A Work Project presented as part of the requirements for the Award of a Master's degree in Economics from the Nova School of Business and Economics

U.S. Government Fiscal Multipliers Across Monetary Policy Regimes

Evidence from Forecast-Error and Narrative Defence-News Shocks

José Tomás Pinto

Work project carried out under the supervision of:

Paulo M. M. Rodrigues

01-2026

Abstract

This work project estimates U.S. government spending multipliers using local projections with external instruments (LP–IV), smooth local projections (SLP–IV), and a time-varying-parameter extension (TVP–SLP–IV). A constant-parameter SLP–IV baseline on 1985Q1–2019Q4 yields modest cumulative dollar multipliers, often below one. A single TVP–SLP–IV model on 1985Q1–2023Q4 delivers regime-averaged impulse responses for five monetary-policy environments; multipliers for the three long pre-Covid regimes cluster around unity, while Covid and post-Covid multipliers are treated as not identified. Identification uses Auerbach–Gorodnichenko forecast-error shocks; Ramey defence-news shocks provide overlap robustness. Inflation and inflation expectations respond mildly.

Keywords: Fiscal Multipliers; Zero Lower Bound; Monetary Policy Regimes; Smooth Local Projections; Time-Varying Parameters; External Instruments; Narrative Shocks; Defence News Shocks; Forecast Error Shocks.

Acknowledgements: I would like to express my deepest gratitude to professor Paulo Rodrigues for his support and guidance throughout this journey. I also thank my family for providing me with this opportunity, along with my friends, who were crucial in helping me complete this work. Above all, I acknowledge my Lord and Savior, Jesus Christ—without Him, I would not be here.

This work was funded by Fundação para a Ciência e a Tecnologia (UID/00124/2025, UID/PRR/124/2025, Nova School of Business and Economics) and LISBOA2030 (DataLab2030 - LISBOA2030-FEDER-01314200).

1 Introduction

Since the emergence of Keynesian economics, fiscal stimulus has been recognised as a central instrument for mitigating major downturns. In the United States (U.S.), the Great Recession (2008–2009) and the Covid-19 pandemic underscored the importance of fiscal policy when conventional monetary tools were constrained. Because policy repeatedly leans on fiscal expansions at such times, reassessing their effectiveness (specifically, the size of the government-spending multiplier at the zero lower bound (ZLB)) is crucial. Those conditions prevailed in 2008Q4–2015Q3 and again in 2020Q2–2021Q4, when stabilising output, employment, and inflation depended more heavily on fiscal measures.

The empirical analysis uses quarterly U.S. data from 1985Q1–2023Q4. A constant-parameter SLP–IV benchmark is estimated on 1985Q1–2019Q4, and time variation is modelled using a single TVP–SLP–IV specification on 1985Q1–2023Q4. Impulse responses are reported as regime averages for the five monetary-policy environments defined below (with 2020Q1 and 2022Q1 excluded from regime averaging but retained in the TVP estimation). Core comparisons emphasise the three long pre-Covid regimes; in the short Covid and post-Covid windows, weak identification of cumulative government spending at long horizons (especially $H = 16$) can render ratios unstable, and such regime-multiplier entries are treated as not identified.

From a macroeconomic perspective, the 1985Q1 start date aligns the baseline window with the Great Moderation and the Volcker–Greenspan monetary-policy era, during which U.S. monetary policy is widely characterised as more forward-looking and more responsive to inflation than in earlier decades, consistent with a stronger commitment to price stability (Goodfriend and King 2005; Clarida, Galí, and Gertler 2000; Stock and Watson 2003). The Pre–ZLB (1985Q1–2008Q3) defined below therefore corresponds to a mature, price-stability-oriented regime with an active, rule-like monetary policy, providing a natural benchmark against which the zero lower bound and post-Covid regimes can be compared.

The government spending multiplier (the change in output generated by an additional unit of government expenditure) remains contested. In this work project, the constant-parameter SLP–IV benchmark reports cumulative dollar multipliers (ratios of cumulative per-capita level responses), whereas the TVP–SLP–IV regime summaries report unit-free regime multipliers defined as ratios of cumulative log responses and are therefore not directly interpretable in dollars. Throughout, output is denoted Y (real GDP per capita) and government spending G (real government purchases per capita).

Using a structural vector autoregression (SVAR), Blanchard and Perotti (2002) report peak government purchases multipliers between 0.9 and 1.29. Using a state-dependent empirical framework, Auerbach and Gorodnichenko (2013) find strongly state-dependent peak multipliers of -0.3–0.8 during expansions and 1.0–3.6 during recessions. Using narrative defence-news shocks, Ramey (2011a) reports cumulative multipliers of 0.6–1.2 (ranges compiled in Ramey (2011a, Table 1)). Definitions also vary (peak versus cumulative multipliers; government purchases; identification via timing, news, or forecast errors). Importantly, a multiplier below one implies that total output rises by less than one-for-one with government spending. Whether private activity is crowded out is an empirical question about the composition of responses. These discrepancies motivate analysis of state dependence across monetary regimes under a consistent external-instrument design. Accordingly, the empirical strategy uses Auerbach–Gorodnichenko (AG) forecast-error shocks as the baseline instrument and Ramey's defence-news shocks on the overlap (see Section 3), and addresses precision and time variation by estimating local projections with external instruments (LP–IV) as an unsmoothed benchmark, smoothing responses across horizons via smooth local projections (SLP), and allowing coefficients to evolve over time in a time-varying Kalman-filter specification (see Section 4). To keep cross-regime comparisons focused on state dependence rather than changing identification, identification is held fixed over time and relies exclusively on external instruments: an AG-type forecast-error instrument constructed following Auerbach and Gorodnichenko (2012) and extended through 2023Q4 as the baseline instrument in both the 1985Q1–2019Q4 SLP–IV benchmark and the 1985Q1–2023Q4 TVP–SLP–IV layer, and Ramey defence-spending news shocks as a robustness check on their 1985Q1–2015Q4 overlap (Ramey 2016), with a Both–IV specification that instruments government spending growth with both series and uses Hansen's J-test (Hansen 1982) of overidentifying restrictions (reported at impact).

A key source of heterogeneity is state dependence across the business cycle and monetary regimes. A large literature studies whether multipliers are larger in recessions and when short-term rates are constrained at or near the ZLB; the empirical evidence is mixed (e.g., Ramey and Zubairy 2018). New Keynesian models rationalise this pattern: with an active Taylor rule away from the ZLB, fiscal expansions are partly offset by higher nominal rates, so that standard calibrations deliver modest government-spending multipliers, often below one (Woodford 2011); when the policy rate is constrained at the ZLB, increases in government spending raise expected inflation and the output gap, lowering real interest rates, which can crowd in private demand and can generate multipliers above one (Christiano, Eichenbaum, and Rebelo 2011; Eggertsson 2011). The U.S. experienced two distinct ZLB windows, the first coinciding with the Global Financial Crisis (GFC, 2008Q4–2015Q3) and the second during the Covid-19 pandemic (2020Q2–2021Q4), separated by positive-rate interludes and the latter followed by a post-Covid tightening. To accommodate this evolution without re-estimating very short subsamples, a single TVP–SLP–IV model is estimated over 1985Q1–2023Q4 and five monetary-policy environments are summarised: Pre–ZLB (1985Q1–2008Q3), ZLB–GFC (2008Q4–2015Q3), Positive-rate (Pre-Covid) (2015Q4–2019Q4), ZLB–Covid (2020Q2–2021Q4), and Positive-rate (Post-Covid) (2022Q2–2023Q4). The empirical analysis focuses on regime-averaged responses and on regime-multiplier comparisons for the three long pre-Covid regimes, where time spans are sufficient for reliable identification. The Covid and post-Covid windows enter the TVP estimation for completeness, but where the cumulative response of government spending is too small to support stable ratios at the multiplier horizons (especially at $H = 16$), the corresponding regime-multiplier entries are treated as not identified and are not used as core quantitative evidence.

Following Jordà (2005), LP–IV provides a flexible horizon-by-horizon framework to trace the response of output to government spending. However, because plain LPs can be noisy at longer horizons, the analysis uses SLP to gently pool information across nearby horizons, tightening uncertainty bands without imposing a vector autoregression (VAR) structure (Barnichon and Brownlees 2019). LP–IV is reported only as an unsmoothed benchmark; the baseline constant-parameter estimates and headline multipliers are based on SLP–IV. To allow effects to evolve with the economy, the SLP is embedded in a time-varying-parameter (TVP) framework estimated via a Kalman filter, which allows responses to drift gradually over time. Identification remains external throughout, and inference is weak-IV-aware (first-stage strength and an over-identification check on the overlap).

The contribution is to combine narrative external instruments, SLP precision, and time variation within a two-layer framework: a constant-parameter SLP–IV benchmark on 1985Q1–2019Q4 and a TVP–SLP–IV layer on 1985Q1–2023Q4 for regime-dependent dynamics. On the data side, quarterly real per-capita series are constructed as described in Section 3, and regime comparisons focus on the three long pre-Covid monetary regimes defined above.

Previewing the results, the constant-parameter SLP–IV estimates imply modest medium-run cumulative dollar multipliers, with point estimates below one and fairly wide confidence intervals. In the TVP–SLP–IV model, regime multipliers cluster around unity in the three long pre-Covid regimes, and the 2008Q4–2015Q3 episode yields the clearest evidence of a positive long-horizon multiplier with relatively tight bands. Consumption responses are typically positive, while investment responses are smaller and less precisely estimated; inflation and inflation expectations move only modestly.

Section 2 reviews the related literature. Section 3 describes data and instruments (series, transformations, ZLB flag, and construction of the Auerbach–Gorodnichenko and Ramey instruments; see Appendix E for details). Section 4 details the estimation approach (LP–IV, SLP, and the TVP–SLP framework) and inference, along with diagnostics for first-stage strength and over-identification. Section 5 presents the baseline impulse responses and cumulative dollar multipliers and documents prices, expectations, and mechanisms. Section 6 reports robustness checks (alternative instruments, horizon and Covid-exclusion diagnostics), with detailed first-stage and overlap results in Appendix B.

2 Literature Review

Empirical work on fiscal multipliers is extensive and disagrees mainly over timing and identification. Blanchard and Perotti (2002) is a classic starting point. Using a quarterly U.S. structural vector autoregression (SVAR) in which fiscal variables are treated as predetermined within the quarter (in the sense that discretionary policy does not respond within the quarter) and government purchases are assumed not to respond contemporaneously to output, they find that surprise increases in government spending raise output with peak multipliers around 0.9 to 1.29 across specifications. The composition is distinctive: private consumption rises after a spending shock, whereas private investment falls after both tax and spending shocks, consistent with crowding out of private investment. The results highlight a key fragility: conclusions depend on what is treated as predetermined within the quarter.

A central identification debate concerns timing versus anticipation. The national accounts record spending when outlays occur, but agents may adjust when news arrives rather than when cash is spent. Households and firms often learn about policy changes before the national accounts data register the spending. In response, a complementary identification strand introduces shock series that align econometric timing with the arrival of fiscal information, including narrative or news measures and forecast errors such as professional forecast errors.

Ramey (2011b) was among the first to press this timing critique using narrative evidence. She documents that many large defence spending episodes are anticipated well in advance and that private behaviour adjusts when news arrives, not when outlays appear in the National Income and Product Accounts (NIPA). Constructing a news-variable measure of defence announcements that quantifies the expected path of future spending at the time of announcement, she finds that the private response differs markedly from SVAR estimates that treat spending as unanticipated. In particular, the classic Blanchard–Perotti consumption increase largely disappears (most components of consumption fall after defence news shocks). This contributes to lower multipliers in news based measures (Ramey 2011a); Ramey (2016) documents and extends the defence news shock series.

The forecast error literature emphasises another approach to anticipation. Auerbach and Gorodnichenko (2013) allow the multiplier to differ over the business cycle and identify government spending shocks using professional forecast errors for government spending growth. The idea is to compare realised government spending growth with what professional forecasters expected and treat the unexpected component (after purging predictable variation using lags of output and spending) as the shock, which is then used to trace the impulse response. Their results suggest state dependent multipliers that are much larger in recessions than in expansions. More generally, Ramey (2019) emphasises that identification choices and implementation decisions (including shock construction and multiplier calculation) can materially affect estimated multiplier magnitudes.

A parallel literature asks whether what governments spend on matters for the size of the multiplier. Different spending types (for example, purchases rather than transfers, or military procurement rather than other categories) can be associated with different private responses. Heterogeneous responses of consumption, investment, and labour supply suggest that aggregation of categories with different elasticities can blur the multiplier. These composition and interpretation issues are salient when comparing defence-based identification strategies to broader spending aggregates. Nakamura and Steinsson (2014) emphasise that their estimates based on regional military procurement identify an "open economy relative multiplier" that differs conceptually from the standard closed economy aggregate multiplier, and they develop a framework to relate the two, so treating defence-based evidence as a direct proxy for generic government purchases multipliers requires caution.

Local projections (LP) have become a standard way to estimate impulse responses without imposing the dynamic restrictions required by vector autoregressions (VARs), and local projections with external instruments (LP–IV) combine this flexibility with external instruments that address anticipation and timing concerns. However, horizon-by-horizon estimates can be noisy at medium horizons and produce jagged response profiles with wide uncertainty bands. Smooth local projections (SLP) address this by penalising curvature across horizons (a second-difference penalty), shrinking adjacent horizons toward a smooth curve and improving precision while leaving identification and regressors unchanged. Embedding SLP in a time-varying-parameter (TVP) framework then allows gradual drift in coefficients while preserving horizon smoothing and compatibility with external-instrument identification, which is useful when the sample contains distinct monetary policy regimes and very short subsamples are unattractive for re-estimation. In the empirical implementation, SLP–IV is the constant-parameter baseline estimator, and LP–IV is reported only as an unsmoothed benchmark.

A final methodological point concerns how multipliers are constructed. Peak or impact multipliers can be sensitive to short-run noise, and definitions differ across studies. Cumulative multipliers smooth over early-horizon inventory and measurement noise and provide a compact summary of the total response over a given horizon. Reporting cumulative multipliers at several horizons (headline $H = 8$ and $H = 16$, with a shorter-horizon $H \leq 12$ diagnostic when weak identification prevents meaningful inference at longer horizons) alongside LP/SLP impulse responses helps reconcile differences across methods and facilitates comparisons across regimes without rescaling government spending or output ex post. In the constant-parameter baseline, cumulative multipliers for real-activity outcomes are reported as per-capita dollar ratios of cumulative level responses. In the TVP regime summaries, regime multipliers for the same outcomes are constructed as ratios of cumulative log responses, which are unit-free and typically close to the corresponding per-capita dollar multipliers. Regime multipliers are reported for all five monetary-policy regimes; however, in short windows where the cumulative response of government spending is too small to support stable response ratios at longer horizons (notably $H = 16$), the corresponding regime-multiplier entries are treated as not identified. For prices, expectations, and the unemployment rate, results are summarised as cumulative percentage-point responses to a government-spending shock normalised to raise G by 1 percent of GDP on impact, rather than as per-dollar multipliers. This mitigates sensitivity to deflator and scaling conventions that can affect the construction and interpretation of real spending and fiscal shocks (Ben Zeev, Ramey, and Zubairy 2023).

On the theory side, New Keynesian models provide guidance on how fiscal multipliers vary with the monetary regime. In models with an active Taylor rule (or a strict inflation target away from the ZLB), increases in government spending tend to be offset by higher nominal interest rates, limiting the size of the multiplier and keeping it below one in baseline New Keynesian calculations (Woodford 2011). When the policy rate is constrained at the ZLB, however, the nominal rate cannot rise in response to higher demand, and government spending increases can raise expected inflation and lower real interest rates, generating larger multipliers (Christiano, Eichenbaum, and Rebelo 2011; Eggertsson 2011; Woodford 2011). Subsequent work emphasises that realistic policy inertia (and the implied expectations about policy tightening after exit from the ZLB) tends to pull these multipliers toward more moderate values near one rather than the very large numbers obtained under more extreme assumptions (Hills and Nakata 2018). Empirical surveys such as Ramey (2019) suggest that most aggregate U.S. postwar government purchases multipliers in "normal times" are around one or below, while multipliers above one are most plausible under episodes of strong monetary accommodation (including ZLB environments).

Taken together, these strands motivate the empirical design used here (external instruments based on Auerbach and Gorodnichenko as the baseline, with Ramey used to assess overlap, precision via SLP, and a TVP layer with regime averaged reporting). The starting date and regime definition link the estimates to the post 1979 monetary policy regime shift and the subsequent Great Moderation environment emphasised by Clarida, Galí, and Gertler (2000) and Stock and Watson (2003), and to the Volcker disinflation episode analysed by Goodfriend and King (2005). The emphasis on ZLB versus positive rate regimes reflects the New Keynesian prediction that fiscal multipliers are state dependent and tend to be larger when monetary policy is constrained, as discussed above.

3 Empirical Design

This section describes the quarterly dataset, the external instruments used to identify innovations to government purchases, and the monetary-policy regime definition that organises state-dependent summaries. The underlying dataset spans 1985Q1–2023Q4. The constant-parameter benchmark is a single SLP–IV specification estimated on 1985Q1–2019Q4. State dependence is summarised using a single TVP–SLP–IV model estimated on 1985Q1–2023Q4; regime-specific objects are constructed by averaging time-varying impulse responses over pre-defined regime windows rather than by re-estimating separate regime subsamples. Full series provenance (identifiers, transformations, and retrieval dates) is reported in Appendix A. Instrument-construction details are reported in Appendix E.

3.1 Data, Sources, and Transformations

The dataset combines national accounts aggregates, labour-market indicators, policy rates, and survey expectations. Real activity is summarised by real GDP and real government purchases; Y_t and G_t denote real GDP per capita and real government purchases per capita. Mechanism variables include real personal consumption expenditures and real private domestic investment (denoted C_t and I_t in per-capita terms), together with labour-market and price/expectations measures used in the mechanisms analysis. Headline inflation is measured using the personal consumption expenditures (PCE) price index, and inflation expectations are measured using the University of Michigan (MICH) one-year-ahead median inflation expectations series.

All variables are aligned to quarterly frequency. Higher-frequency series are converted to quarterly averages prior to any differencing. Real aggregates enter in per-capita log levels constructed by dividing real aggregates by quarterly-average population and taking logs. The local-projection regressions use quarterly log differences, e.g. $\Delta\log(Y_t)$ and $\Delta\log(G_t)$. Equivalently, let $y_t \equiv \log(Y_t)$ and $g_t \equiv \log(G_t)$, so that $\Delta\log(Y_t) = \Delta y_t$ and $\Delta\log(G_t) = \Delta g_t$. Because growth rates are constructed by differencing and the baseline specification includes four lags of $\Delta\log(Y_t)$ and $\Delta\log(G_t)$, the effective estimation sample begins in 1986Q2 after the required lag structure and complete-case filtering are imposed; reported sample windows refer to the underlying quarterly data coverage prior to this trimming. Horizon-h equations additionally truncate the terminal base quarters because the dependent variable y_{t+h} must be observed.

Inflation from price indices is computed as $\pi_t = 400 \cdot \Delta\log(\text{index}_t)$ (annualised percent rates). The unemployment rate and one-year-ahead inflation expectations enter in percent levels. Complete series-by-series identifiers and transformations are deferred to Appendix A (Table 4).

3.2 External Instruments for Fiscal Shocks

Unexpected movements in government purchases are identified using external instruments for innovations to government spending growth. The baseline instrument is an AG-type Survey of Professional Forecasters (SPF) forecast-error shock defined as the one-quarter-ahead forecast error in log growth of real government purchases:

$$z_t^{AG} = \Delta\log(G_t) - \Delta\log\bigl(G_{t|t-1}^{SPF}\bigr),$$

where $\Delta\log\bigl(G_{t|t-1}^{SPF}\bigr)$ is the SPF forecast formed in quarter $t-1$ for quarter t and aligned to the realisation quarter. The AG-type series is constructed using a uniform methodology and extended through 2023Q4. Detailed alignment and construction choices are documented in Appendix E.

As a secondary instrument, the Ramey defence-news shock is used on its original support (1985Q1–2015Q4). In the main specification, AG-type shocks are the sole instrument so that the instrument set is constant over the analysis window. On the overlap sample 1985Q1–2015Q4, robustness specifications include AG-only, Ramey-only, and a Both–IV specification that uses both instruments jointly and supports over-identification checks via Hansen's J statistic. Instruments are not spliced mid-sample: the AG-type series is extended using a uniform methodology, while the Ramey series is confined to its original sample.

3.3 Monetary Policy Regimes

Monetary-policy regimes are summarised by a binary ZLB indicator ZLB_t based on the quarterly average of the federal funds target-range upper bound (DFEDTARU). Specifically, $ZLB_t = 1$ when the quarterly-average upper bound is $\leq 0.25\%$, and $ZLB_t = 0$ otherwise. To limit within-quarter mixing around entry into and exit from the Covid ZLB episode when forming regime averages, boundary quarters are treated as transition quarters and excluded from regime averaging (while retained in the underlying dataset). With this convention, the sample is partitioned into five regimes: Pre–ZLB (1985Q1–2008Q3), ZLB–GFC (2008Q4–2015Q3), Positive-rate (Pre-Covid) (2015Q4–2019Q4), ZLB–Covid (2020Q2–2021Q4), and Positive-rate (Post-Covid) (2022Q2–2023Q4). Exact window definitions and constructed indicators are reported in Appendix A.

Regime-specific impulse responses and mechanism responses are computed from the unified TVP–SLP–IV model, with regime averages formed by averaging time-varying impulse responses over the quarters belonging to each regime. Regimes therefore enter only through averaging windows, and the model is not re-estimated separately by regime. Because the Covid and post-Covid regimes are short, long-horizon objects (notably at $H = 16$) can be unstable when the cumulative response of government spending is too small to support stable response ratios; in such cases, long-horizon regime-multiplier entries are treated as not identified and interpreted cautiously. Supporting diagnostics are provided in Appendix B.

4 Estimation and Inference

This section describes the estimators used to recover impulse response functions (IRFs) and fiscal multipliers. All specifications share the same identification strategy (Section 3.2), control set, and series transformations (Section 3.1). Three complementary local-projection IV approaches are implemented: (i) unsmoothed LP–IV as a benchmark, (ii) SLP–IV as the constant-parameter baseline, and (iii) TVP–SLP–IV as a unified time-varying specification summarised by regime averages. Constant-parameter models are estimated on 1985Q1–2019Q4. The TVP–SLP–IV model is estimated on 1985Q1–2023Q4 and summarised by averages over the monetary-policy regimes in Section 3.3.

4.1 Local Projections with External Instruments (LP–IV)

For each transformed outcome y_t and horizon $h = 0, \ldots, H$ (baseline $H = 16$; robustness uses $H \leq 12$), the LP–IV benchmark is,

$$y_{t+h} = \beta_h \Delta \widehat{g}_t + \mathbf{X}_t' \mathbf{\Gamma}_h + \varepsilon_{t+h}, \tag{1}$$

where $\Delta \widehat{g}_t$ is the fitted innovation in government-spending growth from the first stage using the external instrument(s) and the same lagged controls, and ε_{t+h} denoting the regression error term. The regressor vector $\mathbf{X}_t$ includes a constant and p lags of the transformed outcome and government-spending growth (baseline $p = 4$). For level-type outcomes (unemployment, inflation, and inflation expectations), $\mathbf{X}_t$ additionally includes p lags of the outcome in levels. The policy-rate series enters in first differences in percentage points.

Estimation is implemented in stacked form by pooling all $h = 0, \ldots, H$ equations and interacting horizon indicators with $\Delta \widehat{g}_t$ and $\mathbf{X}_t$. Horizon-by-horizon first-stage diagnostics and weak-IV-robust procedures are documented in Appendix B.

4.2 Smooth Local Projections (SLP–IV)

To increase precision without changing identification, impulse responses are re-estimated using Smooth Local Projections with instruments (SLP–IV), following Barnichon and Brownlees (2019). SLP–IV retains the LP–IV design but constrains the fiscal-shock response sequence to vary smoothly with the horizon, while leaving the coefficients on lagged controls unrestricted across horizons. The set of lagged controls $\boldsymbol{X_t}$ is identical to that used in the LP–IV specification.

Let $\beta(h)$ denote the response at horizon h, with $\beta_h = \beta(h)$ for $h = 0, \ldots, H$. Over $[0, H]$, $\beta(h)$ is represented using a cubic B-spline basis, $\beta(h) = B(h)^\top c$, where $B(h)$ collects spline basis functions and c is the associated coefficient vector. The resulting SLP–IV specification is

$$y_{t+h} = \beta(h)\, \Delta\widehat{g}_t + \boldsymbol{X_t'\Gamma_h} + \varepsilon_{t+h}, \qquad h = 0, \ldots, H, \tag{2}$$

where $\Delta\widehat{g}_t$ is the fitted innovation in government-spending growth from the first stage, and $\boldsymbol{\Gamma_h}$ remains horizon-specific and unrestricted.

A roughness penalty is applied to the spline coefficients to discipline medium- and long-horizon sampling noise, while leaving the control coefficients unpenalised. The smoothing parameter is selected by contiguous time-series cross-validation, and the baseline uses modest undersmoothing relative to the prediction-optimal value to improve interval coverage. Basis construction, penalty details, cross-validation, and tuning sweeps are reported in Appendix C.

4.3 Time-Varying Smooth Local Projections (TVP–SLP–IV)

To summarise state dependence across monetary-policy regimes, SLP–IV is extended to a time-varying parameter specification estimated on 1985Q1–2023Q4. At date t, the response is $\beta_t(h) = \mathbf{B}(h)^\top \mathbf{b}_t$, where $\mathbf{B}(h)$ is a cubic B-spline basis and $\mathbf{b}_t$ collects time-varying spline coefficients. The observation equation is

$$y_{t+h} = \beta_t(h)\, \Delta\widehat{g}_t + \mathbf{X}_t'\mathbf{\Gamma}_t + \varepsilon_{t+h}, \qquad h = 0, \ldots, H,$$

where $\mathbf{\Gamma}_t$ denotes time-varying control coefficients that are restricted to be common across horizons at date t.

Time variation is modelled in a linear Gaussian state-space form by stacking $\boldsymbol{\alpha}_t \equiv (\mathbf{b}_t', \mathbf{\Gamma}_t')'$ and imposing a diagonal AR(1) law of motion with a common persistence parameter $\delta \in (0, 1)$ and diagonal innovation covariance. Estimation is conducted via Kalman filtering and smoothing, with δ selected by predictive cross-validation based on innovation fit. Uncertainty is quantified using Durbin–Koopman (Durbin and Koopman 2002) simulation-smoother full-path draws, mapped into horizon-specific impulse responses.

Regime-specific TVP impulse responses are obtained by averaging draw-level time-varying responses over the regime windows in Section 3.3. Regimes enter only through these averaging windows; the model is not re-estimated separately by regime. Additional implementation details and robustness checks are reported in Appendix C.

4.4 Inference, Diagnostics, and Multiplier Construction

For LP–IV and SLP–IV, uncertainty bands are constructed using a circular moving-block bootstrap over base quarters (baseline block length $\ell = 6$, $B = 999$ replications), reporting 68% and 90% percentile bands. Newey–West HAC (Newey and West 1987) covariance matrices are used for point estimates and for first-stage diagnostics.

Weak-instrument diagnostics are computed horizon-by-horizon using the Montiel–Olea–Pflueger effective F statistic and partial R^2 (Montiel Olea and Pflueger 2013); horizons with $F_{\text{MOP}}(h) < 10$ are flagged as weak (Stock and Yogo 2005) and, where relevant, weak-IV-robust Anderson–Rubin confidence sets (Anderson and Rubin 1949) are reported. Over-identification on the AG–Ramey overlap sample under the Both–IV specification is assessed using Hansen's J statistic. Diagnostic and bootstrap details are reported in Appendix B.

For TVP–SLP–IV, uncertainty bands are obtained from simulation-smoother draws of the state vector, mapped into time-varying impulse responses and then into regime-specific averages by averaging draw-level responses over quarters in each regime window.

Cumulative multipliers at horizon H for LP–IV and SLP–IV are

$$M(H) = \frac{\sum_{h=0}^{H} \Delta Y_h}{\sum_{h=0}^{H} \Delta G_h}, \quad H \in \{8, 16\}. \tag{3}$$

Log impulse responses (which approximate percentage changes) are converted into level changes by scaling each response by the sample-average real per-capita level of the corresponding variable over 1985Q1–2019Q4. This conversion expresses the effects in real per-capita units (level changes) rather than in log points. The same scaling is applied within each bootstrap draw so that the reported uncertainty bands are also stated in level units.

For TVP–SLP–IV, regime-specific multipliers are constructed analogously from regime-averaged log responses; point estimates and 68% and 90% bands are obtained from the median and empirical percentiles of the draw-level multiplier distribution. Exact mapping conventions and window treatments are reported in Appendix C.4.

5 Results

This section presents the empirical findings. The analysis first summarises LP–IV impulse responses and first-stage diagnostics as a transparent but noisy benchmark. It then turns to the smoothed SLP–IV baseline, for which a single mechanism grid provides constant-parameter impulse responses for output, government spending, and key components. State dependence is studied using the time-varying TVP–SLP–IV specification. The final subsections summarise cumulative output multipliers and document prices, expectations, and mechanisms. The constant-parameter SLP–IV baseline reports cumulative multipliers in dollar units, constructed from level responses (log IRFs mapped into level changes using the standard small-response approximation). The TVP–SLP–IV specification instead reports regime multipliers as unit-free ratios of cumulative log responses. These regime multipliers are percent-to-percent objects: a value near one indicates that output and government spending exhibit similar cumulative percentage responses through horizon H, rather than implying a one-for-one dollar effect. Under small responses and slowly varying base levels, the unit-free ratios can be numerically close to level-based (dollar) counterparts obtained by rescaling with regime-specific means $Y_0^{(r)}$ and $G_0^{(r)}$ (Appendix C.4).

5.1 LP–IV Benchmark Results

In the unsmoothed LP–IV benchmark, impulse responses to an Auerbach–Gorodnichenko (AG) spending shock on the structural sample 1985Q1–2019Q4 are characterised by small and imprecisely estimated movements in real per-capita GDP. The quarter-0 point estimate for output is mildly positive, but the confidence bands are wide and include zero. Over the first few quarters, point estimates remain close to zero in magnitude and fluctuate in sign, while bands stay wide. At medium and longer horizons, estimates continue to oscillate around zero and uncertainty increases, a pattern typical of unsmoothed local projections with limited information per horizon.

A shorter-horizon run with $H \leq 12$ removes the terminal oscillations seen in the $H = 16$ specification, confirming that these fluctuations reflect sampling noise. LP–IV is therefore used primarily as a diagnostic benchmark for the smoothed estimators rather than as the preferred specification. The full LP–IV panel for $H = 16$, overlaid with the shorter-horizon $H \leq 12$ run, is reported in Figure 7 in Appendix B as a visual diagnostic.

5.2 First-Stage Diagnostics

Horizon-by-horizon first-stage diagnostics for the AG instrument are reported in Appendix B, Table 5. The Montiel–Olea–Pflueger effective F statistics, $F_{\text{MOP}}(h)$, are well above conventional weak-IV thresholds at all horizons $h = 0, \ldots, 16$, and the partial R^2 is high and stable across horizons, indicating a strong and stable first stage.

On the AG–Ramey overlap sample (1985Q1–2015Q4), Appendix B, Table 6 shows that the Ramey-only specification is weak at all horizons, with $F_{\text{MOP}}(h)$ below the weak-instrument cutoff throughout. For this design, Appendix B therefore reports Anderson–Rubin confidence sets by horizon (Table 7) to provide weak-instrument-robust inference. The Both–IV (AG+Ramey) specification strengthens identification relative to Ramey-only, but the AG-only design remains the strongest first-stage configuration.

Accordingly, the AG-based LP–IV and SLP–IV results in the main text are interpreted as supported by a robust first stage over the full range of horizons considered, while weak-IV concerns are confined to the Ramey-only robustness specification.

5.3 Baseline SLP–IV Multipliers

Cumulative dollar multipliers $M(H)$ are defined in Section 4.4. In the constant-parameter baseline, $M(8)$ and $M(16)$ are constructed from SLP–IV impulse responses estimated on the structural sample 1985Q1–2019Q4 using the AG instrument; LP–IV multipliers on the same window are reported only for comparison.

Table 1: Cumulative dollar multipliers, SLP–IV baseline (AG; 1985Q1–2019Q4).

	Horizon 8			Horizon 16		
Method	Point	68% CI	90% CI	Point	68% CI	90% CI
LP–IV	0.50	[-0.48, 1.21]	[-1.15, 1.95]	1.13	[-0.88, 1.75]	[-1.62, 3.21]
SLP–IV	0.59	[-0.28, 1.60]	[-0.94, 2.65]	0.80	[-0.43, 3.14]	[-1.40, 5.28]

Table 1 shows that SLP–IV point estimates for the output multiplier are positive and below one at both horizons and are similar in sign and order of magnitude to the corresponding LP–

IV values. Pooling information across horizons via SLP–IV primarily smooths the underlying impulse responses, but the resulting cumulative multipliers are still estimated with substantial uncertainty: the 68% and 90% intervals for the SLP–IV multipliers include zero at both horizons, so the data are consistent with small to moderately positive medium-run multipliers and do not pin down their exact magnitude.

Relative to the existing literature, these estimates lie toward the lower part of reported ranges but remain compatible with structural VAR and narrative results once differences in identification and definitions are taken into account. The baseline window 1985Q1–2019Q4 covers primarily the post-Volcker Great Moderation and its aftermath, when monetary policy is typically described by an active, inflation-stabilising rule. In this sense, the SLP–IV multipliers provide a data-driven "normal-times" benchmark of a positive but sub-unity government-purchases multiplier.

5.4 Smooth Local Projections (SLP–IV)

Recall from Section 4.2 that smoothed local projections pool information across horizons to sharpen impulse responses while preserving the LP–IV identification and control structure. Figure 1 reports the constant-parameter mechanism grid for the SLP–IV baseline on 1985Q1–2019Q4, together with the corresponding LP–IV profiles. Each panel displays the responses of a key variable (output, government spending, private consumption, private investment, headline PCE inflation, one-year-ahead inflation expectations, unemployment, and the policy rate) to an Auerbach–Gorodnichenko spending shock. Real activity variables are reported in log changes, while inflation, unemployment, inflation expectations, and the policy rate are reported in percentage points. In Figure 1, real activity variables—output, government spending, consumption, and investment—are shown in log changes. Inflation, expectations, unemployment, and the policy rate are instead scaled to correspond to a fiscal shock that raises government purchases by 1 percent of GDP on impact. This normalization is applied only to rate variables, not to real quantities. Relative to the unsmoothed LP–IV benchmark, the SLP–IV responses are smoother, especially at longer horizons, and damp spurious terminal oscillations while leaving the medium-run shape and level broadly unchanged. The 68% and 90% bootstrap bands (reported only for SLP–IV) remain wide, so medium-run effects (particularly for output and government spending) are modest and imprecisely estimated, in line with the cumulative multipliers in Table 1; the SLP–IV responses in Figure 1 are therefore used as the baseline constant-parameter impulse responses for the multiplier and mechanism analysis.

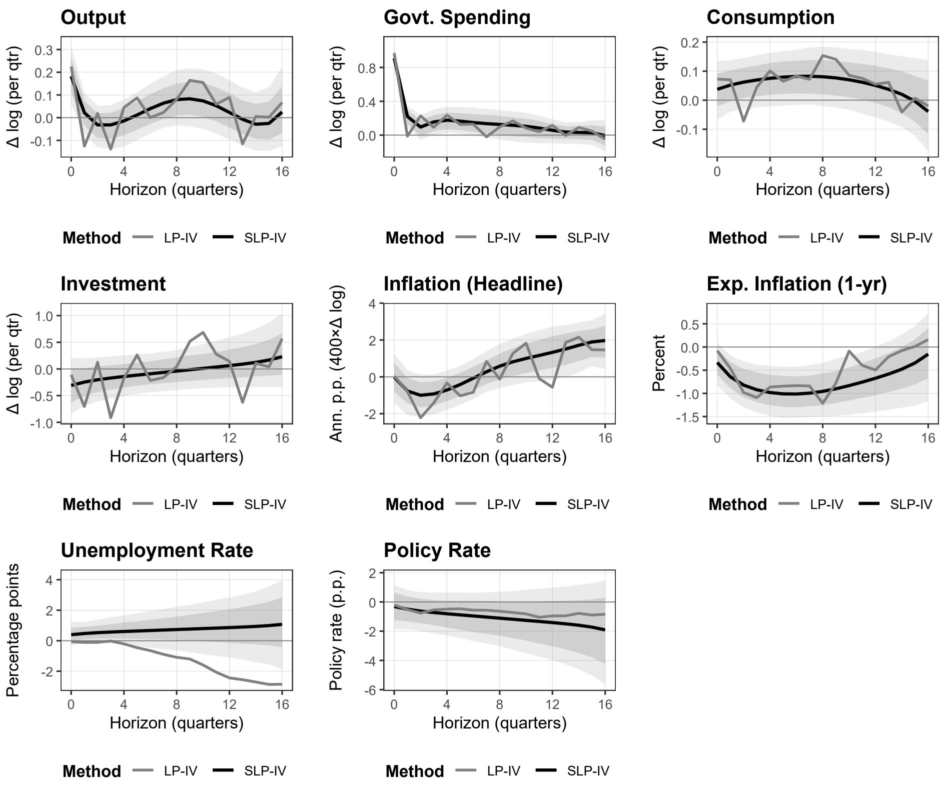


Figure 1: Baseline LP–IV and SLP–IV mechanism impulse responses and cumulative responses (AG instrument; 1985Q1–2019Q4). Notes: Each panel plots LP–IV and SLP–IV responses for a key variable; shaded bands depict 68% and 90% bootstrap intervals for SLP–IV.

5.5 State Dependence: TVP–SLP–IV

State dependence across monetary-policy regimes is analysed with the TVP–SLP–IV model introduced in Section 4.3, estimated on 1985Q1–2023Q4. In this specification the spline coefficients in the impulse responses evolve gradually over time, delivering time-varying impulse responses $\beta_t(h)$ for output, spending, and the mechanism variables, identified throughout by the Auerbach–Gorodnichenko instrument.

Regime-averaged responses in Figure 2 are obtained by averaging the time-varying coefficients over the monetary-policy windows defined in Section 3.3. The top row summarises the responses of output, government spending, and consumption. In all regimes, government spending jumps on impact and then decays over subsequent quarters. Output and consumption point estimates are generally close to zero at most horizons, and the associated uncertainty bands are wide. The point estimates nonetheless display a recurring pattern: small negative responses over the first few quarters after the shock, followed by a modest positive hump peaking around five to six quarters, and a gradual return toward baseline thereafter. Differences across regimes in the level and persistence of these responses are small relative to the 68% and 90% bands, indicating limited evidence of regime heterogeneity in the impulse-response profiles.

The bottom row highlights unemployment and the policy rate. Unemployment responses are modest and not stable across horizons: unemployment declines on impact but subsequently rises toward (and, at some horizons, slightly above) baseline, with wide uncertainty bands. These labour-market responses are therefore interpreted cautiously and are not treated as a central mechanism for the multiplier estimates. For the policy rate, estimated responses are modest and imprecisely estimated, and regime-specific paths are nearly identical. The point estimates suggest a small increase on impact followed by a gradual return toward baseline, but uncertainty bands are wide throughout. Consequently, the policy-rate evidence provides limited support for distinguishing monetary offset across regimes, including between ZLB and non-ZLB episodes, and the discussion of regime heterogeneity is therefore based primarily on the price and expectations responses rather than on the funds-rate response itself. Point estimates in the long non-ZLB regimes are consistent with partial monetary offset, but the policy-rate responses are imprecisely estimated, so the evidence is suggestive rather than conclusive.

Regime-specific cumulative multipliers in Table 2 are unit-free log-response multipliers, computed as ratios of cumulative regime-averaged log responses of output and government spending as in Section 4.4. These objects should be interpreted as percent-to-percent ratios: a multiplier near one indicates similar cumulative percentage responses of output and spending (not dollars-per-dollar), while a positive value indicates that the cumulative responses share the same sign. At $H = 8$, point estimates in the three long pre-Covid regimes are positive but below one, with substantial uncertainty. At $H = 16$, point estimates are close to one in the long pre-Covid regimes. At $H = 16$, the ZLB–GFC (2008Q4–2015Q3) estimate is strictly positive (its 68% and 90% intervals exclude zero), providing the clearest evidence that cumulative output and spending move in the same direction over longer horizons in percent-to-percent terms. Differences across regimes appear mainly in precision rather than in large shifts in the underlying impulse responses or in the policy-rate response. To facilitate interpretation, Appendix D.2 reports separate TVP–SLP–IV impulse-response grids for each of the three long pre-Covid regime.

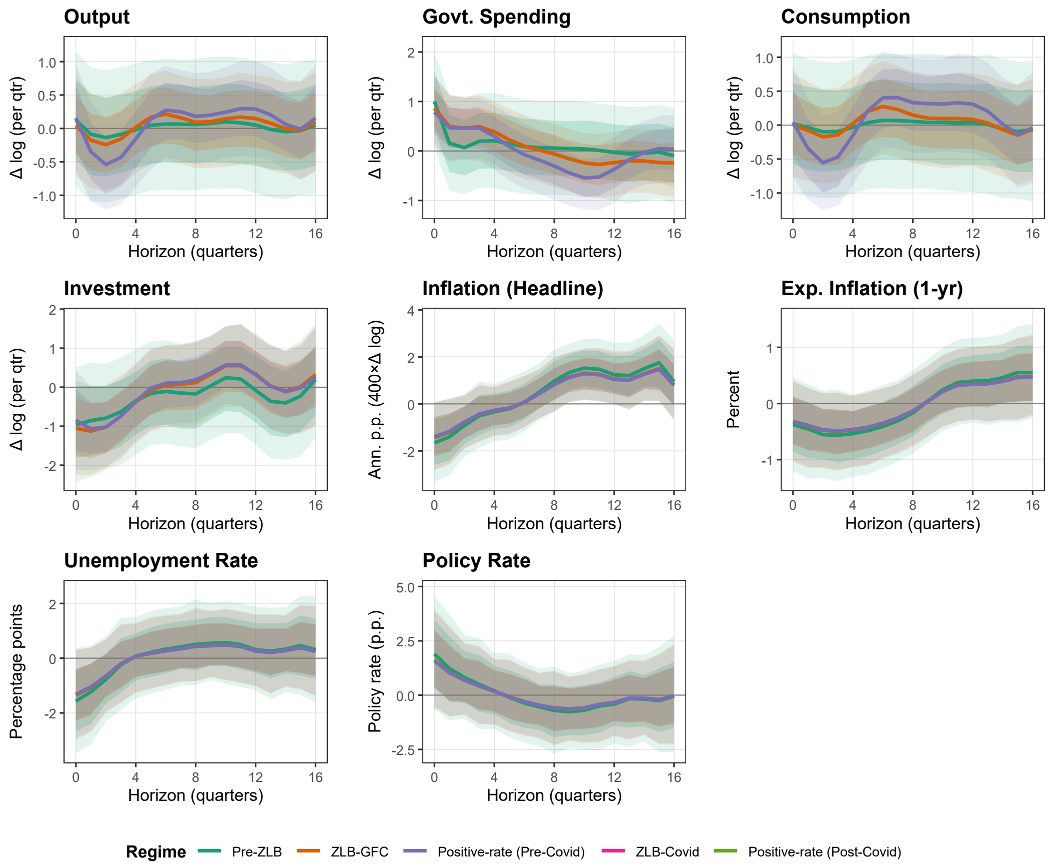


Figure 2: Regime-averaged IRFs from TVP–SLP–IV (1985Q1–2023Q4): Pre–ZLB (1985Q1–2008Q3), ZLB–GFC (2008Q4–2015Q3), Positive-rate (Pre-Covid) (2015Q4–2019Q4), ZLB–Covid (2020Q2–2021Q4), Positive-rate (Post-Covid) (2022Q2–2023Q4).

Table 2: Regime multipliers (unit-free log-response ratios) by monetary-policy regime (TVP–SLP–IV; 1985Q1–2023Q4).

	Horizon 8			Horizon 16		
Regime	Point	68% CI	90% CI	Point	68% CI	90% CI
Pre–ZLB	0.77	[0.14, 1.95]	[-1.45, 4.27]	0.96	[0.62, 1.40]	[-0.15, 1.94]
ZLB–GFC	0.41	[-0.62, 1.96]	[-5.49, 5.90]	1.02	[0.81, 1.26]	[0.23, 1.78]
Positive-rate (Pre-Covid)	0.48	[-0.52, 2.40]	[-3.28, 6.66]	1.04	[0.63, 1.43]	[-0.79, 2.57]

Notes: Entries are unit-free log-response multipliers (ratios of cumulative regime-averaged log responses). A value of 1 indicates similar cumulative percentage responses of output and government spending through horizon H (percent-to-percent, not dollars-per-dollar).

5.6 Prices, Expectations, and Mechanisms

The TVP regime grid in Figure 2 also summarises the behaviour of prices, expectations, and expenditure components, and Appendix D.2 reports separate grids for each of the three long pre-Covid regimes. The panels for headline PCE inflation and one-year-ahead inflation expectations can be read together, while the remaining panels contain investment and other mechanism variables. Across regimes, headline inflation responds only mildly: point estimates are small, can change sign across horizons, and the associated uncertainty bands include zero. One-year-ahead MICH inflation expectations also exhibit small movements and are estimated imprecisely, with wide bands that span zero in all regimes. Regime heterogeneity in prices and expectations is therefore limited and statistically weak; the data do not provide strong evidence of state dependence in the inflation or expectations responses to government spending shocks. Even when cumulative sums are non-zero, the quarter-by-quarter price and expectations responses are small and imprecisely estimated.

Mechanism results in Figure 2 and Table 3 indicate that private demand (consumption and investment) contributes importantly to the transmission of government spending shocks. Mechanism entries are approximate because log IRFs are mapped into level units using a first-order approximation (Appendix C.4). These level-rescaled mechanism summaries are reported solely to aid economic interpretation and are conceptually distinct from the unit-free regime multipliers in Table 2. In the constant-parameter SLP–IV baseline, Figure 1 shows that medium-run consumption responses, while modest in magnitude, are generally positive, whereas investment responses exhibit greater horizon sensitivity. The corresponding cumulative responses of private output, $Y_t - G_t$, reported in Appendix D, Table 15, are also close to zero in cumulative terms on average. The regime-specific cumulative responses from the TVP model in Table 3 display a similar pattern across the three long pre-Covid regimes. At horizons $H = 8$ and $H = 16$, the cumulative response of consumption is positive and economically non-negligible in each regime, while the corresponding investment responses are small in magnitude and frequently statistically indistinguishable from zero. Cumulative private-output responses are also positive but only moderately sized. Cumulative unemployment responses are negative at $H = 8$, reflecting a temporary decline in the unemployment rate after the shock, but are small and positive in cumulative terms at $H = 16$, so long-run labour-market effects are weakly identified in level terms. Differences in these mechanism responses across the Pre–ZLB, ZLB–GFC, and Positive-rate (Pre-Covid) regimes are modest relative to their uncertainty bands, although long-horizon ($H = 16$) point estimates for consumption and private output tend to be somewhat larger in the ZLB–GFC and Positive-rate (Pre-Covid) regimes. Taken together with the regime log-response multipliers in Table 2, these patterns are consistent with modest positive output multipliers that arise mainly from the direct increase in government purchases coupled with a moderate rise in consumption, rather than from large crowding-in of private investment.

The policy-rate responses in Figure 2 provide only limited evidence on monetary offset across regimes. Point estimates in the long non-ZLB regimes suggest a small, temporary increase in the funds rate (in p.p.) following a spending shock, which is consistent with partial monetary offset, but the associated uncertainty implies that this evidence is suggestive rather than conclusive.

Table 3: Cumulative responses of mechanisms by regime (TVP–SLP–IV; 1985Q1–2023Q4).

Regime	Outcome	$R(8)$	$R(16)$
Pre–ZLB	C (PCE)	2.26	2.77
Pre–ZLB	I (GPDIC1)	0.38	0.67
Pre–ZLB	$Y - G$ (private output)	2.44	3.30
Pre–ZLB	UNRATE (p.p.)	-2.14	0.94
Pre–ZLB	Inflation, headline (p.p. annualised)	-3.27	7.02
Pre–ZLB	MICH expectations (p.p.)	-3.62	-0.77
ZLB–GFC	C (PCE)	1.72	3.53
ZLB–GFC	I (GPDIC1)	-0.26	0.86
ZLB–GFC	$Y - G$ (private output)	1.12	4.25
ZLB–GFC	UNRATE (p.p.)	-2.13	0.88
ZLB–GFC	Inflation, headline (p.p. annualised)	-3.23	6.91
ZLB–GFC	MICH expectations (p.p.)	-3.61	-0.77
Positive-rate (Pre-Covid)	C (PCE)	2.18	4.09
Positive-rate (Pre-Covid)	I (GPDIC1)	-0.06	1.18
Positive-rate (Pre-Covid)	$Y - G$ (private output)	1.76	4.98
Positive-rate (Pre-Covid)	UNRATE (p.p.)	-2.32	1.01
Positive-rate (Pre-Covid)	Inflation, headline (p.p. annualised)	-3.49	7.65
Positive-rate (Pre-Covid)	MICH expectations (p.p.)	-3.96	-0.83

Notes: Columns $R(8)$ and $R(16)$ report cumulative responses through horizons $H = 8$ and $H = 16$ quarters, computed as $\sum_{h=0}^{H}$ of the regime-specific TVP–SLP–IV impulse responses. Only the three long pre-Covid regimes are shown; the Covid and post-Covid regimes are omitted because long-horizon level rescalings may be weakly identified. For C (PCE), I (GPDIC1), and $Y-G$ (private output), log responses are mapped into real per-capita level changes using regime-specific base levels (Appendix C.4), and entries equal the ratio of cumulative level changes to cumulative government-spending level changes, in real dollars per person per $1 of cumulative government spending. For UNRATE, headline PCE inflation, and MICH inflation expectations, entries are cumulative percentage-point responses to a shock normalised to raise government consumption and investment by 1 percent of GDP on impact. Level rescalings rely on a first-order log-to-level approximation; the headline regime multipliers remain the unit-free log-response ratios in Table 2.

6 Robustness and Sensitivity

This section summarises a compact set of robustness checks designed to verify that the main conclusions are not driven by particular choices of instruments, sample window, or tuning parameters. Three dimensions are varied systematically. First, alternative external instruments are considered by re-estimating overlap-sample specifications with the Ramey defence-news shock and with a Both–IV set that combines AG and Ramey shocks. For the combined-instrument design, Hansen J over-identification diagnostics provide no evidence against the over-identifying restrictions on the overlap sample. Second, sensitivity to horizon choice and to the treatment of the pandemic period is assessed using a shorter-horizon robustness run with $H \leq 12$ and a Covid-exclusion diagnostic sample that drops 2020Q2–2021Q4 in constant-parameter checks; consistent with the regime convention used elsewhere, the post-Covid segment begins in 2022Q2 (excluding the transition quarter 2022Q1) to avoid within-quarter mixing around liftoff. Third, SLP and TVP tuning parameters (including the spline smoothness penalty, knot count in $\{3, 5, 7\}$, and the TVP persistence parameter) are varied over pre-specified grids. Formal first-stage diagnostics (MOP–F statistics, partial R^2, and weak-IV-robust Anderson–Rubin confidence sets for horizons flagged as weak by the configured first-stage threshold) and the associated horizon-by-horizon tables and first-stage plots are reported in Appendix B, while Appendix C documents the SLP and TVP tuning choices and the λ–knot grids underlying the smoothing sensitivity discussion.

Across these variants, AG-based specifications remain strongly identified and deliver impulse-response profiles that are qualitatively consistent with the baseline conclusions. On the AG–Ramey overlap sample, the Both–IV results are comparable in magnitude to the AG-only overlap estimates, whereas the Ramey-only design is weak and yields much wider uncertainty and weak-IV-robust Anderson–Rubin confidence sets. The shorter-horizon $H \leq 12$ run and the Covid-exclusion diagnostic do not change the qualitative medium-run shape of the output response. Tuning sweeps over spline smoothness, knot placement, and the TVP persistence parameter primarily affect smoothness and the width of uncertainty bands rather than the qualitative regime comparisons. No specification overturns the headline findings of modest average output multipliers, clearer long-horizon evidence in the ZLB–GFC regime, and weak responses of inflation and inflation expectations.

7 Conclusion

This work project analyses U.S. government-spending multipliers over 1985Q1–2023Q4 using an external-instrument design, with Auerbach–Gorodnichenko forecast-error shocks as the baseline instrument and Ramey defence-news shocks used only for overlap-sample robustness checks. The headline constant-parameter benchmark is a smooth local projection IV (SLP–IV) specification estimated on 1985Q1–2019Q4 (with LP–IV retained as a diagnostic comparator). State dependence is summarised by a single TVP–SLP–IV model estimated on 1985Q1–2023Q4, with monetary-policy regimes defined as Pre–ZLB, ZLB–GFC, Positive-rate (Pre-Covid), ZLB–Covid, and Positive-rate (Post-Covid); headline regime-average results are reported for the three long pre-Covid regimes.

In the constant-parameter SLP–IV baseline, medium-run cumulative dollar multipliers are below one and estimated imprecisely. In the TVP–SLP–IV specification, regime-averaged multipliers cluster around unity across the three long pre-Covid regimes. The clearest evidence of a positive long-run multiplier arises during the ZLB–GFC (2008Q4–2015Q3) episode, where the $H = 16$ multiplier is slightly above one with strictly positive confidence bands, whereas Pre–ZLB and Positive-rate (Pre-Covid) multipliers at $H = 16$ are of similar magnitude but less precisely estimated. The TVP layer indicates gradual rather than abrupt evolution in multipliers, consistent with New Keynesian environments in which an active Taylor rule is associated with more modest multipliers in normal times, while ZLB episodes are associated with larger and more precisely estimated long-horizon multipliers.

Headline inflation and one-year-ahead inflation expectations respond only weakly to government-spending shocks across regimes, with limited evidence of destabilising price dynamics. Mechanism impulse responses indicate that consumption tends to increase following government-spending shocks. Private output exhibits moderate positive responses in the TVP–SLP–IV model but remains close to zero in the constant-parameter SLP–IV baseline, whereas investment responses are generally small and more weakly identified. Modest output multipliers are therefore associated with relatively stable inflation and expectations, and transmission operates primarily through private demand (especially consumption) rather than through large or persistent movements in prices or unemployment.

Robustness checks across instruments, samples, horizon caps, Covid-exclusion windows, and tuning choices indicate that the qualitative conclusions are not driven by any single specification; the most precisely estimated long-run multipliers continue to arise when policy rates are at or near the ZLB.

These findings suggest that government spending is most effective when monetary policy is constrained at or near the ZLB, without evidence of unanchored inflation in the sample, and that moderate multipliers need not imply substantial crowding-out of private demand. Limitations include reliance on aggregate quarterly U.S. data and on shocks captured by the Auerbach–Gorodnichenko and Ramey instruments, as well as the short length of the Covid and post-Covid regimes relative to the 16-quarter horizon; in these short regimes, long-horizon regime multipliers (notably at $H = 16$) can be weakly identified when the cumulative government-spending response is too small to support stable response ratios, and such entries are treated as not identified in the regime tables. More generally, regime implications for Covid and post-Covid windows are best interpreted as descriptive features of the unified TVP model rather than fully identified long-run effects.

References

Anderson, T. W., and Herman Rubin. 1949. "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations." *The Annals of Mathematical Statistics* 20 (1): 46–63.

Auerbach, Alan J., and Yuriy Gorodnichenko. 2012. "Measuring the Output Responses to Fiscal Policy." *American Economic Journal: Economic Policy* 4, no. 2 (May): 1–27.

Auerbach, Alan J., and Yuriy Gorodnichenko. 2013. "Fiscal Multipliers in Recession and Expansion." In *Fiscal Policy after the Financial Crisis,* edited by Alberto Alesina and Francesco Giavazzi, 63–98. Chicago, IL: University of Chicago Press, June.

Barnichon, Régis, and Christian Brownlees. 2019. "Impulse Response Estimation by Smooth Local Projections." *The Review of Economics and Statistics* 101 (3): 522–530.

Ben Zeev, Nadav, Valerie A. Ramey, and Sarah Zubairy. 2023. "Do Government Spending Multipliers Depend on the Sign of the Shock?" *AEA Papers and Proceedings* 113:382–387.

Blanchard, Olivier J., and Roberto Perotti. 2002. "An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output." *The Quarterly Journal of Economics* 117 (4): 1329–1368.

Christiano, Lawrence J., Martin Eichenbaum, and Sergio Rebelo. 2011. "When Is the Government Spending Multiplier Large?" *Journal of Political Economy* 119 (1): 78–121.

Clarida, Richard, Jordi Galí, and Mark Gertler. 2000. "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory." *The Quarterly Journal of Economics* 115 (1): 147–180.

Durbin, James, and Siem Jan Koopman. 2002. "A Simple and Efficient Simulation Smoother for State Space Time Series Analysis." *Biometrika* 89 (3): 603–616.

Eggertsson, Gauti B. 2011. "What Fiscal Policy Is Effective at Zero Interest Rates?" In *NBER Macroeconomics Annual 2010, Volume 25,* edited by Daron Acemoglu and Michael Woodford, 59–112. Chicago: University of Chicago Press.

Goodfriend, Marvin, and Robert G. King. 2005. "The Incredible Volcker Disinflation." *Journal of Monetary Economics* 52 (5): 981–1015.

Hansen, Lars Peter. 1982. "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica* 50 (4): 1029–1054.

Hills, Timothy S., and Taisuke Nakata. 2018. "Fiscal Multipliers at the Zero Lower Bound: The Role of Policy Inertia." *Journal of Money, Credit and Banking* 50 (1): 155–172.

Jordà, Òscar. 2005. "Estimation and Inference of Impulse Responses by Local Projections." *American Economic Review* 95 (1): 161–182.

Montiel Olea, José Luis, and Carolin Pflueger. 2013. "A Robust Test for Weak Instruments." *Journal of Business & Economic Statistics* 31 (3): 358–369.

Nakamura, Emi, and Jón Steinsson. 2014. "Fiscal Stimulus in a Monetary Union: Evidence from US Regions." *American Economic Review* 104 (3): 753–792.

Newey, Whitney K., and Kenneth D. West. 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55 (3): 703–708.

Ramey, Valerie A. 2011a. "Can Government Purchases Stimulate the Economy?" *Journal of Economic Literature* 49 (3): 673–685.

Ramey, Valerie A. 2011b. "Identifying Government Spending Shocks: It's all in the Timing." *The Quarterly Journal of Economics* 126, no. 1 (February): 1–50.

Ramey, Valerie A. 2016. "Defense News Shocks, 1889–2015: Estimates Based on News Sources." Unpublished manuscript, University of California, San Diego.

Ramey, Valerie A. 2019. "Ten Years after the Financial Crisis: What Have We Learned from the Renaissance in Fiscal Research?" *Journal of Economic Perspectives* 33 (2): 89–114.

Ramey, Valerie A., and Sarah Zubairy. 2018. "Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data." *Journal of Political Economy* 126 (2): 850–901.

Stock, James H., and Mark W. Watson. 2003. "Has the Business Cycle Changed and Why?" In *NBER Macroeconomics Annual 2002, Volume 17,* edited by Mark Gertler and Kenneth S. Rogoff, 159–230. Cambridge, MA: MIT Press.

Stock, James H., and Motohiro Yogo. 2005. "Testing for Weak Instruments in Linear IV Regression." In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg,* edited by Donald W. K. Andrews and James H. Stock, 80–108. Cambridge: Cambridge University Press.

Woodford, Michael. 2011. "Simple Analytics of the Government Expenditure Multiplier." *American Economic Journal: Macroeconomics* 3 (1): 1–35.

A Data and Provenance

This Appendix documents the provenance and construction of all data series used in the analysis. Each variable is listed in Table 4 with its source identifier, originating agency, transformation, and retrieval date as logged by the project pipeline. Unless otherwise noted, series use the latest available vintage as of the retrieval dates in Table 4. The constant-parameter SLP–IV specification uses the 1985Q1–2019Q4 structural baseline window; constant-parameter LP–IV is implemented on the same window as a diagnostic and robustness check. The TVP–SLP–IV model is estimated once on the full 1985Q1–2023Q4 sample, and regime summaries are obtained by averaging within pre-defined regime windows rather than by re-estimation on short subsamples.

Notation versus pipeline identifiers.

In the main text and model specification, uppercase symbols (e.g., Y_t, G_t, C_t, I_t) denote the conceptual macroeconomic aggregates, expressed as real per-capita quantities (log levels or log differences depending on the regression). In Table 4, the variable names recorded in the Alias (or identifier) column correspond to dataset and code identifiers used in the replication pipeline and are therefore typically lowercase (e.g., `y`, `g`, `pce`, `inv`). These aliases map one-for-one into the paper notation:

$$Y_t \leftrightarrow \texttt{y}, \qquad G_t \leftrightarrow \texttt{g}, \qquad C_t \leftrightarrow \texttt{pce}, \qquad I_t \leftrightarrow \texttt{inv}.$$

Accordingly, capitalisation differences reflect a notational distinction between mathematical symbols and code identifiers rather than a difference in definition or scaling. The same convention applies to indicators (e.g., the ZLB dummy is stored as `zlb` for contract compatibility within the codebase).

A.1 Sources and Series IDs

The dataset combines quarterly macroeconomic aggregates and higher-frequency indicators (monthly or daily) aggregated to the quarterly frequency. All macroeconomic series are taken in seasonally adjusted form where such adjustments are provided by the originating agency. The main sources are:

- **Bureau of Economic Analysis (BEA):** national accounts aggregates, including real GDP, real government purchases, real private investment, real consumption, and price indices.
- **Bureau of Labor Statistics (BLS):** unemployment and population age 16+.
- **Federal Reserve System:** policy rates and the target-range upper bound used in the ZLB indicator.
- **University of Michigan:** one-year-ahead median inflation expectations.

Series identifiers, aliases, and source codes are provided in Table 4.

Vintage note. Realized government purchases used to construct the SPF forecast-error instrument are taken from the latest-available BEA/FRED series as of the pipeline retrieval dates (Table 4), rather than real-time/first-release vintages; consequently, the measured forecast error may incorporate subsequent data revisions, and a vintage-matched real-time construction is left for robustness work.

A.2 Transformations and Units

All real aggregates are converted to per-capita log levels using the quarterly average of the relevant population series. Higher-frequency series are first averaged to the quarterly frequency so that all series share a common quarterly time index and are then differenced as required by the specification. Inflation from price indices is computed as

$$\pi_t \;=\; 400 \cdot \Delta \log(\text{index}_t),$$

which yields annualised percent rates. Real activity variables enter the regressions in log differences.

For the constant-parameter LP–IV and SLP–IV specifications, and for the activity mechanisms, impulse responses are mapped into level changes (real dollars per capita) when computing dollar multipliers and mechanism entries. In contrast, TVP–SLP–IV regime multipliers are reported as unit-free log-response ratios. Unemployment, inflation, and expectations mechanisms are reported in percentage points (or percent units for level series) as cumulative responses to a government-spending shock normalised to raise real government spending by 1 percent of GDP on impact.

A.3 Regime Indicator Windows

The ZLB dummy equals one when the federal funds target-range upper bound is less than or equal to 0.25% on a quarterly-average basis. This rule divides the sample into five regimes:

- Pre–ZLB (1985Q1–2008Q3)
- ZLB–GFC (2008Q4–2015Q3)
- Positive-rate (Pre-Covid) (2015Q4–2019Q4)
- ZLB–Covid (2020Q2–2021Q4)
- Positive-rate (Post-Covid) (2022Q2–2023Q4)

For reference, the quarter windows used in regime averaging are Pre–ZLB (1985Q1–2008Q3), ZLB–GFC (2008Q4–2015Q3), Positive-rate (Pre-Covid) (2015Q4–2019Q4), ZLB–Covid (2020Q2–2021Q4), and Positive-rate (Post-Covid) (2022Q2–2023Q4). Boundary quarters are treated as transition quarters and coded as NA in the regime indicator for regime averaging.

These windows are used to construct regime indicator variables for the TVP–SLP–IV design and to compute regime-averaged summaries from a single full-sample estimation. The underlying parameters are never re-estimated on short subsamples.

A.4 Provenance Table and Revision Policy

Table 4 records, for every series in the dataset, the variable identifier used in the analysis, the original source and series identifier, the transformation applied in the pipeline, and the retrieval date. The same provenance information is stored in machine-readable form within the replication package and is summarised in the panel-construction and validation logs.

Table 4: Data provenance and transformations

Alias	Series ID	Source	Transform pipeline	Sample start	Sample end	Retrieval date
y	GDPC1	FRED	Quarterly real GDP. Log real per-capita $y_t = \log(\text{GDPC1}_t/\text{POP\_Qavg}_t)$, where POP_Qavg is the quarterly average of CNP16OV. Growth series $dy_t = \Delta y_t$; IRFs are in Δ log units and are mapped back to level changes for dollar multipliers.	1985Q1	2023Q4	2025-10-15
g	GCEC1	FRED	Same treatment as GDPC1: log real per-capita $g_t = \log(\text{GCEC1}_t/\text{POP\_Qavg}_t)$, with $dg_t = \Delta g_t$. Level change in g is the denominator for dollar multipliers.	1985Q1	2023Q4	2025-10-15
pce	PCECC96	FRED	Log real per-capita $c_t = \log(\text{PCECC96}_t/\text{POP\_Qavg}_t)$. Growth series $dc_t = \Delta c_t$; IRFs use Δ log units and are converted to level changes when computing consumption multipliers.	1985Q1	2023Q4	2025-10-15
inv	GPDIC1	FRED	Log real per-capita $i_t = \log(\text{GPDIC1}_t/\text{POP\_Qavg}_t)$. Growth series $di_t = \Delta i_t$; same log-diff $\to$ level-change logic as for y and g.	1985Q1	2023Q4	2025-10-15
pop	CNP16OV	FRED	Monthly population (thousands of persons). Quarterly average within each quarter to form POP_Qavg. Used as the denominator for per-capita log levels; scaling by 1,000 shifts log levels by a constant and therefore cancels in log differences (and in ratios such as multipliers when applied consistently). No seasonal adjustment step.	1985Q1	2023Q4	2025-10-24
unrate	UNRATE	FRED	Monthly SA unemployment rate. Quarterly average $\to$ unemployment rate per quarter. Used in percent levels (no log, no Δ) as a mechanism outcome; reported IRFs are cumulative responses in percentage points under the spending-shock normalisation that raises real government spending by 1 percent of GDP on impact. A derived series du_t (first difference, p.p.) is used when changes in unemployment are required.	1985Q1	2023Q4	2025-10-15

Continued on next page

Table 4 (continued)

Alias	Series ID	Source	Transform pipeline	Sample start	Sample end	Retrieval date
pcepi	PCECTPI	FRED	Quarterly headline PCE price index P_t^{PCE}. Headline inflation $\pi_t^{\mathrm{PCE}} = 400\,[\log P_t^{\mathrm{PCE}} - \log P_{t-1}^{\mathrm{PCE}}]$, expressed in annualised percent.	1985Q1	2023Q4	2025-10-15
mich	MICH	FRED	Monthly University of Michigan 1-year-ahead median inflation expectations. Quarterly average of the MICH column; kept in percent levels (no log, no Δ) as required by the baseline specification.	1985Q1	2023Q4	2025-11-27
DFEDTARU	DFEDTARU	FRED	Daily target range upper bound for the federal funds rate. Daily values are aggregated to a quarterly average; construct ZLB dummy $ZLB_t = 1$ if the quarterly average ≤ 0.25, and 0 otherwise. The rate itself is not otherwise transformed.	1985Q1	2023Q4	2025-10-15
fedfunds	FEDFUNDS	FRED	Monthly effective federal funds rate (percent, NSA). Quarterly average $\rightarrow$ rate per quarter; the monetary control is the first difference Δi_t^{FF} (percentage points). No logs or seasonal adjustment.	1985Q1	2023Q4	2025-10-15
z_ag	SPF_FE_G_AGtype	Constructed (SPF+BEA)	AG-type SPF forecast-error government spending shock z_t^{AG}, constructed within this project under a uniform procedure and extended through 2023Q4 (log-growth/forecast-error units; values around 0.01 correspond to about 1 percent in log-growth units). Used directly as the baseline external IV over 1985Q1–2023Q4; no logs or differences. Any scaling/standardisation is handled inside the IV code.	1985Q1	2023Q4	2025-09-04
z_ramey	RZDAT	Ramey	Ramey defence news shock constructed as a share of GDP in the original Ramey dataset. Decimal-quarter dates are mapped to standard quarter labels; the IV is used on the 1985Q1–2015Q4 overlap with the main panel. No additional transforms; any standardisation is done inside the estimation code.	1985Q1	2015Q4	2025-10-15

Continued on next page

Table 4 (continued)

Alias	Series ID	Source	Transform pipeline	Sample start	Sample end	Retrieval date
dy	dy	derived	Δ log(GDPC1/POP_Qavg) (real per-capita output growth).	1985Q1	2023Q4	–
dg	dg	derived	Δ log(GCEC1/POP_Qavg) (real per-capita government spending growth).	1985Q1	2023Q4	–
dc	dc	derived	Δ log(PCECC96/POP_Qavg) (real per-capita consumption growth).	1985Q1	2023Q4	–
di	di	derived	Δ log(GPDIC1/POP_Qavg) (real per-capita investment growth).	1985Q1	2023Q4	–
pi_pce	pi_pce	derived	$400 \times \Delta$ log(PCECTPI) (headline PCE inflation).	1985Q1	2023Q4	–
ZLB	ZLB	derived	Indicator from rule DFEDTARU quarterly average ≤ 0.25 (see DFEDTARU row).	1985Q1	2023Q4	–
zlb	zlb	derived	Lowercase alias of ZLB for contract compatibility in the code.	1985Q1	2023Q4	–
regime_tvpslp	regime_tvpslp	derived	Regime labels (Pre–ZLB, ZLB–GFC, Positive-rate (Pre-Covid), ZLB–Covid, Positive-rate (Post-Covid)); NA in transition quarters.	1985Q1	2023Q4	–
regime_slp	regime_slp	derived	Indicator for the Covid ZLB window used in robustness/diagnostic routines (equals ZLB–Covid for 2020Q2–2021Q4; NA otherwise); not used to estimate a separate constant-parameter SLP model.	1985Q1	2023Q4	–
y_pc	y_pc	derived	Real per-capita output level (consistent with dy).	1985Q1	2023Q4	–
g_pc	g_pc	derived	Real per-capita government spending level (consistent with dg).	1985Q1	2023Q4	–
c_pc	c_pc	derived	Real per-capita consumption level (consistent with dc).	1985Q1	2023Q4	–
i_pc	i_pc	derived	Real per-capita investment level (consistent with di).	1985Q1	2023Q4	–

Notes: “Source” indicates the originating database or author. Derived series are constructed within the replication scripts from the listed raw series. Retrieval dates are ISO calendar dates corresponding to timestamps recorded by the project pipeline; unless otherwise noted, all series use the latest available vintage as of those dates.

B Diagnostics and Tests

This Appendix reports the diagnostic and inference details delegated from Section 4. It documents (i) the first-stage specification and fitted $\Delta\widehat{g}_t$, (ii) strength diagnostics and weak-IV-robust Anderson–Rubin confidence sets, (iii) over-identification diagnostics for Both–IV on the overlap sample, and (iv) the bootstrap design used for LP–IV and SLP–IV inference. It also collects supporting robustness materials (overlap multipliers by instrument, TVP alternative-instrument grids, and horizon/Covid-exclusion diagnostics) referenced in Section 6. All outputs are generated automatically by the project pipeline and recorded in the associated validation logs.

B.1 Sanity Report (Validation Summary)

Coverage: horizons $h = 0, \ldots, 16$ in the baseline specification. Selected robustness runs use a shorter terminal horizon $H = 12$. In weak-instrument designs, interpretation places primary emphasis on results at or below $H \leq 12$, while longer-horizon summaries (e.g. $H = 16$) are retained where reported for completeness.

Weak horizons & AR sets: horizons with Montiel–Olea–Pflueger effective F statistic below the configured weak-instrument threshold $F_{\text{weak}} = 10$ are flagged as weak; Anderson–Rubin confidence sets are constructed by horizon for the Ramey-only specification and summarised in Table 7.

SLP tuning: λ_{CV} selected by contiguous 5-fold cross-validation; estimation uses $0.8 \times \lambda_{\text{CV}}$; knots $= 5$.

TVP tuning: discount factor δ^* selected by predictive cross-validation.

Multipliers: baseline SLP–IV multipliers in Table 1 and cumulative dollar multipliers by instrument on the overlap sample in Table 9, with a combined overview export assembled in Table 16 (Appendix D).

B.2 First-Stage Specification and Construction of $\Delta\widehat{g}_t$

For each outcome, the fitted policy innovation $\Delta\widehat{g}_t$ used in the LP–IV second stage in Section 4.1 is obtained from a first-stage projection of Δg_t on the external instrument(s) and the same lag structure used in the corresponding second stage. For notational simplicity this fitted series is denoted by $\Delta\widehat{g}_t$. Because the core variables enter as quarterly log differences and the baseline specification includes four lags of $\Delta\log(Y_t)$ and $\Delta\log(G_t)$ (with additional outcome lags where required), the effective estimation sample begins in 1986Q2 once the lag structure and complete-case filtering are imposed; stated sample windows refer to the underlying quarterly data coverage prior to this trimming.

Let $\Delta g_t \equiv \Delta\log(G_t)$ denote quarterly government spending growth (in the transformed units described in Section 3.1) and let z_t collect the external instrument(s): AG in the baseline, and AG, Ramey, or Both–IV (AG+Ramey) on the overlap sample. Let $\mathbf{X}_t$ stack the same lagged controls used in the local-projection equation, namely four lags of the outcome and government spending, together with additional lags of the outcome level where required (for unemployment, inflation, inflation expectations, and the policy rate). u_t represents the regression error term.

On the structural sample $t = 1985\text{Q1}, \ldots, 2019\text{Q4}$, the baseline first-stage regression is,

$$\Delta g_t \;=\; \alpha \;+\; \mathbf{z}_t'\boldsymbol{\pi} \;+\; \mathbf{X}_t'\boldsymbol{\kappa} \;+\; u_t. \tag{4}$$

The fitted component of government spending growth,

$$\Delta\widehat{g}_t \;=\; \hat{\alpha} \;+\; \mathbf{z}_t'\hat{\boldsymbol{\pi}} \;+\; \mathbf{X}_t'\hat{\boldsymbol{\kappa}}, \tag{5}$$

is referred to as the policy innovation in government spending growth and is the regressor $\Delta\widehat{g}_t$ entering the LP–IV second-stage equation,

$$y_{t+h} \;=\; \beta_h\Delta\widehat{g}_t + \mathbf{X}_t'\boldsymbol{\Gamma}_h + \varepsilon_{t+h}, \qquad h = 0, \ldots, H,$$

reported in Section 4.1.

For diagnostic reporting, the first stage (4) is evaluated on the common base-quarter sample supporting the full horizon set $h = 0, \ldots, H$ (i.e., quarters t such that y_{t+h} is observed for all horizons up to H). This convention yields a horizon-invariant base-quarter sample first-stage diagnostics in Tables 5 and 6. The second-stage local projections remain horizon-specific through the dependent variable y_{t+h}.

This representation is algebraically equivalent to a conventional two-stage least squares regression of y_{t+h} on Δg_t instrumented by $\mathbf{z}_t$, estimated in stacked form over horizons $h = 0, \ldots, H$ with horizon dummies interacted with Δg_t and the controls $\mathbf{X}_t$. Writing the second stage directly in terms of the fitted policy innovation $\Delta \widehat{g}_t$ is adopted for notational and computational convenience; all diagnostics and weak-instrument-robust procedures below are based on the underlying instrumented relationship implied by (4).

B.3 Implementation of J and AR Tests

Hansen's J statistic is implemented for the over-identified Both–IV specification (AG + Ramey) on the overlap sample. The J statistic is computed using the standard residual-regression formulation: IV residuals from the just-identified regression of the outcome on the instrumented regressor are regressed on the full instrument set, and the resulting test statistic is reported together with its degrees of freedom and p-value in Table 8. This provides a conventional over-identification check on the joint validity of AG and Ramey shocks when used together.

Anderson–Rubin (AR) confidence sets are constructed by horizon in the same local-projection IV design and are reported for the Ramey-only specification in the diagnostic tables (Table 7). Horizons with $F_{\text{MOP}} < 10$ are treated as weak-instrument horizons; for those horizons, inference is interpreted using the AR sets rather than relying solely on conventional Wald bands. For SLP–IV, when AR tests are used, the spline penalty is held fixed during refits in order to maintain comparability across bootstrap draws and instrument sets. This design ensures that weak-instrument robustness is handled consistently in specifications where the first stage is weak.

For ease of interpretation, the diagnostic tables report for each horizon the Montiel–Olea–Pflueger effective F statistic (a measure of first-stage strength, with values well above 10 indicating strong instruments), the partial R^2 of the instrument set in the first stage (the share of variation in Δg_t explained by $\mathbf{z}_t$, conditional on the controls $\mathbf{X}_t$), and, for the Ramey-only specification, AR confidence sets for the output response. Wide AR intervals that span both positive and negative values indicate that the sign and magnitude of the response are weakly identified; when the first stage is strong, the AR intervals tend to align closely with conventional Wald intervals.

B.4 First-Stage Diagnostics by Horizon

Table 5 reports first-stage diagnostics for the baseline AG specification, and Table 6 reports the corresponding results for the overlap designs with Ramey and Both–IV instruments. In the baseline sample the AG instrument delivers Montiel–Olea–Pflueger F statistics well above the weak-IV threshold and stable partial R^2, indicating a strong first stage in (4). On the overlap sample the Ramey-only specification is substantially weaker, while combining AG and Ramey in the Both–IV configuration strengthens identification but does not reach the AG benchmark. These patterns motivate the focus on AG-based estimates in the main analysis and the use of Ramey-only and Both–IV specifications as robustness checks.

Table 5: First-stage diagnostics by horizon (AG baseline).

AG (baseline)

h	MOP-F	partial R^2
0	199.07	0.758
1	199.19	0.758
2	194.86	0.757
3	194.14	0.757
4	192.66	0.754
5	187.77	0.748
6	186.20	0.748
7	185.69	0.751
8	177.14	0.746
9	178.36	0.746
10	163.44	0.745
11	163.55	0.745
12	164.45	0.746
13	167.64	0.750
14	152.72	0.717
15	148.14	0.716
16	149.07	0.716

Table 6: First-stage diagnostics by horizon (overlap instruments).

	Ramey-only			**Both–IV (AG+Ramey)**	
h	MOP-F	partial R^2	h	MOP-F	partial R^2
0	8.30	0.036	0	65.02	0.716
1	8.30	0.036	1	65.02	0.716
2	8.30	0.036	2	65.02	0.716
3	8.30	0.036	3	65.02	0.716
4	8.30	0.036	4	65.02	0.716
5	8.30	0.036	5	65.02	0.716
6	8.30	0.036	6	65.02	0.716
7	8.30	0.036	7	65.02	0.716
8	8.30	0.036	8	65.02	0.716
9	8.30	0.036	9	65.02	0.716
10	8.30	0.036	10	65.02	0.716
11	8.30	0.036	11	65.02	0.716
12	8.30	0.036	12	65.02	0.716
13	8.30	0.036	13	65.02	0.716
14	8.30	0.036	14	65.02	0.716
15	8.30	0.036	15	65.02	0.716
16	8.30	0.036	16	65.02	0.716

B.5 Over-Identification and Weak-IV Checks

Table 7 shows that, for the Ramey-only specification, the Anderson–Rubin confidence sets are typically wide and are often labelled "grid censored" in the note column (this label indicates that, when inverting the AR test, the numerical search reaches the edge of the grid used for the parameter space, so the reported intervals are truncated by the grid bounds). Combined with the low Montiel–Olea–Pflueger statistics in Table 6, these results confirm that the Ramey instrument delivers a weak first stage in this design. The AR sets span both negative and positive values at many horizons, and several horizons include zero; at the same time, the AR sets may exclude zero at particular horizons, reflecting sampling variation under weak identification. By contrast, the AG baseline exhibits strong first-stage diagnostics (Table 5), so weak-IV concerns are confined to the Ramey-only robustness specification rather than to the baseline estimates. The Hansen J test for the Both–IV overlap specification yields no strong evidence against the joint validity of AG and Ramey instruments on the overlap sample.

Table 7: Anderson–Rubin confidence sets by horizon (Ramey-only specification).

h	CI low	CI high	Note
0	-0.17	0.48	grid censored
1	-0.42	0.23	grid censored
2	-0.18	0.31	grid censored
3	-0.44	0.21	grid censored
4	-0.33	0.34	grid censored
5	-0.30	0.37	grid censored
6	-0.36	-0.03	grid censored
7	-0.34	0.34	grid censored
8	-0.03	0.34	grid censored
9	-0.30	0.30	grid censored
10	-0.27	0.43	grid censored
11	-0.32	0.37	grid censored
12	-0.31	0.39	grid censored
13	-0.46	0.24	grid censored
14	-0.29	0.38	grid censored
15	-0.40	0.12	grid censored
16	-0.34	0.39	grid censored

Table 8: Hansen J test (Both–IV overlap).

Statistic	Value
J	0.285
df	1
p	0.594

B.6 Bootstrap Robustness

Bootstrap inference uses circular moving blocks with block length $\ell = 6$ and $B = 999$ draws. The same procedure is applied both to the baseline horizon window $H = 16$ and to the shorter robustness window $H = 12$. The bootstrap confidence intervals for output multipliers and constant-parameter comparisons reported in Section 6 are therefore based throughout on the $\ell = 6$ specification; the same block length and number of draws are used for the constant-parameter designs (LP–IV and SLP–IV). For TVP–SLP–IV, uncertainty bands are constructed from Durbin–Koopman simulation-smoother draws as described in Section 4.3.

B.7 Overlap Multipliers by Instrument

Table 9 reports multipliers at $H = 8$ and $H = 16$ for AG-only, Ramey-only, and Both–IV specifications on the overlap sample. The AG and Both–IV multipliers are similar in magnitude and lie within each other's confidence intervals, whereas Ramey-only estimates are more volatile, with very wide intervals that span a broad range of values. These comparisons reinforce the conclusion that the baseline findings are driven by the AG shocks and that incorporating Ramey as an additional instrument does not overturn the main results. In view of weak-IV diagnostics for Ramey-only, emphasis is placed on the shorter-horizon summaries and on weak-IV-robust inference when interpreting the overlap robustness results.

Table 9: Cumulative dollar multipliers by instrument (overlap sample).

	Horizon 8			Horizon 16		
Method/Regime	Point	68% CI	90% CI	Point	68% CI	90% CI
LP–IV (AG overlap)	0.40	[-0.60, 1.20]	[-1.19, 1.97]	1.06	[-0.91, 1.88]	[-1.73, 3.40]
LP–IV (Ramey sample)	-0.44	[-1.39, 4.90]	[-4.51, 13.23]	-0.54	[-1.00, 5.28]	[-8.47, 14.78]
LP–IV (Both–IV overlap)	0.35	[-0.60, 1.18]	[-1.16, 1.96]	0.98	[-0.91, 1.84]	[-1.80, 3.29]

Notes: This table reports cumulative dollar multipliers at $H = 8$ and $H = 16$ on the AG–Ramey overlap sample (1985Q1–2015Q4) for three specifications: AG-only, Ramey-only, and Both–IV (AG+Ramey). Intervals are percentile bootstrap bands constructed using the baseline moving-block design.

B.8 TVP–SLP–IV Robustness: Alternative Instruments

For completeness, this appendix reports the full TVP–SLP–IV grids for alternative instruments. The main text shows the baseline grid under the Auerbach–Gorodnichenko (AG) instrument (split by the corresponding regimes); Figures 3–4–5–6 display the corresponding TVP impulse responses for the AG (without regimes), Ramey-only, Both–IV, and AG-overlap specifications. These grids are used as instrument-robustness checks.

B.9 Regime Windows, Horizon, and Covid-Exclusion Diagnostics

This section complements the main robustness discussion by summarising how regime-dependent results relate to horizon choices and to the short Covid and post-Covid windows and by documenting the quantitative diagnostics underlying the horizon and window checks referenced in Section 6.

First, the TVP–SLP–IV regime averages reported in the main text are computed over the fixed windows defined in Section 3.3. The TVP model is estimated on the full 1985Q1–2023Q4 sample and delivers regime-averaged impulse responses for all five monetary-policy regimes. At long horizons (e.g. $H = 16$), the short windows (ZLB–Covid and Positive-rate (Post-Covid) contain limited effective support for tracing shocks over the full horizon length, and regime-multiplier entries may therefore be treated as not identified and in summary tables. The main multiplier comparisons focus on the three long pre-Covid regimes, where time spans are sufficient for reliable identification.

Second, the LP–IV horizon diagnostics examine whether the medium-run output response is sensitive to the choice of terminal horizon. Figure 7 displays the LP–IV impulse-response panel for the structural 1985Q1–2019Q4 sample, overlaying the baseline $H = 16$ specification and a shorter-horizon run with $H \leq 12$. The two sets of point estimates lie close to one another over the first few years, and the principal effect of reducing the horizon cap is to remove oscillations and additional noise at the final horizons. Trimming the terminal horizon does not introduce systematic changes in the profile over business-cycle frequencies.

Third, Covid-exclusion diagnostics examine sensitivity in an extended-sample constant-parameter LP–IV specification estimated on 1985Q1–2023Q4 to the inclusion of the Covid block. These diagnostics are distinct from the structural baseline window (1985Q1–2019Q4), which excludes Covid observations by construction. Table 10 reports the LP–IV cumulative output response for the AG instrument on the Covid-exclusion window obtained by excluding 2020Q2–2021Q4 from the 1985Q1–2023Q4 sample, with horizons capped at $H \leq 12$. Relative to the structural-sample LP–IV results, the Covid-exclusion point estimates remain of similar order of magnitude at comparable horizons and display a similar pattern without an additional systematic change in sign or shape over business-cycle frequencies. The associated confidence intervals are broadly comparable to, and in some cases wider than, those in the structural-window specification, consistent with a reduced effective sample once the Covid block is removed. Taken together with the baseline multiplier estimates in Table 1, these results indicate that excluding the Covid period in the extended sample does not overturn the qualitative conclusion of a modestly positive medium-run output multiplier.

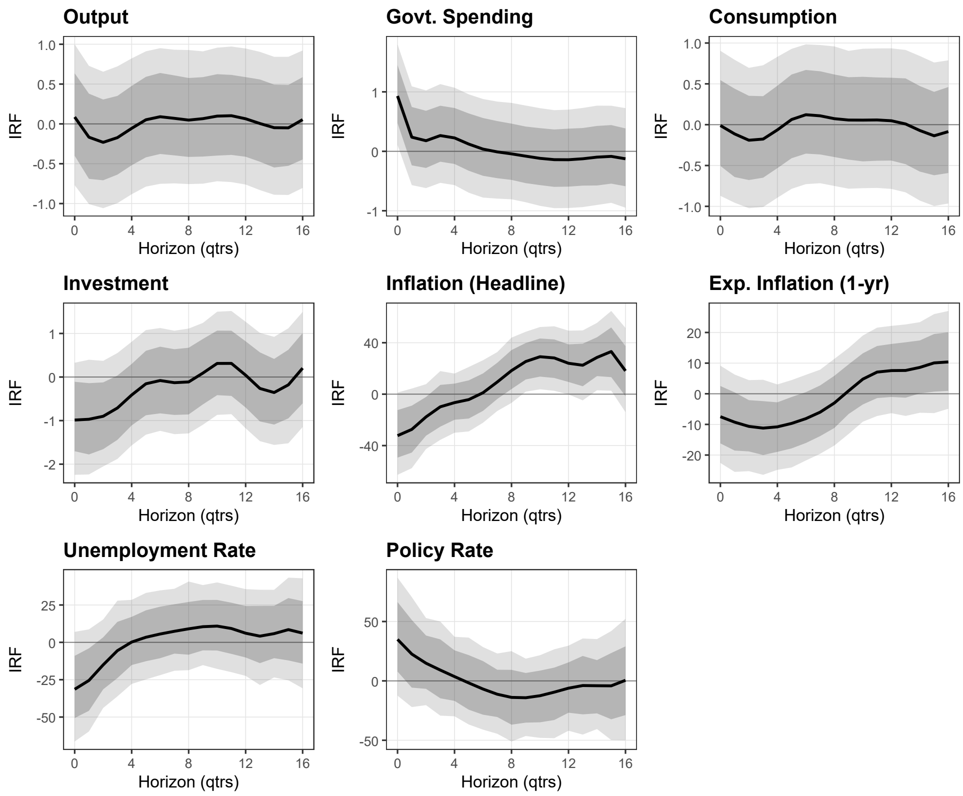


Figure 3: TVP–SLP–IV impulse-response grid with the AG instrument (1985Q1–2023Q4).

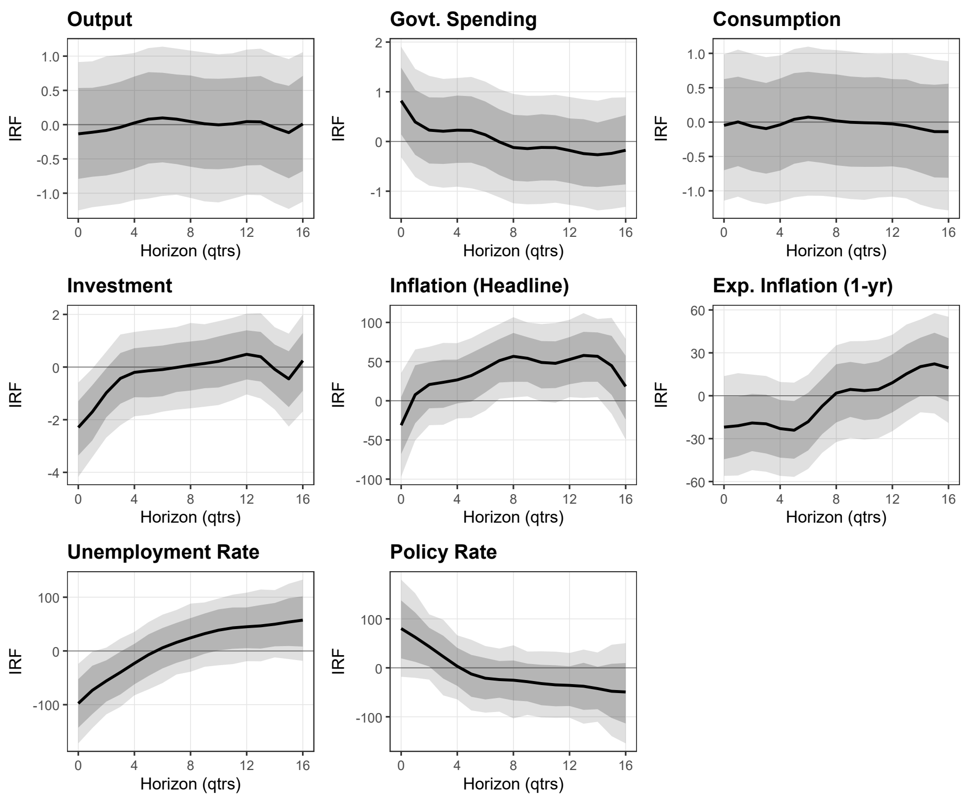


Figure 4: TVP–SLP–IV impulse-response grid with the Ramey instrument (1985Q1–2023Q4).

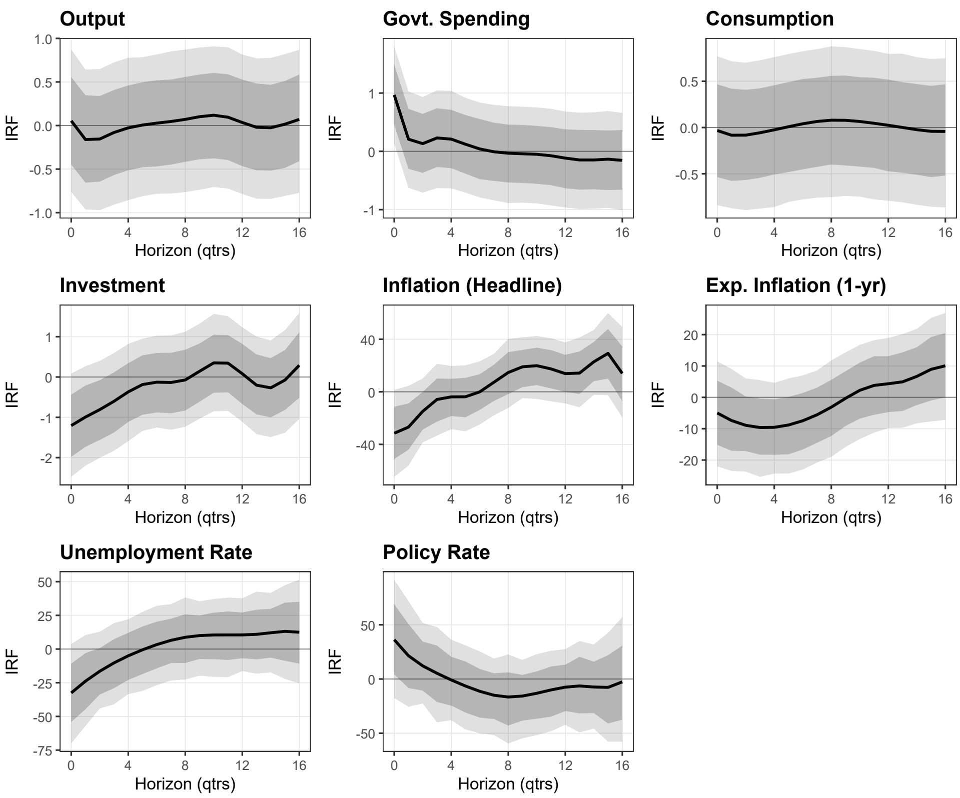


Figure 5: TVP–SLP–IV impulse-response grid with the Both–IV (AG+Ramey) specification.

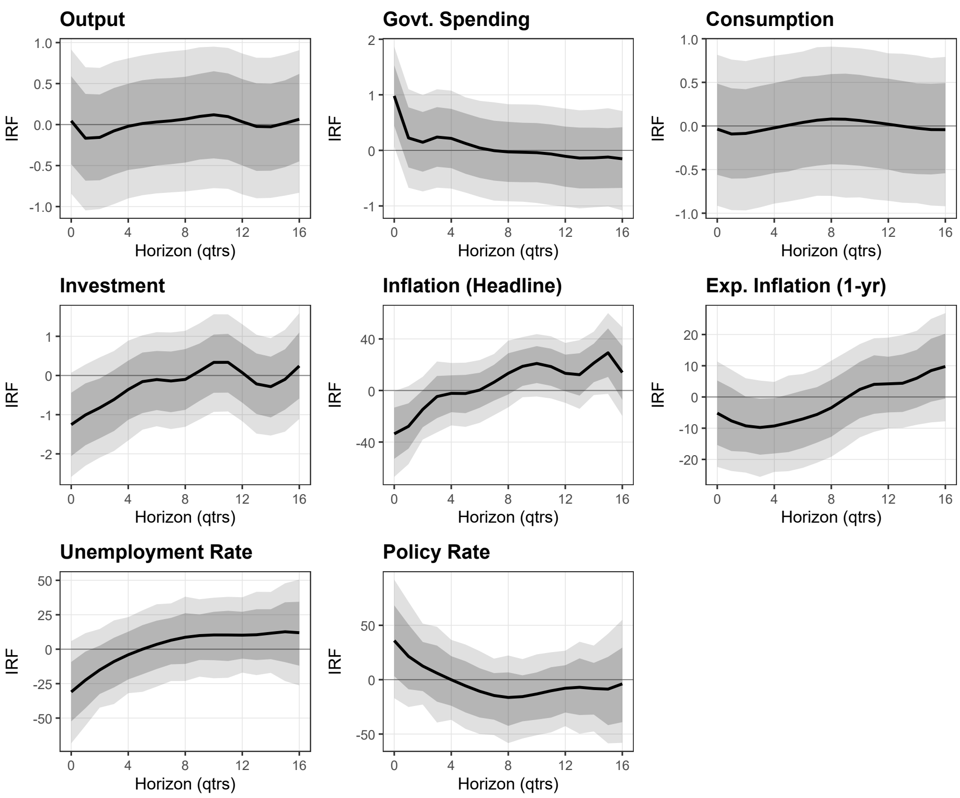


Figure 6: TVP–SLP–IV impulse-response grid on the AG–Ramey overlap sample.

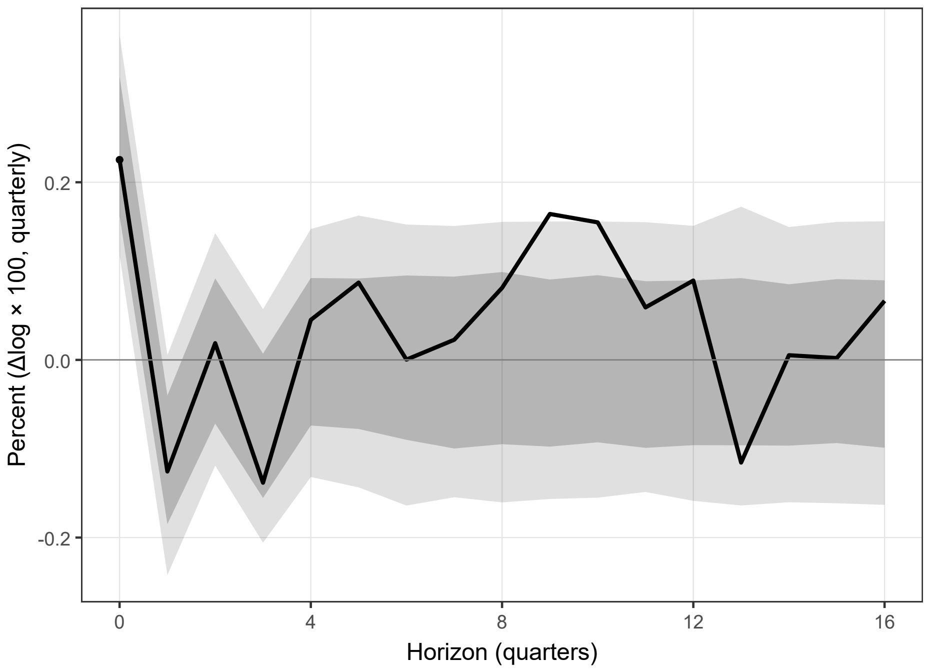


Figure 7: LP–IV output response with the Auerbach–Gorodnichenko instrument on the structural 1985Q1–2019Q4 sample, horizon cap $H = 16$. The figure underlies the horizon diagnostics discussed in the text.

Table 10: Covid-exclusion LP–IV cumulative output responses (AG instrument, horizon $H \leq 12$).

Horizon h	Point estimate	Lower CI	Upper CI
0	0.2472	0.1301	0.3643
1	-0.1079	-0.2124	-0.0035
2	0.0170	-0.0953	0.1292
3	-0.1375	-0.2523	-0.0227
4	0.0170	-0.0997	0.1336
5	0.0704	-0.0358	0.1765
6	-0.0099	-0.1262	0.1064
7	0.0072	-0.1072	0.1215
8	0.0601	-0.0753	0.1955
9	0.1238	-0.0255	0.2731
10	0.1522	0.0363	0.2681
11	0.0568	-0.0696	0.1832
12	0.1016	0.0216	0.1816

Notes: Headline LP–IV specification for output, using the Auerbach–Gorodnichenko instrument, estimated on the 1985Q1–2023Q4 sample excluding 2020Q2–2021Q4. Horizons are capped at $H = 12$ to focus on the medium run. Confidence intervals are percentile bootstrap bands constructed as in the baseline specification.

C Tuning and Sensitivity

This Appendix documents parameter choices for the SLP and TVP estimators. The objective is to make explicit (i) how smoothness penalties and discount factors are selected, (ii) how uncertainty bands are constructed, and (iii) which "nearby" specifications are used for robustness checks.

Throughout this Appendix, the identification strategy and the underlying IV shock series remain the same as in the main text; only the regularisation (smoothing) and time-variation hyperparameters are varied. Three distinct objects should be kept separate:

- **LP–IV:** estimates a separate coefficient at each horizon (no smoothing across horizons).
- **SLP–IV:** estimates a smooth curve of coefficients across horizons by representing the impulse response on a spline basis and penalising excessive curvature. The tuning parameter λ controls smoothness across horizon.
- **TVP–SLP–IV:** allows the spline coefficients (and control coefficients) to drift over time in a state-space model. The discount factor δ controls smoothness over time.

Uncertainty is computed differently across estimators: constant-parameter LP–IV and SLP–IV use a moving-block bootstrap, whereas TVP–SLP–IV uses Durbin–Koopman simulation-smoother draws from the state-space model.

This Appendix provides the implementation details referenced in Sections 4.2 and 4.3. The main text retains the identification, core specification, and intuition of SLP–IV and TVP–SLP–IV while relocating the mechanics of spline construction, roughness penalties, and cross-validation here for brevity.

Sample windows (baseline versus diagnostics).

The structural constant-parameter baseline is defined on the 1985Q1–2019Q4 window (Covid excluded by construction). Extended-sample diagnostics (including the λ and knot sweeps) are run on 1985Q1–2023Q4 and are interpreted as sensitivity checks; they do not redefine the structural baseline. The TVP–SLP–IV model is estimated once per outcome on 1985Q1–2023Q4, and regime-specific objects are constructed only by averaging the resulting time-varying responses over fixed regime windows (the model is not re-estimated separately by regime).

The quantitative summaries in Section 6 are based on the following pre-specified variations, which are documented in detail below and in Appendix B:

- **Alternative instruments:** AG-only baseline, Ramey-only, and Both–IV (AG+Ramey) specifications on the AG–Ramey overlap sample, with diagnostics and multipliers reported in Appendix B.
- **Horizon set:** baseline local projections and SLP–IV with $H = 16$ compared with a shorter $H \leq 12$ run to check that medium-run shapes are not driven by the last few horizons, with the diagnostics reported in Appendix B.
- **Covid-exclusion window:** extended-sample diagnostic re-estimation of constant-parameter LP–IV specifications on 1985Q1–2023Q4 dropping 2020Q2–2021Q4, with horizons capped at $H \leq 12$, documented in Appendix B. This diagnostic is distinct from the structural baseline 1985Q1–2019Q4 window, which excludes Covid observations by construction.
- **SLP parameters:** cubic B-splines with 5 interior knots in the baseline and robustness variations with $\{3, 7\}$ interior knots, combined with penalty weights on the grid $\{0.5, 1, 2\} \times \lambda_{\text{CV}}$.
- **TVP persistence parameter:** δ^* selected over the baseline grid $\{0.97, 0.98, 0.99, 0.995\}$ using a one-step-ahead predictive score from the Kalman filter, with robustness checks on $\{0.97, 0.99, 0.995\}$.
- **Bootstrap settings:** circular moving-block bootstrap with $B = 999$ and block length $\ell = 6$ in the baseline, and robustness checks for $\ell \in \{4, 6, 8\}$.

These dimensions define the neighbourhood of specifications considered when assessing the robustness of the main findings.

C.1 Smooth Local Projections (SLP)

This subsection documents the spline basis, roughness penalty, and λ selection procedure underlying the SLP–IV results in the main text.

Table 11: SLP parameters, baseline values, and rationale (part 1).

Component	**Baseline value**	**Rationale**	**Robustness**
Basis	Cubic B-splines (degree $= 3$)	Smooth but flexible responses across horizons.	—
Interior knots	5 interior knots on $[0, 16]$	Flexibility set by knots; variance controlled by λ.	$\{3, 7\}$
Smoothed parameters	$\beta(h)$ only (fiscal shock)	Identification remains transparent; controls remain horizon-specific.	—
Penalty	Second differences on spline coefficients	Shrinks high-frequency "wiggles" while preserving medium-run curvature.	fixed at order 2
Penalty weight λ	CV-based; use $0.8 \times \lambda_{\text{CV}}$ for inference	Modest undersmoothing tends to improve interval coverage in practice.	$\{0.5, 1, 2\} \times \lambda_{\text{CV}}$

Table 12: SLP parameters, baseline values, and rationale (part 2).

Component	**Baseline value**	**Rationale**	**Robustness**
Tail penalty	Off in baseline (optional light ridge on $\beta(H)$)	Limits long-horizon drift without forcing endpoints to zero.	enable only if diagnostics indicate drift
Horizon set	$H = 16$ (check $H \leq 12$)	Captures medium-run dynamics; long horizons can be noisy.	report both $H = 16$ and $H \leq 12$
Lags	$p = 4$	Standard quarterly lag structure.	$p \in \{2, 4\}$
Bootstrap	Circular MBB, $B = 999$, $\ell = 6$, refit spline with fixed λ each draw	Preserves serial dependence while resampling the "base quarter" units.	$\ell \in \{4, 6, 8\}$
Instrument integration	Plug-in $\Delta\widehat{g}_t$ from first stage	Keeps the shock regressor identical to LP–IV in the baseline.	AG / Ramey / Both–IV overlap

Notes: CV = cross-validation; MBB = moving-block bootstrap.

Tables 11–12 summarise the SLP hyperparameters. The baseline specification uses cubic B-splines with 5 interior knots on the horizon grid $\{0, \ldots, 16\}$ to approximate the impulse response and penalises variation in the response across horizons to remove high-frequency noise. Cross-validation selects the strength of the smoothness penalty, and inference is carried out at a slightly undersmoothed value ($0.8 \times \lambda_{\text{CV}}$) to improve coverage of bootstrap intervals. The robustness column lists alternative choices for knots, lag length, and bootstrap block length that define a neighbourhood of specifications around the baseline. The figures and multiplier tables in the main text are based on the baseline configuration.

Basis and penalty implementation.

Shock notation. Throughout Appendix C, $\Delta\widehat{g}_t$ denotes the first-stage fitted innovation in government spending growth (the external-instrument shock regressor in Section 4.1).

Spline representation (what is being smoothed). In an LP–IV, a separate coefficient β_h is estimated at each horizon h. In SLP–IV, the horizon-indexed coefficient is instead treated as a smooth function $\beta(h)$. Let $h = 0, \ldots, H$ index horizons and define a cubic B-spline basis

$$\mathbf{B}(h) = \big(b_1(h), \ldots, b_K(h)\big)^\top \in \mathbb{R}^K$$

constructed on the horizon interval $[0, H]$ using standard endpoint augmentation for cubic splines (including an intercept term in the spline basis). The fiscal response at horizon h is parametrised as

$$\beta(h) \;=\; \mathbf{B}(h)^\top \mathbf{c},$$

where $\mathbf{c} \in \mathbb{R}^K$ is the vector of spline coefficients specific to the fiscal shock. These constant-parameter coefficients are the time-invariant analogue of the TVP spline coefficients $\mathbf{b}_t$ in Section 4.3, where $\beta_t(h) = \mathbf{B}(h)^\top \mathbf{b}_t$.

Stacked design (how horizons are combined). For each base quarter t in the structural sample, define the horizon-stacked outcome vector

$$\mathbf{y}_t \equiv \big(y_{t+0}, \ldots, y_{t+H}\big)^\top \in \mathbb{R}^{H+1}.$$

Collect the spline basis across horizons in the $(H+1) \times K$ matrix

$$\mathbf{B}_{0:H} \equiv \big(\mathbf{B}(0), \ldots, \mathbf{B}(H)\big)^\top,$$

and define the shock-design block

$$\mathbf{X}_{c,t} \equiv (\Delta \widehat{g}_t)\, \mathbf{B}_{0:H},$$

so that $(\mathbf{X}_{c,t}\mathbf{c})_h = \beta(h)\, \Delta \widehat{g}_t$ for $h = 0, \ldots, H$. Let $\mathbf{X}_{\text{ctrl},t}$ collect the horizon-specific control blocks (including horizon dummies and lagged controls), with coefficients that are unrestricted across horizons.

Stacking over base quarters yields

$$\mathbf{y} \equiv (\mathbf{y}_1^\top, \ldots, \mathbf{y}_T^\top)^\top, \qquad \mathbf{X}_c \equiv (\mathbf{X}_{c,1}^\top, \ldots, \mathbf{X}_{c,T}^\top)^\top, \qquad \mathbf{X}_{\text{ctrl}} \equiv (\mathbf{X}_{\text{ctrl},1}^\top, \ldots, \mathbf{X}_{\text{ctrl},T}^\top)^\top,$$

and the stacked regression $\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$ with $\mathbf{X} = [\mathbf{X}_c, \mathbf{X}_{\text{ctrl}}]$. The coefficient vector is defined as the column-stack

$$\boldsymbol{\theta} \equiv \begin{pmatrix} \mathbf{c} \\ \boldsymbol{\Gamma} \end{pmatrix}, \qquad \boldsymbol{\Gamma} \equiv (\Gamma_0^\top, \ldots, \Gamma_H^\top)^\top,$$

where Γ_h denotes the horizon-specific control coefficients (unrestricted across horizons, as in the main-text SLP specification). This corresponds to the subset-smoothing case in Barnichon and Brownlees (2019): only $\beta(h)$ is represented on the spline basis and penalised, while control coefficients remain unrestricted across horizons.

Unbalanced stacked sample across horizons. In implementation the stacked design is unbalanced across horizons: for each h the regression uses base quarters t such that y_{t+h} is observed, so the effective sample size declines with h. The notation below suppresses this dependence for readability.

Roughness penalty (how wiggles are discouraged). The roughness penalty is implemented as a quadratic form in $\mathbf{c}$,

$$\mathbf{c}^\top \mathbf{P}\mathbf{c}, \qquad \mathbf{P} = \mathbf{D}^\top \mathbf{D},$$

where $\mathbf{D}$ is the second-difference matrix applied to the K-vector $\mathbf{c}$. This is a standard discrete approximation to a second-derivative penalty: large entries of $\mathbf{Dc}$ correspond to sharp changes in the slope of $\beta(h)$ between adjacent horizons. With penalty order $d = 2$, the regularisation shrinks the impulse response toward a first-order polynomial in h (a linear profile). In the baseline, the penalty order is fixed at 2.

An optional tail penalty can be added, of the form

$$\tau\,[\beta(H)]^2 = \tau\,\mathbf{c}^\top \mathbf{B}(H)\mathbf{B}(H)^\top \mathbf{c}, \qquad \tau > 0,$$

to gently shrink the terminal horizon toward zero if diagnostics indicate excessive long-horizon drift. This tail penalty is set to zero in all baseline runs and is only considered in ad hoc checks.

Penalised estimator and λ selection.

Given the stacked regression $\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \varepsilon$, SLP–IV solves

$$\hat{\boldsymbol{\theta}}(\lambda) = \arg\min_{\boldsymbol{\theta}} \Big\{ (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\theta}) + \lambda\,\mathbf{c}^\top \mathbf{P}\mathbf{c} \Big\},$$

so that only the fiscal-shock spline coefficients $\mathbf{c}$ enter the penalty; the control coefficients $\{\Gamma_h\}_{h=0}^{H}$ remain fully flexible across horizons. Writing $\mathbf{X} = [\mathbf{X}_c, \mathbf{X}_{\text{ctrl}}]$ and letting $\mathbf{P}_{\text{full}}$ denote the block-diagonal extension of $\mathbf{P}$ with zeros in the control block, the normal equations take the form

$$(\mathbf{X}^\top \mathbf{X} + \lambda \mathbf{P}_{\text{full}})\,\hat{\boldsymbol{\theta}}(\lambda) = \mathbf{X}^\top \mathbf{y},$$

which are solved numerically.

Contiguous time-series cross-validation (what the CV score measures). The smoothing parameter λ is chosen by contiguous time-series cross-validation over base quarters. For baseline SLP–IV, $t = 1, \ldots, T$ indexes the base quarters in the structural 1985Q1–2019Q4 sample. For tuning diagnostics reported on the full 1985Q1–2023Q4 window (e.g., the SLP sweep table in Section C.3), the same procedure is applied with t indexing base quarters in that extended sample.

The set $\{1, \ldots, T\}$ is partitioned into $K_{\text{fold}} = 5$ contiguous time blocks (folds). For a given candidate λ, the estimator is refit K_{fold} times, each time leaving out one fold and computing the mean squared prediction error on the held-out base quarters. In each fold, all stacked rows associated with the held-out base quarters (i.e., all horizons $h = 0, \ldots, H$ for those t) are excluded from estimation and used only for prediction. The cross-validation score for λ is the average of these prediction errors across folds. The candidate grid is taken on a log scale, and the minimiser is denoted λ_{CV}. In rare cases where the cross-validation surface is numerically degenerate, λ_{CV} is set to the median grid value.

Baseline undersmoothing (how λ is set for inference). In the baseline output specification, the cross-validation profile in Figure 8 is relatively flat in a neighbourhood of λ_{CV}, and neighbouring points on the candidate grid deliver very similar prediction errors. This implies that moderate perturbations of λ are unlikely to materially affect the estimated impulse-response shape. The baseline SLP specification therefore uses modest undersmoothing relative to the prediction-optimal value by setting

$$\lambda \;=\; \max\left\{0.8\,\lambda_{\text{CV}},\; \lambda_{\min}\right\},$$

where $\lambda_{\min}$ is the minimum value on the candidate grid. This λ is then held fixed both in the reported SLP estimates and in the bootstrap. The robustness range $\{0.5, 1, 2\} \times \lambda_{\text{CV}}$ in Table 11 defines a natural grid of nearby penalty strengths.

Bootstrap and multipliers (constant-parameter SLP–IV).

For SLP–IV, uncertainty bands in the main figures and tables are based on a circular moving-block bootstrap over base quarters. Let $t = 1, \ldots, T$ index distinct base quarters in the structural 1985Q1–2019Q4 sample. Blocks of length ℓ are drawn from $\{1, \ldots, T\}$ with replacement and wrapped around the endpoints, yielding a bootstrap sequence of quarter indices of length T. For each such sequence, all stacked rows corresponding to those base quarters are included (so all horizons associated with a given base quarter move together). On this resampled stacked dataset, the penalised estimator is refit with the spline basis, penalty matrix, and smoothing parameter λ fixed at their baseline values. Repeating this procedure B times yields a collection of bootstrap draws of the response $\beta^{(b)}(h)$, from which 68% and 90% percentile bands are constructed horizon-by-horizon. Block lengths $\ell \in \{4, 6, 8\}$ are considered; $\ell = 6$ is the baseline.

Cumulative multipliers are computed both in levels (using the sample-average real per-capita output and spending) and within each bootstrap draw using the same log-to-level conversion, so that multiplier confidence intervals reflect sampling uncertainty in both the numerator and the denominator conditional on the smoothing parameter. Uncertainty about the choice of λ itself is not bootstrapped; instead, the undersmoothing scheme described above is used to improve coverage.

C.2 Time-Varying Parameters (TVP)

The TVP–SLP–IV specification is implemented as a linear Gaussian state-space model in which the spline coefficients governing the impulse-response curve and the coefficients on controls evolve according to a diagonal AR(1) transition with persistence parameter $\delta \in (0, 1)$. The parameter δ is tuned outcome by outcome on the full 1985Q1–2023Q4 TVP estimation sample using a one-step-ahead predictive score computed from the Kalman filter.

Tuning of δ (interpretation of the predictive score).

For each candidate δ on a discrete grid (baseline grid $\{0.97, 0.98, 0.99, 0.995\}$), the TVP state-space model is constructed and one-step-ahead forecast errors ("innovations") from the Kalman filter are obtained. These innovations measure the discrepancy between observed outcomes and model-implied one-step-ahead predictions. A predictive log-likelihood score is then computed from the innovations and their forecast error covariance matrices. Concretely, if $v_t(\delta)$ denotes the one-step-ahead innovation and $F_t(\delta)$ its covariance, the (Gaussian) contribution is proportional to

$$-\tfrac{1}{2}\Big(\log|F_t(\delta)| + v_t(\delta)^\top F_t(\delta)^{-1} v_t(\delta) \Big),$$

and the predictive score sums (or averages) these contributions over time.

To reduce sensitivity to short-lived local changes in fit, the innovation sequence is partitioned into $K_{\text{fold}} = 5$ contiguous time blocks and the predictive score is averaged across blocks. The selected value δ^* maximises this average predictive score. Robustness checks inspect nearby values on the grid $\{0.97, 0.99, 0.995\}$.

When numerical issues render some candidate values infeasible (e.g., due to non-invertible covariance matrices), δ^* is chosen among the remaining candidates with finite predictive scores. In rare cases in which all predictive scores are undefined, the selection falls back to the median of the candidate grid. A single persistence parameter is imposed for all states (spline and control coefficients) for a given outcome, and δ is not re-tuned by regime.

Table 13: TVP parameters, baseline values, and rationale.

Component	**Baseline value**	**Rationale**	**Robustness**
Persistence δ	$\delta^* \in \{0.97, 0.98, 0.99, 0.995\}$ (selected by predictive score)	Controls how gradually coefficients drift over time.	Grid $\{0.97, 0.99, 0.995\}$ and inspection of predictive scores
State evolution	Diagonal AR(1) for spline and control coefficients	Allows gradual time variation while discouraging abrupt changes.	Not varied beyond δ grid
Band construction	Durbin–Koopman simulation-smoother percentiles	Provides state-space-consistent uncertainty bands.	No bootstrap alternative; robustness judged via δ variations
Regime averaging	Fixed ZLB and Covid windows from Section 3.3	Ensures consistent regime definitions across outcomes and specifications.	Windows held fixed; short regimes interpreted as descriptive

Notes: IRF = impulse response function. Regime objects are formed by averaging time-varying IRFs; the TVP model is not re-estimated by regime.

State evolution and initialisation (what δ does mechanically).

Let $\boldsymbol{\alpha}_t = (\mathbf{b}_t', \boldsymbol{\Gamma}_t')'$ collect the time-varying spline coefficients $\mathbf{b}_t$ and the time-varying coefficients on controls $\boldsymbol{\Gamma}_t$. Time variation is modelled as

$$\boldsymbol{\alpha}_t = T(\delta)\,\boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim \mathcal{N}\big(0, Q(\delta)\big),$$

where $T(\delta)$ is diagonal and applies the same persistence δ to all state elements for a given outcome. The covariance matrix $Q(\delta)$ is diagonal and, in implementation, its diagonal elements are set proportional to $1 - \delta^2$ (subject to small numerical floors and ridges to ensure positive definiteness). Under this parameterisation, values of δ closer to one imply more persistent states and smoother time variation, whereas smaller values imply less persistence and relatively larger innovations.

The state vector is initialised with a diffuse Gaussian prior (large diagonal prior variance) and zero prior mean for all state elements. The diffuse prior implies that early observations quickly dominate the initial conditions. The TVP layer remains comparable to the constant-parameter baseline because the spline basis, control set, and IV design match the SLP specification; deviations from the baseline are governed by the time variation induced by δ rather than by redefinition of the shock or controls.

Estimation, draws, and regime averaging (post-processing, not re-estimation).

For each outcome, the linear Gaussian state-space model is estimated by Kalman filtering and smoothing. Uncertainty bands are constructed from Durbin–Koopman simulation-smoother full-path draws of the state vector. For each draw, spline coefficients are mapped into horizon-specific impulse responses at every date t. These date-specific IRFs are then averaged over the regime windows defined in Section 3.3. Regime-specific point estimates correspond to the median across draws, and 68% and 90% bands are obtained from the corresponding empirical percentiles. These TVP bands are fully model-based and do not rely on bootstrap resampling.

C.3 SLP–IV λ and Knot Sweeps

This subsection reports the SLP–IV penalty and knot-sweep diagnostics for the AG design. These diagnostics are generated on the full 1985Q1–2023Q4 sample under the AG instrument and are interpreted as sensitivity checks; the structural baseline SLP–IV specification used in the main text remains defined on 1985Q1–2019Q4.

Figure 8 displays the five-fold contiguous-block cross-validation criterion as a function of the smoothness-penalty weight λ for the AG design at $H = 16$ (using $K = 5$ interior knots). The profile exhibits a pronounced improvement in predictive fit when moving from very low penalty weights toward modest regularisation, consistent with the penalty removing high-frequency sampling noise that harms out-of-sample performance. Beyond this initial region, the curve becomes nearly flat over a wide range of λ values, indicating that many penalty choices deliver essentially indistinguishable cross-validated losses. The minimiser λ_{CV} therefore lies on a broad plateau rather than at a sharply identified optimum, implying that moderate perturbations of λ are unlikely to materially affect the estimated impulse-response shape. This near-degeneracy motivates reporting robustness checks over $\{0.5, 1, 2\} \times \lambda_{\text{CV}}$ and adopting modest undersmoothing ($0.8 \times \lambda_{\text{CV}}$) for inference, since predictive performance is largely unchanged while slightly weaker regularisation reduces attenuation of medium-run curvature.

Table 14: SLP–IV λ and knot sweeps for the AG design (output, $H = 16$).

Inst.	Var.	H	λ_{CV}	Factor	λ	Knots	K_{fold}	$M(8)$	$M(16)$
ag	dy	16	0.2154	0.5	0.0862	3	5	-0.3680	0.7611
ag	dy	16	0.2154	1.0	0.1724	3	5	-0.3649	0.7652
ag	dy	16	0.2154	2.0	0.3447	3	5	-0.3695	0.7704
ag	dy	16	0.4642	0.5	0.1857	5	5	-0.3950	0.7594
ag	dy	16	0.4642	1.0	0.3713	5	5	-0.3883	0.7639
ag	dy	16	0.4642	2.0	0.7427	5	5	-0.3889	0.7695
ag	dy	16	2.1544	0.5	0.8618	7	5	-0.3963	0.7666
ag	dy	16	2.1544	1.0	1.7235	7	5	-0.3975	0.7728
ag	dy	16	2.1544	2.0	3.4471	7	5	-0.4066	0.7799

Notes: Each row corresponds to a combination of cross-validated penalty λ_{CV}, multiplicative factor, and number of interior knots. The column K_{fold} reports the number of contiguous time blocks used in cross-validation ($K_{\text{fold}} = 5$ throughout); the spline basis is cubic (degree 3) in all rows. The columns $M(8)$ and $M(16)$ report the 8- and 16-quarter cumulative output multipliers implied by the corresponding SLP–IV specification, constructed as in Appendix C. Across the grid, these multipliers vary only modestly, indicating that the reported SLP impulse-response shapes are not sensitive to moderate changes in smoothing and knot placement.

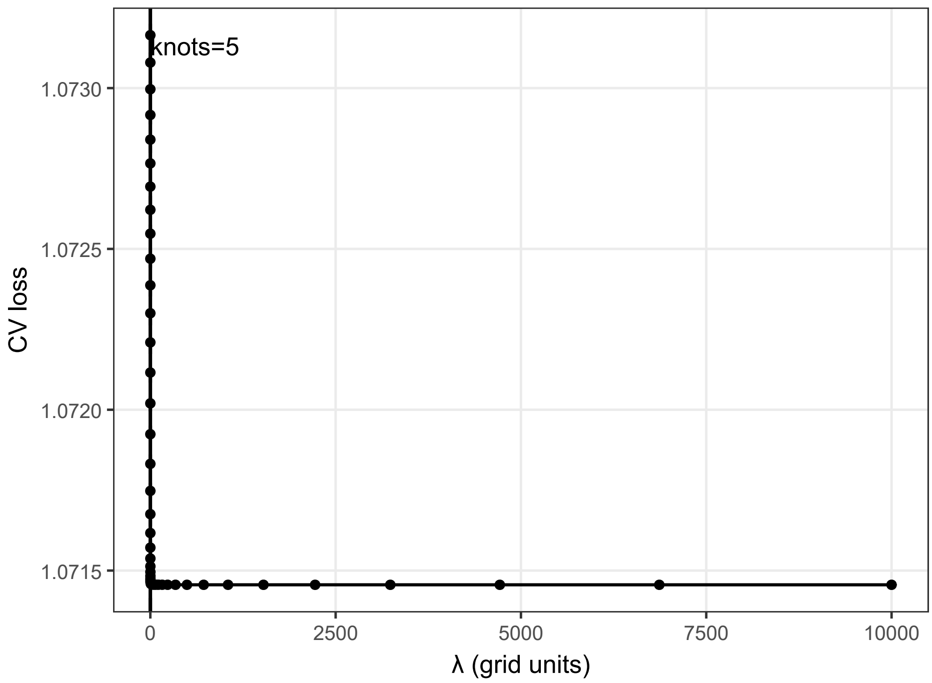


Figure 8: SLP–IV cross-validation profile and candidate λ grid for the AG design (output, horizon $H = 16$).

C.4 Multiplier Implementation Across Specifications

This section records how the estimation pipeline constructs multipliers for the different specifications used in the paper.

Constant-parameter LP–IV and SLP–IV (dollar multipliers).

For each specification and horizon h, the LP–IV and SLP–IV estimators deliver impulse responses β_h^Y and β_h^G for the transformed outcomes $\Delta \log Y$ and $\Delta \log G$. Let Y_0 and G_0 denote the sample means of the corresponding real per-capita levels over the relevant sample window (structural baseline, overlap window, or full sample, depending on the table). In the pipeline the level responses used to form multipliers are

$$\Delta Y_h \equiv Y_0\big(e^{\beta_h^Y} - 1\big), \qquad \Delta G_h \equiv G_0\big(e^{\beta_h^G} - 1\big),$$

so that the horizon-H multiplier reported in the baseline SLP–IV tables is

$$M(H) \equiv \frac{\sum_{h=0}^{H} \Delta Y_h}{\sum_{h=0}^{H} \Delta G_h}, \qquad H \in \{8, 16\}.$$

The same log-to-level mapping is applied within each bootstrap draw using the same Y_0 and G_0, so that multiplier confidence intervals reflect uncertainty in both numerator and denominator conditional on the smoothing parameter.

TVP–SLP–IV regime multipliers (log-response multipliers).

For each outcome and each state draw s from the simulation smoother, the TVP layer delivers time-varying impulse responses $\beta_t^{Y,(s)}(h)$ and $\beta_t^{G,(s)}(h)$ for $\Delta \log Y$ and $\Delta \log G$ at every date t and horizon h. Given the regime windows in Section 3.3, the regime-r average log responses for draw s are

$$\bar{\beta}_{r,h}^{Y,(s)} \equiv \frac{1}{|\mathcal{T}_r|} \sum_{t \in \mathcal{T}_r} \beta_t^{Y,(s)}(h), \qquad \bar{\beta}_{r,h}^{G,(s)} \equiv \frac{1}{|\mathcal{T}_r|} \sum_{t \in \mathcal{T}_r} \beta_t^{G,(s)}(h),$$

where $\mathcal{T}_r$ denotes the set of quarters belonging to regime r.

Regime-specific multipliers are constructed as unit-free ratios of cumulative log responses:

$$M_r^{(s)}(H) \equiv \frac{\displaystyle\sum_{h=0}^{H} \bar{\beta}_{r,h}^{Y,(s)}}{\displaystyle\sum_{h=0}^{H} \bar{\beta}_{r,h}^{G,(s)}}, \qquad H \in \{8, 16\}.$$

These objects are referred to as log-response multipliers. This construction avoids introducing additional regime-specific base levels into the definition of regime multipliers. When responses are small (as is typical for quarterly log responses) and base levels evolve gradually, log-response multipliers are numerically close to the per-capita dollar multipliers that would be obtained by mapping log responses into levels using regime-specific base levels $Y_0^{(r)}$ and $G_0^{(r)}$.

The pipeline stores the draws $\{M_r^{(s)}(H)\}_s$ as `multipliers_tvp_regimes`. When regime-level effects are discussed in the main text, point estimates correspond to the median of this distribution for each regime and horizon, and the reported bands reflect the corresponding empirical percentiles. As noted in the main text, $H = 16$ multipliers for the shortest regimes are interpreted cautiously and are not used as core quantitative evidence.

Mechanism decompositions.

When decomposing the output response into components (consumption, investment, and $Y - G$) in the mechanisms tables, the pipeline uses a first-order approximation to map log responses into level changes,

$$\Delta X_h \approx X_0 \, \beta_h^X,$$

where X_0 is the sample mean of the corresponding real per-capita level for component X.

Overview export.

For completeness, the pipeline also exports an object `multipliers_overview`, which combines the main multiplier estimates into a single table with three panels. Panel A reports the baseline constant-parameter LP–IV and SLP–IV multipliers on the structural 1985Q1–2019Q4 sample; Panel B reports the TVP–SLP–IV regime multipliers for all monetary-policy regimes on 1985Q1–2023Q4; Panel C reports overlap multipliers for alternative instrument configurations on the AG–Ramey overlap sample. The overview table therefore aggregates constant-parameter, TVP, and overlap multipliers in a unified format. Appendix D reproduces this object in Table 16, and the regime-specific entries in Table 2 in the main text are drawn from Panel B.

D Mechanism Figures and Summary Tables

This Appendix collects the baseline SLP–IV cumulative mechanism responses and links them to the mechanism grid in the main text, provides a summary table of cumulative output multipliers across specifications, and reports regime-specific impulse-response grids from the TVP–SLP–IV specification.

Baseline SLP–IV mechanism impulse responses and cumulative responses on 1985Q1–2019Q4 are summarised in Table 15. The table reports cumulative responses of the main components of output (consumption, investment, and other private demand) to an Auerbach–Gorodnichenko (AG) spending shock over the baseline horizons $H \in \{8, 16\}$. Activity components (C, I, and $Y - G$) are reported in per-dollar units, while rate and expectations outcomes are reported in percentage points under the 1%-of-GDP-on-impact normalisation, as stated in the table notes.

Table 15: Cumulative responses of mechanisms, SLP–IV baseline (AG; 1985Q1–2019Q4).

Outcome	$R(8)$	$R(16)$
C	0.91	1.08
I (GPDIC1)	-0.48	-0.19
$Y - G$ (private output)	-0.28	-0.05
UNRATE (p.p.)	5.31	12.46
Inflation, headline (p.p. annualised)	-3.07	8.33
MICH expectations (p.p.)	-7.68	-12.40

Notes: For C, I, and $Y - G$, the columns $R(8)$ and $R(16)$ report cumulative component responses over horizons 0–8 and 0–16, expressed in real dollars per capita per $1 of cumulative real government spending. For UNRATE, headline PCE inflation, and Michigan inflation expectations, the columns $R(8)$ and $R(16)$ report cumulative percentage-point responses over the same horizons to a government-spending shock normalised to raise government spending by 1 percent of GDP on impact. Units and transformations follow Section 3.1 and Appendix C.

The corresponding baseline mechanism impulse-response grid for output, government spending, consumption, investment, inflation, expectations, unemployment, and the policy rate is plotted in Figure 1 in the main text. Joint consideration of Figure 1 and Table 15 provides a compact summary of the constant-parameter mechanisms underlying the baseline multipliers.

The baseline mechanism decomposition indicates that the modest positive output multipliers primarily reflect the direct increase in government purchases, accompanied by a moderate but non-negligible consumption response, while private investment and $Y - G$ remain close to zero and are often estimated to be slightly negative over the medium run. The cumulative responses for the unemployment rate reported in Table 15 are positive over both horizons. Because Table 15 reports cumulative responses over the horizon rather than level responses at a single date, these entries summarise the net effect of the underlying impulse-response path over the horizon and should not be interpreted as evidence of a monotonic change in unemployment at every horizon. Both headline inflation and inflation expectations exhibit only mild quarter-by-quarter movements in Figure 1 and are estimated imprecisely; nevertheless, persistent deviations can cumulate to non-zero sums over the horizons reported in Table 15, consistent with the subdued price and expectations effects emphasised in the main text.

D.1 Summary of Cumulative Output Multipliers

For completeness, this subsection collects the main multiplier estimates across specifications into a single overview table. Panel A reports the baseline constant-parameter SLP–IV multipliers (headline) together with LP–IV multipliers (diagnostic) for the structural 1985Q1–2019Q4 sample; Panel B reports the TVP–SLP–IV regime multipliers for the full 1985Q1–2023Q4 sample; and Panel C reports overlap multipliers for alternative instrument configurations on the AG–Ramey overlap sample. The underlying estimates are constructed as described in Appendix C. Panel B contains entries for all five monetary-policy regimes defined in Section 3.3; the main text reports regime multipliers for the three long pre-Covid regimes as the primary quantitative comparison, while the two short Covid and post-Covid regimes are treated as descriptive and may be blanked at horizons where ratios are weakly identified.

Table 16: Overview of cumulative output multipliers across specifications.

	Horizon 8			Horizon 16		
Method/Regime	Point	68% CI	90% CI	Point	68% CI	90% CI
A: Baseline (1985–2019)						
LP–IV	0.50	[-0.48, 1.21]	[-1.15, 1.95]	1.13	[-0.88, 1.75]	[-1.62, 3.21]
SLP–IV	0.59	[-0.28, 1.60]	[-0.94, 2.65]	0.80	[-0.43, 3.14]	[-1.40, 5.28]
B: TVP Regimes (1985–2023)						
Pre–ZLB	0.77	[0.14, 1.95]	[-1.45, 4.27]	0.96	[0.62, 1.40]	[-0.15, 1.94]
ZLB–GFC	0.41	[-0.62, 1.96]	[-5.49, 5.90]	1.02	[0.81, 1.26]	[0.23, 1.78]
Positive-rate (Pre-Covid)	0.48	[-0.52, 2.40]	[-3.28, 6.66]	1.04	[0.63, 1.43]	[-0.79, 2.57]
ZLB–Covid	–	–	–	–	–	–
Positive-rate (Post-Covid)	–	–	–	–	–	–
C: Overlap robustness						
LP–IV (AG overlap)	0.40	[-0.60, 1.20]	[-1.19, 1.97]	1.06	[-0.91, 1.88]	[-1.73, 3.40]
LP–IV (Ramey sample)	-0.44	[-1.39, 4.90]	[-4.51, 13.23]	-0.54	[-1.00, 5.28]	[-8.47, 14.78]
LP–IV (Both–IV overlap)	0.35	[-0.60, 1.18]	[-1.16, 1.96]	0.98	[-0.91, 1.84]	[-1.80, 3.29]

Notes: Panel A reports baseline constant-parameter SLP–IV multipliers (headline) together with LP–IV multipliers (diagnostic) on 1985Q1–2019Q4. Panel B reports TVP–SLP–IV regime multipliers on 1985Q1–2023Q4; these entries are unit-free ratios of cumulative log responses computed from the unified TVP–SLP–IV model estimated on the full 1985Q1–2023Q4 sample. For the ZLB–Covid and Positive-rate (Post-Covid) regimes, multipliers are (denoted "–") wherever the cumulative response of government spending at the relevant horizon is too small and weakly identified to support stable ratios; these entries are not used as core quantitative evidence in the main text. Panel C reports overlap multipliers for alternative instrument configurations on the AG–Ramey overlap sample.

D.2 Regime-specific TVP–SLP–IV impulse responses

This subsection reports regime-averaged impulse responses from the TVP–SLP–IV model to an Auerbach–Gorodnichenko government-spending shock for the three long pre-Covid regimes defined in Section 3.3. All regime-specific objects are computed as averages from the single full-sample TVP model; parameters are not re-estimated by regime. Regime-averaged objects are computed for all five regimes in the replication outputs, but only the three long pre-Covid regimes are displayed here because the short Covid and post-Covid windows provide limited effective support for medium-horizon summaries.

Figure 9: TVP–SLP–IV impulse responses to an AG spending shock, Pre–ZLB (1985Q1–2008Q3)

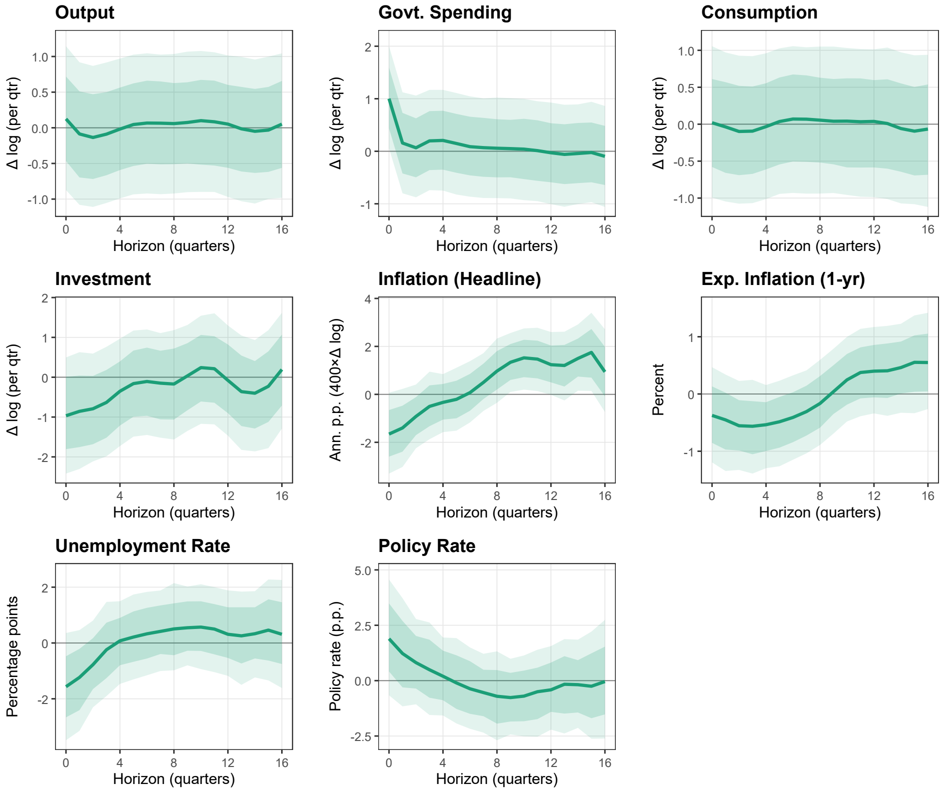


Figure 10: TVP–SLP–IV impulse responses to an AG spending shock, ZLB–GFC (2008Q4–2015Q3)

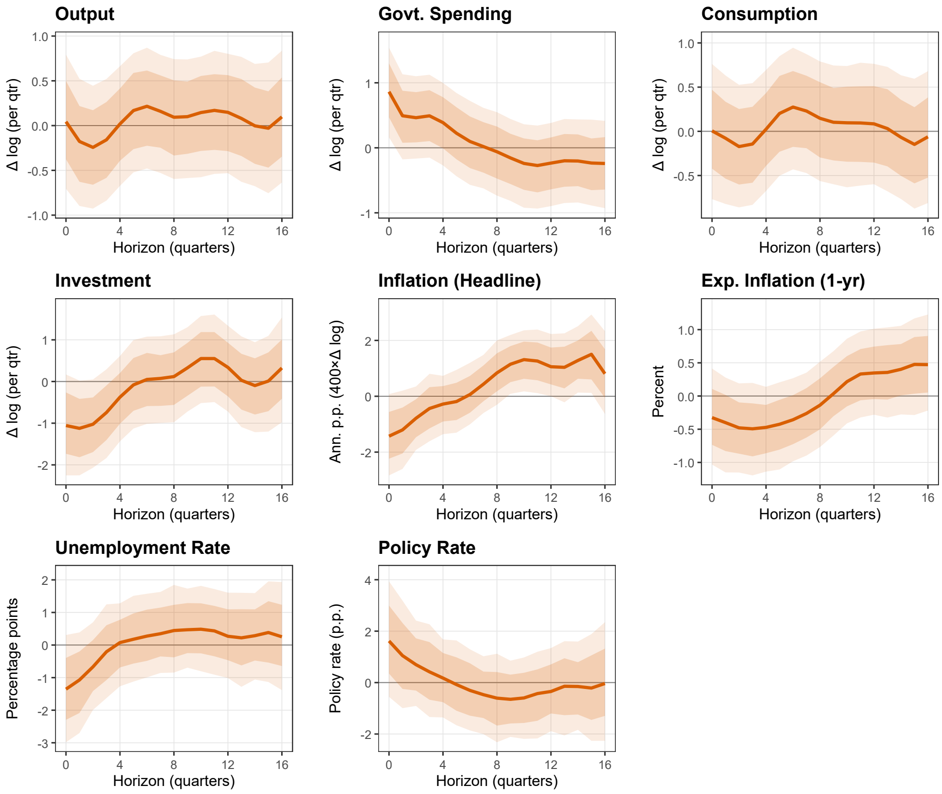
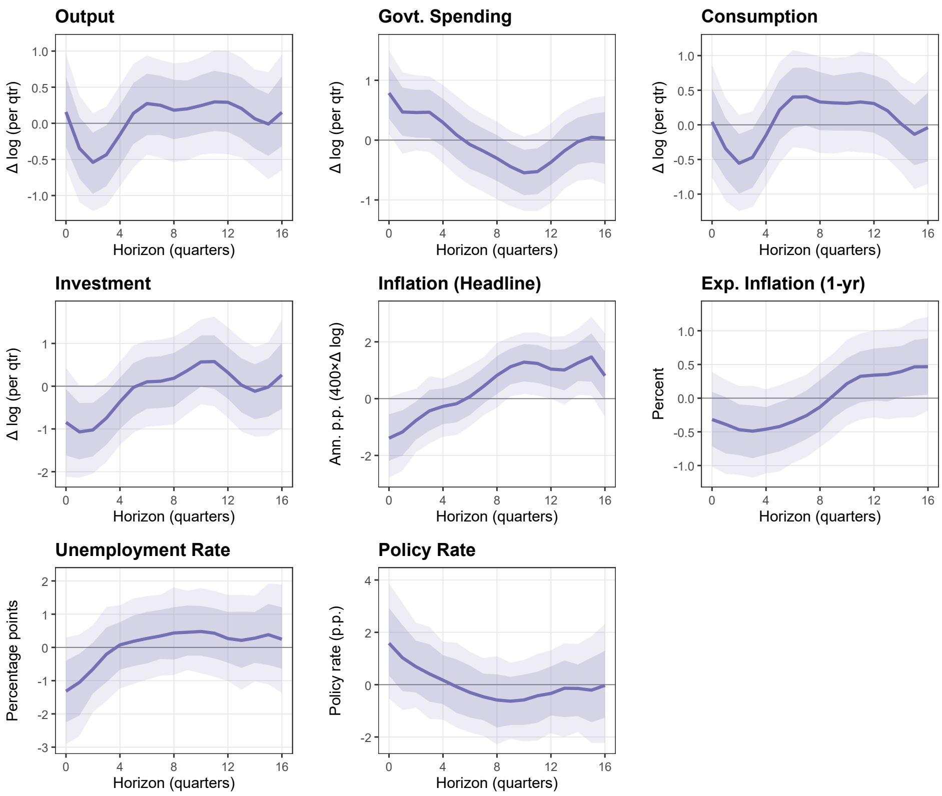


Figure 11: TVP–SLP–IV impulse responses to an AG spending shock, Positive-rate (Pre-Covid) (2015Q4–2019Q4)



E AG and Ramey Shock Construction

This Appendix documents the construction of the external instruments used to identify innovations to government purchases: (i) an Auerbach–Gorodnichenko AG-type (one-quarter-ahead forecast-error) spending shock constructed from the Survey of Professional Forecasters (SPF) and extended through 2023Q4 using a uniform methodology, and (ii) the Ramey defence-news shock series used on its original published support (restricted to the overlap with the baseline sample).

E.1 AG-type (SPF) Forecast-Error Shock

Concept.

Unexpected fiscal innovations are measured as the one-quarter-ahead forecast error in real government purchases growth. Let $\widetilde{G}_t$ denote aggregate real government consumption expenditures and gross investment, and let pop_t denote quarterly-average population. Define per-capita purchases as $G_t \equiv \widetilde{G}_t/\text{pop}_t$. In log growth,

$$\Delta \log G_t \;=\; \Delta \log \widetilde{G}_t \;-\; \Delta \log(\text{pop}_t). \tag{6}$$

Let $\Delta \log(G^{SPF}_{t|t-1})$ denote the SPF-based proxy for the expectation formed in quarter $t-1$ of per-capita government purchases growth in quarter t. The AG-type shock is the forecast error:

$$z^{AG}_t \;=\; \Delta \log G_t \;-\; \Delta \log(G^{SPF}_{t|t-1}). \tag{7}$$

Because the per-capita conversion uses the realised population growth rate for the realisation quarter t in both realised and SPF-implied growth (see (10)), population growth cancels in (7). Equivalently,

$$z^{AG}_t = \left[\Delta \log \widetilde{G}_t - \Delta \log(\text{pop}_t)\right] - \left[\Delta \log(\widetilde{G}^{SPF}_{t|t-1}) - \Delta \log(\text{pop}_t)\right] = \Delta \log \widetilde{G}_t - \Delta \log(\widetilde{G}^{SPF}_{t|t-1}), \tag{8}$$

so the shock can be viewed either as a per-capita forecast error or as an aggregate forecast error in $\Delta \log \widetilde{G}_t$.

Implementation (SPF alignment and horizons).

The implementation uses SPF level forecasts of government purchases. Let the SPF survey in quarter $t-1$ report mean level forecasts for federal purchases and state/local purchases at SPF horizon labels $m=1$ and $m=2$. Total purchases forecasts are formed as

$$\widehat{\widetilde{G}}^{SPF}_{t-1,m} \;=\; \widehat{\widetilde{G}}^{F}_{t-1,m} + \widehat{\widetilde{G}}^{SL}_{t-1,m}.$$

The pipeline aligns the SPF horizon labels so that $m=1$ corresponds to the forecasted level for quarter $t-1$ (the survey quarter) and $m=2$ corresponds to the forecasted level for quarter t (the subsequent quarter). Under this alignment, the one-quarter-ahead forecast of aggregate growth for the realisation quarter t is proxied by the log change between the two consecutive level forecasts:

$$\Delta\log(\widetilde{G}^{SPF}_{t|t-1}) \;\equiv\; \log\widehat{\widetilde{G}}^{SPF}_{t-1,2} - \log\widehat{\widetilde{G}}^{SPF}_{t-1,1}, \qquad \text{dated to quarter } t. \tag{9}$$

Monthly population is averaged to the quarter prior to differencing. The SPF forecast is converted to a per-capita growth forecast using realised quarterly population growth for the same realisation quarter:

$$\Delta\log(G^{SPF}_{t|t-1}) \;=\; \Delta\log(\widetilde{G}^{SPF}_{t|t-1}) - \Delta\log(\text{pop}_t). \tag{10}$$

Finally, z_t^{AG} is computed from (7) (equivalently, from (8)). The baseline series is constructed uniformly over 1985Q1–2023Q4 and stored under the alias `z_ag`; no additional logs or differences are applied, and any scaling/standardisation is handled within the IV code.

Identification in IV.

In the LP–IV specifications, identification requires (i) relevance of z_t^{AG} for innovations in government purchases, and (ii) exogeneity with respect to the second-stage error, i.e. $E\big[z_t^{AG}\,\varepsilon_{t+h} \mid X_t\big]=0$ for $h=0,\dots,H$, where X_t denotes the baseline control vector used in both stages. The forecast-error construction ensures orthogonality to the SPF information set dated $t-1$; remaining threats are addressed by the baseline control set described in the main text and Appendix B.

E.2 Ramey Defence-News Shock

Concept.

The defence-news measure is a narrative proxy for changes in expectations about future U.S. defence spending. It is intended to capture the quarter t information arrival—i.e., when agents learn about prospective defence buildups or drawdowns—rather than the quarter when the associated outlays are recorded in the NIPA accounts. In the original construction, episodes are identified and dated using contemporaneous news sources (e.g., periodicals and major newspapers) and then translated into revisions in the expected path of defence spending.

Construction and normalisation.

Let D_{t+i} denote nominal defence outlays i quarters ahead and r_t a discount rate at time t. The news measure is constructed as the present discounted value (PDV) of revisions in expected future spending:

$$\text{PDV}_t = \sum_{i=0}^{I} \frac{\Delta E_t D_{t+i}}{(1+r_t)^i},$$

where $\Delta E_t D_{t+i}$ is the revision in expectations at time t. In the original series, r_t is taken from market interest rates (commonly the three-year Treasury rate in the postwar sample), and I typically corresponds to a medium-horizon projection window (about 3–5 years, with longer horizons only in special episodes). News arriving very late in a quarter is conventionally assigned to the subsequent quarter.

The shock used in the analysis scales the nominal PDV by lagged nominal GDP:

$$z_t^R = \frac{\text{PDV}_t}{\text{NGDP}_{t-1}}.$$

Implementation in the analysis.

The quarterly defence-news series is taken from the published Ramey dataset, aligned to NIPA quarters, and used as provided (i.e., in PDV terms and normalised by lagged nominal GDP as in z_t^R). In the baseline analysis the 1985Q1–2015Q4 segment of the published series is used to match the overlap with the main sample; observations outside this window are treated as missing in the overlap-sample exercises. No extension, splicing, or mid-sample reconstruction is performed beyond the published series. In Appendix A it appears under the alias `z_ramey`. No additional transformations are applied beyond the original construction, and any standardisation is performed inside the estimation code.

Complementarity on the overlap sample.

The AG-type SPF forecast-error shocks capture short-horizon surprises in realised government purchases, whereas the defence-news shocks capture medium-horizon news about future defence spending. On the overlap sample, the two instruments are combined in an over-identified specification (Both–IV) that supports over-identification diagnostics (e.g. Hansen J) and Anderson–Rubin confidence sets.