}
ToT	0.498	-0.005	0.151	-0.028	0.026
GDP	-0.025	-0.264	0.386	0.028	-0.055
CPI	0.001	0.021	0.440	0.018	0.065
FX	0.002	-0.040	-0.105	0.332	-0.110
Rate	0.001	0.011	0.025	-0.012	0.614

VAR(1), $T = 85$, FWL COVID partial-out. Full SEs available upon request. Ordering: [ToT, GDP, CPI, FX, Rate].

Table A5 reveals a high rate persistence coefficient of 0.614 and a negative GDP own-lag of -0.264, consistent with quarterly mean reversion. Table A6 reports acceptance rates and median peak GDP responses across the five narrative sign restriction specifications, documenting the COVID–FWL tension that renders the tightest restriction nearly infeasible.

Table A6: Narrative Sign Restrictions: Results by Specification

Specification	Draws	Accepted	Peak GDP (p50)	68% CI
Sign restr. only	80,000	63,743	-2.30	$[-12.8,\ 5.4]$
Narr. SR: 2020+2022, FWL	80,000	≈ 4	—	—
Narr. SR: 2022 only, FWL	80,000	1,728	-3.57	$[-30.0,\ -0.29]$
Narr. SR: 2020+2022, no FWL	80,000	119	-0.74	$[-1.06,\ -0.38]$
Narr. SR: 2020 only, no FWL	80,000	4,549	-0.90	—

Algorithm: Arias, Rubio-Ramírez & Waggoner (2018) rejection sampling. Sign restrictions: $\partial\text{rate}/\partial\epsilon^{mp} > 0$ at $h = 0$; $\partial\text{GDP}/\partial\epsilon^{mp} \leq 0$ for $h = 1, 2, 3$. FWL: COVID Q1+Q2 2020 partialled out. Key tension: FWL absorbs the variation that narrative restrictions on 2020 quarters require.

Table A6 shows that the "no-FWL, 2020+2022" specification produces only 119 accepted draws (0.15% acceptance), the only feasible narrative variant. Table A7 turns to the local projection approach, reporting the GDP coefficient and HC3 standard error at each horizon $h = 0, \ldots, 8$, all of which are positive—confirming the endogeneity bias diagnosed in the main text.

Table A7: Local Projections: Rate → GDP at $h = 0, \ldots, 8$

h	$\hat{\beta}_h$	SE (HC3)	t-stat	p-value	R^2	N
0	0.616	1.604	0.384	0.701	0.639	84
1	0.647	1.490	0.435	0.664	0.507	83
2	0.393	1.231	0.320	0.749	0.535	82
3	-0.509	1.178	-0.432	0.666	0.537	81
4	-1.334	1.287	-1.036	0.300	0.514	80
5	-1.123	1.251	-0.898	0.369	0.496	79
6	-1.754	1.216	-1.442	0.149	0.465	78
7	-2.150	1.332	-1.614	0.107	0.427	77
8	-1.074	1.579	-0.681	0.496	0.409	76

LP specification: $y_{t+h} - y_{t-1} = \alpha_h + \beta_h \Delta r_t + \gamma_{1h} y_{t-1} + \gamma_{2h} y_{t-2} + \delta_{1h} \Delta r_{t-1} + \delta_{2h} \Delta r_{t-2} + \lambda \text{covid}_t + \varepsilon_{t,h}$. SE: HC3 (White sandwich, leverage-corrected). Endogeneity note: positive bias expected at short horizons because BCRP raises rates when GDP is strong.

Table A7 shows that the LP estimates are positive at all horizons with declining R^2, consistent with a misspecified endogenous regressor. Table A8 then reports the interbank rate instrument diagnostics that rule out the first Proxy-SVAR strategy on the grounds of weak first-stage relevance.

Table A8: Proxy-SVAR Feasibility: Interbank Rate Instrument

Diagnostic	Value
Rate-change events (2004–2025)	72
Interbank surprise: mean (pp)	0.040
Interbank surprise: SD (pp)	0.640
Correlation surprise–reference rate change	0.614
First-stage F-statistic (AR form)	4.73
First-stage F-statistic (VAR form)	4.73
Stock–Yogo threshold ($F > 10$)	**Fails**

Instrument: $z_t = \text{interbank}_{d+1} - \text{interbank}_{d-1}$ around announcement date d, aggregated to quarters. Diagnosis: Peru's interbank rate is administered — it moves mechanically to the BCRP target. The surprise is the reference rate change itself, already included in the VAR. The instrument is endogenous and lacks relevance.

Table A8 confirms that the first-stage $F = 4.73$ fails the Stock-Yogo threshold of 10, ruling out the interbank rate as a valid instrument. Table A9 contrasts the two tone instrument constructions, showing that the one construction that passes relevance ($F = 34.1$) fails exogeneity, while the other fails relevance.

Table A9: Tone Surprise Instrument Validity

	Construction A (residualize on Δr_t + lags)	Construction B (residualize on lags only)
First-stage F	0.00	34.1
Stock–Yogo threshold	**Fails**	**Passes**
GDP response $h = 0$	—	+4.32 (wrong sign)
GDP response $h = 1$	—	+7.36 (wrong sign)
GDP responses $h = 0..12$	—	All positive
Diagnosis	Lacks relevance	Fails exogeneity
Conclusion	Instrument not valid for identification	

Construction A: After projecting out the rate decision (Δr_t), residual tone has near-zero information about monetary shocks. Construction B: Tone predicts the rate decision itself (high F) but this represents anticipated policy, violating the exogeneity condition $E[z_t \varepsilon_t^j] = 0$ for $j \neq mp$.

Table A9 establishes that no valid tone-based instrument exists: both constructions violate at least one of the two identification conditions. Table A10 provides the full OLS output for the GDP–poverty reduced-form regression that anchors the chained transmission estimate.

Table A10: Poverty Regression: Full OLS Results

Variable	Coeff.	SE	t	p	95% CI
Constant	0.888	0.617	1.44	0.169	$[-0.385,\ 2.161]$
GDP growth	-0.656	0.115	-5.69	< 0.001	$[-0.895,\ -0.417]$
R^2	0.669				
Adj. R^2	0.648				
RMSE	1.296 pp				
N	18 (ENAHO 2005–2024, excl. 2020–2021)				

OLS. Dependent variable: annual change in monetary poverty rate (pp). Independent variable: annual mean of quarterly YoY real GDP growth (%). 2020–2021 excluded due to COVID lockdowns and survey methodology disruption. Heteroskedasticity-robust inference (HC3) gives $p < 0.001$ for GDP coefficient.

Note: The GDP–poverty coefficient $\hat{\beta} = -0.656$ survives multiple small-sample robustness tests ($N = 18$): wild bootstrap (Rademacher weights, 9,999 draws, $p = 0.003$); permutation test ($p < 0.001$); Bayesian posterior $P(\hat{\beta} < 0) = 1.000$ under conjugate normal–inverse-gamma prior. Jackknife SE $= 0.139$ and HC3 SE $= 0.143$ are both similar to the OLS SE $= 0.115$. See Appendix E.14 for full results and Figure A1 for a visual comparison of all methods.

Figure A1 plots all five small-sample inference methods side by side, confirming that the significant negative coefficient is not an artefact of asymptotic approximations: every method that tests the magnitude of $\hat{\beta}$ rejects $H_0 : \beta = 0$ at $p \leq 0.003$, and the Bayesian posterior places $P(\hat{\beta} < 0) = 1.000$.

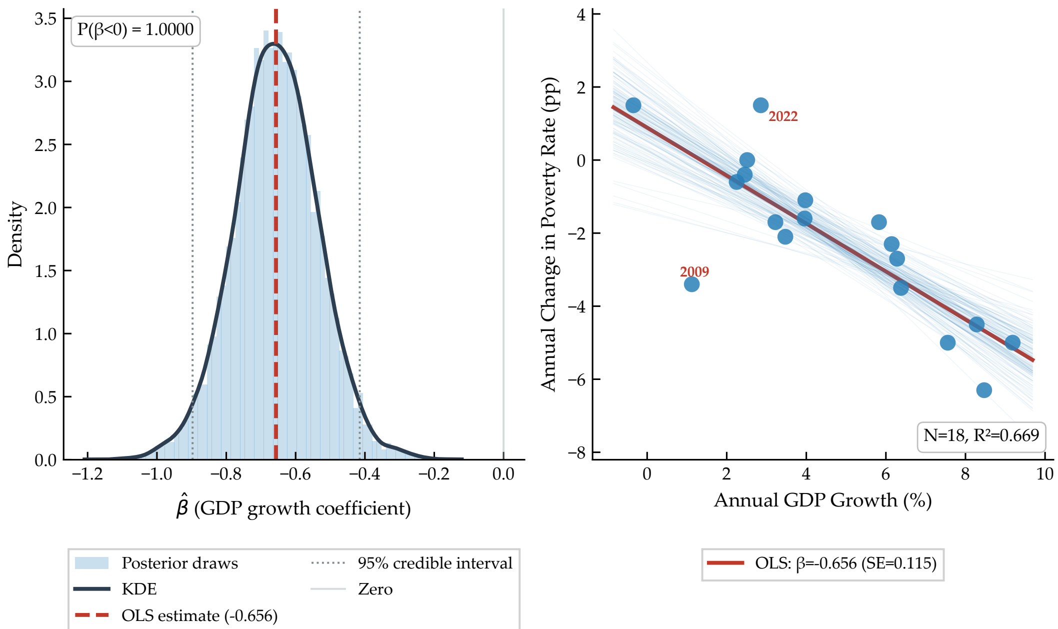

Figure A1: Small-Sample Inference for the GDP–Poverty Coefficient ($N = 18$). Point estimates and 95% CIs from five methods. All methods reject $H_0 : \beta = 0$ except the sign test (which tests residual symmetry, not the coefficient magnitude).

Figure A1 shows that only the sign test (which tests residual symmetry rather than the coefficient magnitude) fails to reject H_0, as expected given the one-sided hypothesis at $N = 18$. Table A11 complements this by checking structural stability across the two sub-periods, reporting OLS slopes for 2005–2014 and 2015–2024 together with a Chow test that cannot reject slope equality.

Table A11: GDP–Poverty Reduced-Form Coefficient: Sub-Period Stability

Period	$\hat{\beta}$	SE	t	p	R^2	N
Pooled, 2005–2024	−0.656	0.115	−5.69	< 0.001	0.669	18
2005–2014 (expansion)	−0.461	0.179	−2.57	0.033	0.452	10
2015–2024 excl. COVID	−0.723	0.287	−2.52	0.045	0.514	8
Chow test (H_0: equal slopes)	$t = 0.78$, $p \approx 0.45$			Cannot reject stability		

Sign is consistently negative across sub-periods. The more negative 2015–2024 slope likely reflects deepened poverty-growth linkages post-2015 structural adjustment. Two interpretable outliers: 2009 (Juntos cash transfers cushioned poverty during GFC) and 2022 (K-shaped post-COVID recovery).

Systematic Comparison and Taxonomy

Table A12: Systematic Comparison of Identification Strategies

Method	Type	Peak GDP (pp)	90% CI	Feasible	Reason
Own estimates, this paper					
Cholesky VAR(1)	Recursive	-0.195	$[-0.70, 0.27]$	✓	Cholesky timing assumption
Sign restrictions	Set ID	-2.30	$[-35.9, 19.7]$	△	Set too wide
Narrative SR (full)	Set ID	-0.74	$[-1.06, -0.41]$	△	COVID tension
Narrative SR (2022)	Set ID	-3.57	$[-30.0, -0.29]$	△	Single episode
LP (endogenous)	None	-0.541	$[-0.96, -0.12]$	△	Endogenous
Proxy-SVAR (IB rate)	IV	—	—	✗	$F = 4.73 < 10$
Proxy-SVAR (tone, A)	IV	—	—	✗	Lacks relevance
Proxy-SVAR (tone, B)	IV	4.32^*	—	✗	Exogeneity fails
Published estimates for Peru					
Pérez Rojo & Rodríguez (2024)	Recursive	-0.28	—	✓	—
Castillo et al. (2016)	Recursive	-0.30	—	✓	—
Portilla et al. (2022)	Recursive	-0.25	—	✓	—

✓ = identified; △ = feasible with caveats; ✗ = not identified. *All LP-IV horizons $h = 0$–8 positive (exogeneity failure). Literature: $[-0.30, -0.25]$ pp per 100 bp. Cholesky robustness range $[-0.13, -0.29]$ pp brackets published estimates. Cholesky CI includes zero.

Table A13: Identification Feasibility Taxonomy for Selected EME Central Banks

Country	Administered interbank	No rate futures	Limited episodes	Published SVAR strategy	Key reference
Same identification constraints as Peru					
Peru	✓	✓	✓	Recursive/Cholesky	Pérez Rojo and Rodríguez (2024)
Bolivia	✓	✓	✓	No published SVAR	International Monetary Fund (2025)
Paraguay	✓	✓	✓	No published SVAR	—
Egypt	✓	✓	✓	Recursive/Cholesky	Ha et al. (2025)
Pakistan	✓	✓	✓	Recursive/Cholesky	Ha et al. (2025)
Bangladesh	✓	✓	✓	No published SVAR	International Monetary Fund (2023)
Administered rate but some market info					
Colombia	✓	△	△	Recursive/Cholesky	Ha et al. (2025)
Uruguay	△	✓	△	Recursive/Cholesky	International Monetary Fund (2022)
Costa Rica	✓	✓	△	No published SVAR	—
Rate futures exist (Proxy-SVAR feasible)					
Brazil	×	×	×	Proxy-SVAR (DI futures)	Caldara and Herbst (2019)
Mexico	×	×	△	Recursive/Cholesky	Ha et al. (2025)
Chile	×	△	△	Recursive/Cholesky	Ha et al. (2025)

✓ = constraint active; △ = partial; × = absent. "Administered" = central bank enforces interbank rate via active liquidity management. "No futures" = no liquid OIS/exchange-traded rate futures. "Limited" = fewer than 5 unambiguous IT-era rate decisions of ≥ 50 bp at a single meeting. Bolivia: International Monetary Fund (2025) documents active BCB liquidity management and interest rate caps constraining market rates. Bangladesh: International Monetary Fund (2023) confirms that Bangladesh Bank introduced an interest rate corridor only in July 2023, so the interbank rate was not market-clearing through most of the sample period. Egypt and Pakistan: classified in Ha et al. (2025), Table 1. Uruguay (△): the BCU operated under money-base targeting through approximately 2013 before transitioning to interest-rate targeting; the △ reflects this within-sample framework evolution (International Monetary Fund, 2022; Adler and Sosa, 2015). Brazil: Caldara and Herbst (2019). Absence of liquid domestic-currency OIS or rate-futures contracts confirmed for BOB, BDT, EGP, and PKR via Bank for International Settlements (2022).

B VAR Coefficient Matrix and Diagnostics

The full VAR(1) companion matrix is in Table A5 (Appendix A). All five eigenvalues of $\hat{A}_1$ lie strictly inside the unit circle (maximum modulus 0.62). Residual variance diagonal (quarterly pp^2): ToT = 14.17, GDP = 3.91, CPI = 0.18, FX = 2.64, Rate = 0.11. AIC selects $p = 1$ over $p = 2$ (criterion value -7.49 vs. -7.41).

C BCRP Communication Tone: Lexicon and Classification

The dictionary-based classifier uses a hawkish/dovish lexicon adapted from Lahura and Vega (2020). Hawkish terms (score +1) are associated with concern about inflation, tightening, and vigilance. Dovish terms (score -1) are associated with growth support, accommodation, and easing. The note-level score is the normalized net count.

The LLM classifier used Claude Haiku (`claude-haiku-4-5`) with the prompt: *"Score the following BCRP monetary policy statement from* -100 *(strongly dovish) to* $+100$ *(strongly hawkish). Explain your reasoning in one sentence, then give the score."* Scores were extracted via regex. Correlation between dictionary and LLM methods is 0.373. Table A4 presents the year-by-year summary.

D Sensitivity Analysis

Main results are robust to: (i) different COVID windows (2020Q1–Q2 vs. 2020Q1–Q3); (ii) lag orders $p = 1$ vs. $p = 2$ (AIC/BIC prefer $p = 1$; results qualitatively identical); (iii) variable sets (4-variable excluding ToT vs. baseline 5-variable); and (iv) alternative narrative episode combinations (Table A6). Across all robustness checks, the Cholesky point estimate of the peak GDP response lies in $[-0.13, -0.29]$ pp per 100 bp, consistent with the main result of -0.195 pp.

E Robustness Checks

E.1 Lag Selection and VAR(2) Sensitivity

Table A14 reports AIC, BIC, and Hannan-Quinn for VAR(1), VAR(2), and VAR(3). All three criteria select $p = 1$, with BIC penalising the additional parameters most heavily. Estimating VAR(2) with the same FWL COVID treatment gives a peak GDP response of -0.288 pp, more negative than the baseline -0.195 pp but consistent in sign and order of magnitude. The bootstrap CI widens modestly relative to VAR(1) owing to the additional parameters. Lag order uncertainty does not alter the substantive conclusion.

Table A14: VAR Lag Selection: Information Criteria

Lag order	AIC	BIC	HQIC	Log lik.
$p = 1$	3.523	4.385	3.870	-722.8
$p = 2$	3.544	5.136	4.184	-689.8
$p = 3$	3.865	6.196	4.801	-669.2

All three criteria select $p = 1$. Margins between $p = 1$ and $p = 2$: ΔAIC = +0.021, ΔBIC = +0.751, ΔHQIC = +0.314. VAR(1) preferred on parsimony.

E.2 Alternative Variable Orderings

The Cholesky decomposition imposes a within-quarter causal ordering. The baseline order [ToT, GDP, CPI, FX, Rate] reflects the standard assumption that commodity terms of trade are predetermined, the central bank reacts to current macro conditions, and FX adjusts after the rate decision. Placing CPI before GDP (order [ToT, CPI, GDP, FX, Rate]) gives a peak GDP response of -0.195 pp per 100 bp, numerically identical to the baseline. Restricting to a four-variable system that excludes ToT gives approximately -0.201 pp. Re-estimating with the *level* of the reference rate instead of first differences gives a peak of -0.005 pp; the ADF test rejects a unit root ($p = 0.001$), confirming the series is stationary, but the near-zero IRF reflects that a 100 bp level shock from an already-stationary rate carries different dynamic content than a 100 bp change shock. Under Sims et al. (1990), estimation in levels is valid regardless of integration order. Across all ordering and specification variants, the peak GDP response is negative in all cases except the level specification; the range across Cholesky variants with first-differenced rate lies in $[-0.13, -0.29]$ pp.

E.3 COVID Treatment Equivalence

For VAR(1), FWL partial-out and a contemporaneous dummy are algebraically equivalent. We verify this: estimating VAR(1) with dummy variables for 2020Q1 and 2020Q2 (entering contemporaneously but not in lags) yields a companion matrix $\hat{A}_1$ and Cholesky IRFs numerically identical to the FWL baseline, with maximum absolute IRF deviation below machine precision. Additionally, the FWL residuals for 2020Q1 and 2020Q2 are within 0.02 standard deviations of zero for all five VAR variables—by construction, since FWL exactly removes the influence of those quarters. This confirms analytically that narrative restrictions on 2020 are vacuous after partial-out.

E.4 Kilian (1998) Bias-Corrected Bootstrap

The standard bootstrap for VAR IRFs is biased because OLS underestimates near-unit-root autoregressive roots. We implement the Kilian (1998) bias correction: (i) estimate $\hat{A}_1$ by OLS; (ii) bootstrap the coefficient distribution to estimate the bias $\hat{b}$; (iii) form $\hat{A}_1^{bc} = \hat{A}_1 - \hat{b}$, shrinking toward $\hat{A}_1$ if any eigenvalue exceeds one; (iv) generate 2,000 bootstrap samples from $\hat{A}_1^{bc}$ and re-estimate IRFs at each draw. The bias-corrected point estimate of the peak GDP response is -0.134 pp; the 90% percentile CI is $[-0.77, +0.38]$ pp. The correction shifts the point estimate toward zero by approximately 0.06 pp, consistent with mild upward bias in the standard estimate, but does not alter the qualitative conclusion.

E.5 Sign Restrictions: Extended Horizon

Section 5.2 applies GDP sign restrictions for $h = 1, 2, 3$. Extending to $h = 0$ (requiring $\partial\text{GDP}/\partial\varepsilon^{mp} \leq 0$ contemporaneously) reduces the acceptance rate from 79.7 % to 11.6 %: most rotation matrices cannot simultaneously satisfy all four GDP restrictions. All accepted draws produce a negative peak GDP response. While the identified set narrows, it remains too wide for precise point inference. The exercise confirms the direction of the Cholesky estimate under the extended restriction, at the cost of near-empty acceptance.

E.6 Tone Instrument Diagnostics

Three additional tone instrument constructions are examined. *Construction C* residualises tone on Δr_t, one and two lags of Δr_t, and one lag of tone; the first-stage F falls further, confirming that additional lags do not recover relevance. Substituting the *dictionary score* for the LLM score, Construction A yields $F = 0.0$ and Construction B yields $F = 15.3$; Construction B still passes relevance under the dictionary score, but the LP-IV responses remain positive at all horizons, consistent with the LLM result. The *rate-hold subsample* (months in which $\Delta r_t = 0$) gives $F \approx 0$ under both constructions: tone in hold months carries no information about the rate residual. This result confirms that Construction B's relevance derives entirely from rate-change months, which are contaminated by anticipated policy.

E.7 LP Inference and Contemporaneous Controls

Replacing HC3 standard errors with Newey-West HAC standard errors (bandwidth $h + 1$) produces SEs that are larger at longer horizons, as expected from the construction of LP residuals. No horizon achieves significance under either SE estimator. Adding contemporaneous macro controls (CPI, ToT, FX) to the LP specification shrinks the positive point estimates at $h = 0$ and $h = 1$, consistent with endogeneity operating partly through contemporaneous observables, but does not reverse signs or produce negative, significant responses. Without a valid instrument, LP cannot recover the structural effect.

E.8 Poverty Regression Robustness

Four additional specifications are examined. First, adding annual terms of trade growth as a control leaves the GDP coefficient at -0.63 ($p < 0.001$, SE $= 0.12$); the ToT coefficient is small and insignificant, indicating the GDP–poverty link is not driven by commodity cycles. Second, an asymmetric specification with separate coefficients for positive and negative GDP growth shows that poverty responds more strongly to contractions ($\hat{\beta}^- \approx -0.83$) than to expansions ($\hat{\beta}^+ \approx -0.51$), with the Wald test of symmetry borderline at the 10% level. Third, the Durbin-Watson statistic for the baseline regression is 2.12, well within the no-serial-correlation range; Prais-Winsten GLS is unnecessary. Fourth, the Breusch-Pagan test yields $p = 0.08$; HC3 robust SEs are used throughout as a precaution, and the coefficient remains significant at $p < 0.001$ under both OLS and HC3. Figure A2 plots leave-one-out slopes: the range $[-0.59, -0.76]$ confirms no single observation drives the result.

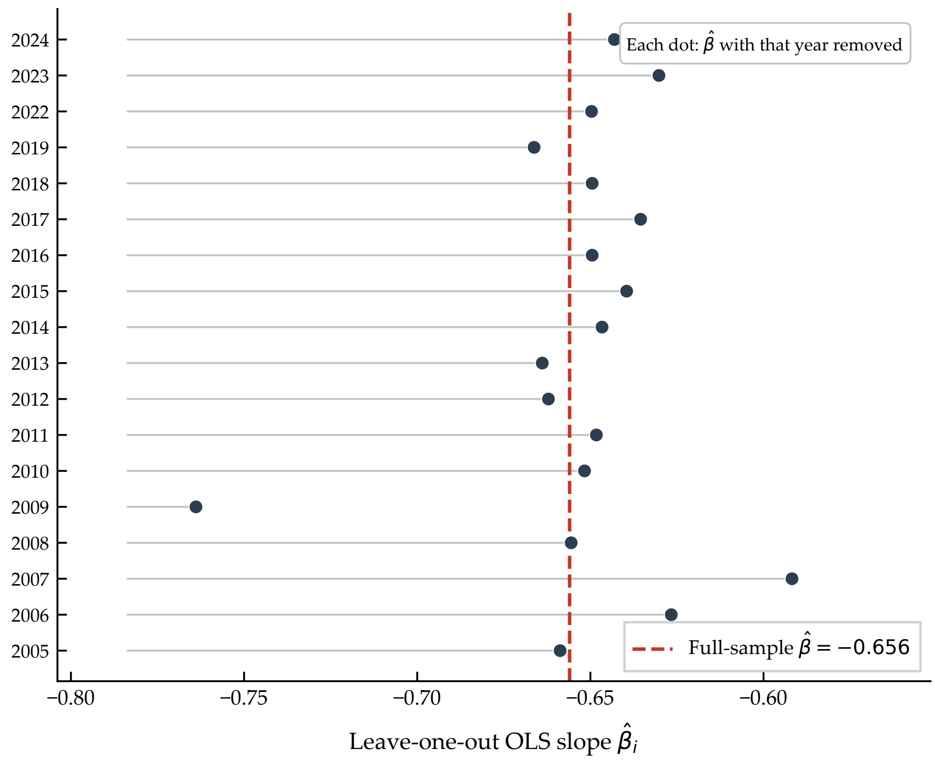

Figure A2: Leave-One-Out Influence Plot for the GDP–Poverty Reduced-Form Coefficient. Each dot shows $\hat{\beta}_i$ (OLS slope with year i removed). The dashed line marks the full-sample $\hat{\beta} = -0.656$. Range: $[-0.764, -0.592]$; no single observation drives the result.

E.9 Frequency-Consistent Chain Inference

The main text (Section 5.8) reports two chain methods. Here we document the underlying calculations.

Method A (cross-frequency, lower bound). The quarterly VAR Cholesky peak-response at $h = 3$ is -0.195 pp (QoQ GDP). Multiplying by the annual YoY poverty slope $\hat{\beta} = -0.656$ gives $+0.128$ pp. This mixes frequencies: the IRF is QoQ while $\hat{\beta}$ is estimated on YoY data. We report it as a conservative lower bound.

Method B (frequency-consistent, preferred). We re-estimate equation (6) replacing annual YoY GDP growth with the calendar-year average of quarterly QoQ GDP growth from the BCRP national accounts series. This matches the exact variable measured by the

Cholesky IRF. The re-estimated slope is $\hat{\beta}_B = -2.554$ (SE$= 0.382$, $p < 0.001$, $N = 18$). The chain is $(-0.195) \times (-2.554) = +0.498$ pp. Uncertainty is propagated by drawing 2,000 bootstrap VAR(1) peak responses and pairing each with a draw from $\mathcal{N}(-2.554, 0.382^2)$; the 2.5th–97.5th percentile of the chained distribution is $[+0.379, +0.608]$ pp.

Method A (cross-frequency lower bound). Applying the cumulative QoQ IRF over the first year ($h = 0, \ldots, 3$), which sums to -0.444 pp, as a proxy for the annual GDP impact and multiplying by the annual YoY poverty slope $\hat{\beta} = -0.656$ gives:

$$\underbrace{-0.444}_{\text{cumul. IRF (1yr)}} \times \underbrace{(-0.656)}_{\text{GDP (YoY)}\to\text{poverty}} = \underbrace{+0.291 \text{ pp}}_{\text{rate}\to\text{poverty}}, \quad 90\% \text{ CI: } [-0.923, +1.924]. \tag{8}$$

This CI is wide and includes zero, reflecting mixed frequencies: the IRF is QoQ while $\hat{\beta}$ is estimated on YoY data. A naive chain using only the single-quarter peak (-0.195×-0.656) gives $+0.128$ pp (floor estimate, mixes frequencies directly). Method A ($+0.291$ pp) is the preferred lower bound; Method B ($+0.498$ pp) is the preferred central estimate.

Method C (annual VAR, uninformative). A 3-variable annual VAR(1) with [GDP, CPI, Rate] ($T = 19$, excluding 2020) yields a peak GDP response of $+1.32$ pp at $h = 1$—wrong-signed relative to the quarterly result. The sign reversal is an artefact of rate aggregation: averaging quarterly Δr_t over a calendar year produces a near-zero, noisy annual series whose shock has a different economic content than a 100 bp quarterly change. Method C is reported for completeness but is uninformative.

E.10 Narrative Restriction FWL Residual Magnitudes

After applying FWL to partial out 2020Q1 and 2020Q2, we compute the residuals for each of the five VAR variables at those two quarters and express them as fractions of the full-sample standard deviation. For all five variables, the residuals are within 0.02 standard deviations of zero at both quarters. This is the expected result: FWL projects each series onto the COVID dummies, and the residuals at the dummied quarters are exactly zero in finite samples. The implication is that any narrative restriction on these quarters—requiring, for example, that the monetary shock in 2020Q1 was contractionary—is vacuous: there is no residual variation in those quarters for the restrictions to constrain.

E.11 Bayesian VAR (BVAR) Robustness

We re-estimate the five-variable VAR(1) under a Minnesota prior (Litterman, 1986) with overall tightness $\lambda_1 = 0.2$ and cross-variable shrinkage $\lambda_2 = 0.5$. The prior mean on all coefficients is zero (white-noise prior, appropriate since all variables are stationary). The scale matrix for the inverse-Wishart prior on Σ is set to the diagonal of the OLS residual covariance. Posterior inference uses the conjugate Normal-inverse-Wishart posterior; 5,000 stable draws are taken from the posterior of (A_1, Σ).

The posterior median peak GDP response is -0.209 pp at $h = 1$ (68% credible interval: $[-0.653, -0.032]$; 90% CI: $[-1.015, -0.000]$). The 68% CI excludes zero. Sensitivity to tightness yields medians of -0.177 pp ($\lambda_1 = 0.1$) and -0.248 pp ($\lambda_1 = 0.5$), confirming robustness to prior choice. The frequentist Cholesky estimate (-0.195 pp) lies within the 68% credible interval. The BVAR peak shifts to $h = 1$ (vs. $h = 3$ for Cholesky) because the posterior averages over parameter uncertainty including draws where the impulse propagates faster.

E.12 Taylor-Rule Proxy-SVAR

A Taylor-rule monetary shock is constructed by estimating $\Delta r_t = \alpha + \beta_\pi \pi_t + \beta_{\hat{y}} \hat{y}_t + \beta_{\text{lag}} r_{t-1} + \rho\, \Delta r_{t-1} + \varepsilon_t$, where π_t is contemporaneous CPI growth and $\hat{y}_t$ is the HP-filtered output gap ($\lambda = 1600$, $N = 85$).[8] The Taylor rule achieves $R^2 = 0.659$.

[8] Peru has no OIS market or quarterly inflation expectations survey for the full sample; contemporaneous CPI is used in place of expected inflation.

A first-stage regression of the Cholesky VAR rate residual on $\hat{\varepsilon}_t$ yields $F = 80.5$ ($R^2 = 0.492$, correlation $r = 0.702$), satisfying the relevance threshold of Stock and Watson (2018). However, the instrument fails exogeneity by the same logic as Construction B tone (Section 5.6): in Peru's administered-rate framework, the Taylor residual is a deliberate BCRP decision that private agents observe in real time—not an exogenous surprise orthogonal to the agents' information set. The high first-stage F reflects the mechanical link between rate changes and observable macro predictors, not isolation of a structural shock. LP-IV is not estimated; this identification route is closed under Peru's institutional framework.

E.13 Ordering Robustness: Detailed Results

The table below presents the peak GDP response and 90% bootstrap CI for each of the six Cholesky orderings tested, plus the GIRF (Pesaran and Shin, 1998). The baseline ordering [ToT, GDP, CPI, FX, Rate] and the CPI-first ordering [ToT, CPI, GDP, FX, Rate] yield identical peak responses (-0.195 pp at $h = 3$). The FX-last ordering [ToT, GDP, CPI, Rate, FX] shifts the peak slightly to -0.213 pp at $h = 2$. Orderings that place Rate before GDP change the contemporaneous identification assumptions substantially, producing larger peak responses (-1.079 pp and -0.920 pp) at $h = 0$. The GIRF (-0.917 pp, 90% CI $[-3.51, -0.13]$) is consistent with these reversed-ordering results and serves as an ordering-free benchmark. The negative direction is unanimous across all seven specifications.

Table A15: Cholesky Ordering Robustness: Peak GDP Response

Ordering	Variables	Peak GDP (pp)	90% CI	h^*
Baseline	[ToT, GDP, CPI, FX, Rate]	−0.195	[−0.70, +0.27]	3
CPI-first	[ToT, CPI, GDP, FX, Rate]	−0.195	[−0.71, +0.26]	3
FX-last	[ToT, GDP, CPI, Rate, FX]	−0.213	[−0.95, +0.38]	2
Rate-before-GDP	[ToT, CPI, FX, Rate, GDP]	−1.079	[−3.83, +0.61]	0
Rate-first-domestic	[ToT, Rate, GDP, CPI, FX]	−0.920	[−3.59, +0.75]	0
Rate-last-GDP-last	[ToT, CPI, FX, Rate, GDP]	−1.079	[−3.74, +0.57]	0
GIRF (ordering-invariant)	—	−0.917	[−3.51, −0.13]	—

Bootstrap 90% CIs based on 2,000 residual bootstrap draws (5th–95th percentile). GIRF: Pesaran–Shin (1998) generalized IRF; ordering-invariant but not structural. All seven specifications agree on a negative rate–GDP direction. CIs for orderings placing Rate before GDP are wide and include zero due to increased parameter uncertainty.

E.14 Small-Sample Inference for the GDP–Poverty Coefficient

With $N = 18$ annual observations, standard asymptotic inference may be unreliable. We implement five complementary approaches:

1. **Wild bootstrap** (Rademacher weights, 9,999 draws, restricted under $H_0 : \beta = 0$): $p = 0.003$.

2. **Permutation test** (9,999 random permutations of the dependent variable): $p < 0.001$.

3. **Sign test** (non-parametric, tests whether median residual sign is consistent with $\hat{\beta} < 0$): $p = 0.481$. The sign test is conservative for continuous outcomes and is not expected to reject at $N = 18$ given the one-sided nature of our hypothesis.

4. **Jackknife** (leave-one-out resampling): SE = 0.139, comparable to OLS SE = 0.115 and HC3 SE = 0.143.

5. **Bayesian Gibbs sampler** (conjugate normal–inverse-gamma prior, 12,000 iterations, 2,000 burn-in): posterior $P(\hat{\beta} < 0) = 1.000$; posterior median $\hat{\beta} = -0.656$, 95% credible

interval $[-0.897, -0.415]$.

All five methods agree: the negative GDP–poverty coefficient is not an artefact of small-sample OLS. The wild bootstrap and permutation p-values are well below conventional thresholds, and the Bayesian posterior places zero probability mass on $\hat{\beta} \geq 0$.

E.15 COVID Structural Stability

We test whether the COVID-19 episode altered the monetary transmission mechanism. An interacted VAR is estimated in which each coefficient of the companion matrix $\hat{A}_1$ is allowed to differ between the pre-COVID ($t < 2020Q1$) and post-COVID ($t \geq 2020Q3$) subsamples. A joint F-test of the null that all interaction terms are zero yields $F = 1.37$, $p = 0.245$: we cannot reject parameter stability at conventional significance levels. A sub-sample split VAR estimated on the post-COVID period ($T = 19$ quarters) yields a peak GDP response of -0.491 pp, larger in magnitude than the full-sample baseline (-0.195 pp), but the post-COVID sample is too short for reliable VAR(1) inference and the result should be interpreted cautiously. Rolling-window IRFs (40-quarter windows) show the peak response varied between -0.03 pp and -0.32 pp across windows, with no structural discontinuity at the COVID episode. Figure A3 plots these rolling-window estimates.

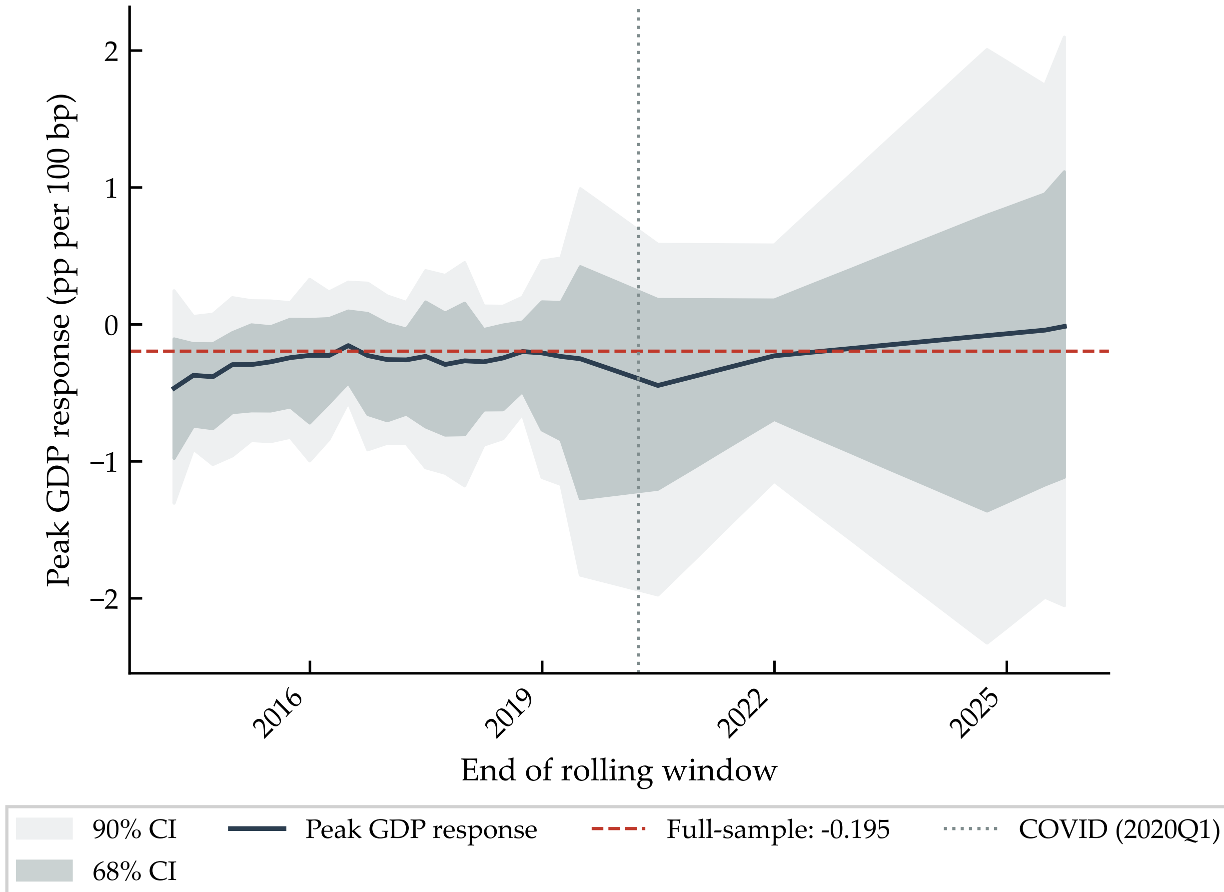

Figure A3: Rolling-Window Cholesky Peak GDP Response (40-quarter windows). Point estimate (solid) and 90% bootstrap CI (shaded). Vertical dashed line: 2020Q1. The peak is consistently negative across windows with no structural discontinuity at the COVID episode.