

LOCAL PROJECTIONS WITH IMPERFECT INSTRUMENTS AND CONTROLS*

Christian Matthes

Department of Economics, University of Notre Dame

October 12, 2026

Abstract

Local projections with instrumental variables estimate policy effects as the ratio of two regressions on the instrument, one for the outcome and one for the policy variable. Exclusion requires only that the instrument be uncorrelated, given controls, with the outcome net of the policy variable's contribution. This condition depends on the outcome, horizon, normalization, and controls: an instrument can identify a policy shock's initial effect but not its persistence, and the first stage cannot tell. Shock-by-shock exclusion restrictions imply the condition for every response but are necessary only for a full set of responses. Where exclusion is in doubt, I construct weak-instrument-robust confidence sets under bounds on contamination of both regressions. Ramey–Zubairy spending multipliers stay below one only if contamination biases them by less than 0.03. A Smets–Wouters Monte Carlo shows how controls change the estimated size and persistence of monetary policy effects while barely moving the first stage.

JEL Codes: C22, C26, E32, E52, E62.

Keywords: local projections; instrumental variables; fiscal multipliers; monetary transmission; information controls.

*Correspondence: cmatthes@nd.edu.

1 INTRODUCTION

How large are the effects of a policy shock, and how long do they last? Local projections with external instruments (LP-IV) answer both questions with the same instrument, the same controls, and the same first stage (Jordà, 2005; Stock and Watson, 2018). This paper shows that the instrument can identify one answer and not the other. It can identify the initial response but not its persistence, or one variable's response but not another's, and the first stage does not reveal which. What decides each case is whether the other shocks correlated with the instrument move the outcome and the normalization variable in the proportion given by the true response; when they do, their contribution cancels. In the Smets–Wouters model, an instrument uncorrelated with every other current shock understates the first-year price decline by 55% and the subsequent decline by 98%. In the fiscal design of Ramey and Zubairy (2018), the conclusion that spending multipliers lie below one survives only if contamination biases the multiplier by less than 0.03 at two years and 0.045 at four.

Consider LP-IV estimates of the first-year price response to a monetary policy shock and the additional price change in years two and three. Suppose the policy-announcement instrument also contains news about fiscal policy, say an indirect-tax increase scheduled for the second year. For this thought experiment, assume that the tax news leaves the current policy rate unchanged and affects consumer prices only when the tax takes effect.¹ If the instrument otherwise satisfies exclusion, the first-year estimate identifies the causal effect of the monetary policy shock. The later estimate is biased because it attributes part of the tax-induced price increase to monetary policy. Both regressions have the same first stage when they use the same current interest rate, instrument, controls, and sample. First-stage strength therefore cannot establish that both price responses are identified.

Stock and Watson (2012) document substantial correlations among shock estimates obtained using external instruments intended to identify different structural shocks. Their findings raise the possibility that an instrument is correlated with more than one structural shock. Following Stock and Watson (2018), I express the policy effect as an outcome's response to a specified shock divided by that same shock's effect on the variable used for normalization. Another shock that remains correlated with the instrument after controlling for observed variables can affect both parts of this ratio. Its contribution cancels when its effect on the outcome equals the target coefficient times its effect on the normalization variable: for a spending multiplier of one, a shock that raises output and government purchases by equal amounts causes no bias. This cancellation need not hold at another horizon or for another outcome.

I derive three results. The first is the necessary and sufficient condition for one normalized response: after controls, the instrument must be uncorrelated with the outcome net of the policy shock's contribution through the normalization. The second characterizes when a collection of responses is identified. The shock-by-shock exclusion restrictions of Stock and Watson (2018) and Plagborg-Møller and Wolf (2021) are the special case in which every residual covariance is zero. They are always sufficient, and they are necessary exactly when the net loadings of the responses have full column rank. Identifying a transmission mechanism therefore requires the conventional conditions; identifying a single response does

¹For the interest-rate assumption, see González-Páramo (2005), who explains why monetary policy need not respond to a temporary increase in inflation caused by indirect-tax changes.

not. The third is a confidence set for the responses under bounded violations of exclusion that remains valid with weak instruments. The relevant question is which conclusions survive which violations, and it can be answered response by response. The same conditions govern proxy-SVAR identification, and they say what controls must accomplish: lagged variables that stand in for an omitted index leave a bias that I characterize, and a control dated before the policy decision can create a bias where none existed.

I apply these restrictions to the military-news design of Ramey and Zubairy (2018). Confidence sets for cumulative government-spending multipliers lie below one at two and four years, even when inference allows for weak instruments. I examine how this conclusion changes when military news is correlated with other shocks that affect output and government purchases. At a multiplier of one, these shocks matter through their effects on GDP less government purchases: consumption, investment, and net exports. I vary bounds on their contributions to the coefficients on news for cumulative output and purchases, and calculate when the confidence sets admit a multiplier of one. The calculation makes explicit which departures from exclusion overturn the conclusion at each horizon.

I use the model of Smets and Wouters (2007) to study contamination by past shocks. I construct an instrument that combines the current monetary policy innovation, an inflation forecast formed before the policy decision, and independent measurement noise. In this experiment, the instrument is uncorrelated with every current structural innovation other than the monetary policy innovation. Its contamination arises from past shocks reflected in the inflation forecast. Those shocks continue to affect prices and interest rates. Lagged macroeconomic controls explain much of the forecast, but leave enough of this component in the instrument to produce substantial bias. In the reference specification with four lags of macroeconomic variables, the population coefficients understate the first-year price decline by 55% and the additional decline in years two and three by 98%. Adding the exact forecast as a control eliminates these biases while changing the first-stage covariance by about 3%.

When the forecast is measured with error, adding it as a control need not eliminate the bias. In the experiment, the model's exact forecast accounts for 80% of the variance of its noisy measure. Lagged macroeconomic variables already explain most of this signal. After conditioning on these lags, the signal accounts for only 34.8% of the remaining variance, so adding the measured forecast removes only part of the bias. A forecast formed after the policy decision presents a different problem: it responds to the monetary policy shock itself. Controlling for that forecast can therefore bias the estimates even when the instrument is otherwise valid.

Valid exclusion does not guarantee accurate inference in small samples. In 2,000 Monte Carlo replications with the exact forecast control, nominal 95% joint confidence sets contain both true responses in 85.8% of samples with 120 quarters and 92.2% with 240 quarters. Coverage rises to 94.7% with 960 quarters. When the instrument is contaminated and the regressions omit the controls, increasing the sample from 120 to 240 quarters instead reduces coverage from 50.5% to 17.0%. In this experiment, larger samples improve the sampling approximation when exclusion holds but concentrate estimates around biased population coefficients when it fails.

Related work. Stock and Watson (2018) develop conditions under which external instruments iden-

tify dynamic causal effects, including normalized and cumulative responses. Plagborg-Møller and Wolf (2021) establish LP-VAR equivalence under unrestricted lag structures and show that correlation with shocks irrelevant to an outcome need not prevent identification up to scale. I show that their restrictions are the special case of my conditions in which every residual covariance is zero, and that this case is necessary only when a full set of responses is required. Miranda-Agrippino and Ricco (2023) also weaken the conditions for external-instrument SVARs: they require only that the shock of interest be recoverable from the VAR innovations and characterize the bias when the instrument is correlated with other shocks. Their conditions are specific to the shock of interest; mine are specific to the response of interest, so they can hold for one outcome or horizon and fail for another. Under the conditions for equivalence between LP-IV and internal-instrument VARs, these restrictions apply to both estimators. Agreement between their estimates therefore does not establish identification. Jordà and Taylor (2025) and Montiel Olea et al. (2026) review LP and VAR methods.

Miranda-Agrippino and Ricco (2021) and Jarociński and Karadi (2020) study monetary surprises that combine policy innovations with information about fundamentals. Lloyd and Manuel (2024) compare using controls directly with first constructing a shock series. In work outside the local-projection setting, Deuchert and Huber (2017) and Glynn, Rueda, and Schuessler (2024) show how adding controls can introduce bias in IV estimation. I examine these control problems in LP-IV, where their consequences for identification depend on the outcome, horizon, and normalization. Lewis and Mertens (2026) study weak-instrument bias when exclusion holds. My Monte Carlo experiment varies instrument strength and exclusion separately to distinguish the consequences of weak instruments from those of invalid instruments.

Khalaf, Lin, and Tang (2026) show that omitted or mismeasured controls can invalidate LP-IV and cause Anderson–Rubin tests to overreject. They also propose a “total causal effect” for an alternative policy innovation, with a causal interpretation based on recursive timing restrictions. My conditions concern the original policy shock: they establish when contamination cancels for one response but biases another, and when requiring several responses to be correct rules out such cancellations, so that the shock-by-shock restrictions of Stock and Watson (2018) become necessary as well as sufficient. I relate this distinction to identification in internal and external proxy-SVARs. I also derive confidence sets for the original policy responses under joint bounds on other shocks’ contributions to the outcome and normalization variable. These sets quantify which conclusions about policy transmission survive specified departures from exclusion.

The fiscal analysis uses the cumulative LP-IV estimator introduced by Ramey and Zubairy (2018) and follows Conley et al. (2012) in assessing sensitivity to departures from exclusion. News about future fiscal policy can affect current economic activity (Ramey, 2011; Leeper et al., 2013). Briganti et al. (2026) show that delays in recording government purchases can also make output appear to respond before purchases. I retain Ramey and Zubairy’s purchases measure and examine contamination of the output and purchases coefficients by other shocks correlated with military news. I report which joint bounds on these contributions keep the confidence sets below one at each horizon.

Section 2 uses a simple government-spending example to show why identification can differ across

horizons and why adding a control can introduce bias. Section 3 derives the general identification conditions, relates them to assumptions used in internal and external proxy-SVARs, and constructs confidence sets under bounds on contamination. Section 4 applies these confidence sets to the military-news data of Ramey and Zubairy (2018) and asks how much contamination would overturn the conclusion that cumulative spending multipliers are below one. Section 5 uses the Smets–Wouters model to examine how the choice of controls affects the size and persistence of estimated price responses. Its Monte Carlo experiment distinguishes identification bias from sampling error. Section 6 concludes.

2 A GOVERNMENT-SPENDING EXAMPLE

I begin with a simple government-spending example to explain why identification must be assessed separately for each horizon and control specification. Consider GDP responses one and two periods after a spending shock, each divided by the shock’s effect on current government purchases. The true multiplier is one at both horizons. The instrument also reflects variation unrelated to the spending shock. This variation raises current purchases and next-period GDP equally, leaving the first IV ratio unchanged. It does not affect GDP two periods ahead, so the second ratio understates the multiplier. Adding a control observed before the shock can prevent this cancellation and bias both estimates. These differences arise even though the two regressions share the same first stage.

Let G_t denote government purchases and Y_t GDP, both measured as deviations from their means in the same real-dollar units. Let m_t denote the government-spending shock. Two variables, $s_{1,t}$ and $s_{2,t}$, are determined before this shock is realized. The variable $s_{1,t}$ affects current government purchases and next-period GDP and also enters the instrument Z_t . The instrument therefore reflects both the spending shock and this other source of variation in purchases and output. The variable $s_{2,t}$ affects next-period GDP but enters neither current purchases nor the instrument. Neither variable affects GDP two periods ahead. Government purchases, future output, and the instrument satisfy

$$\begin{aligned} G_t &= m_t + s_{1,t}, \\ Y_{t+1} &= m_t + s_{1,t} + s_{2,t}, \\ Y_{t+2} &= m_t + \varepsilon_t, \\ Z_t &= m_t + s_{1,t} + v_t. \end{aligned} \tag{1}$$

The term v_t is noise in the instrument, and ε_t is an additional disturbance to GDP two periods ahead. For each t , $m_t, s_{1,t}, s_{2,t}, v_t$ are mutually independent, with mean zero and unit variance. The disturbance ε_t is independent of these variables, with mean zero and finite variance. Appendix A.1 derives the results below.

A one-unit increase in m_t , holding $s_{1,t}, s_{2,t}$, and all other innovations fixed, raises current government purchases and GDP at each of the next two dates by one unit. The structural multipliers are therefore $\beta_1 = \beta_2 = 1$. The fiscal application in Section 4 instead divides the cumulative GDP response by the cumulative government-purchases response.

2.1 THE REGRESSIONS AND THEIR IDENTIFYING CONDITIONS The normalization variable in this example is current government purchases, $X_t = G_t$. The first stage regresses G_t on the instrument Z_t and controls W_t :

$$G_t = a_G + pZ_t + \gamma'_G W_t + e_{G,t}. \quad (2)$$

Estimating this regression by OLS gives fitted government purchases $\hat{G}_t = \hat{a}_G + \hat{p}Z_t + \hat{\gamma}'_G W_t$. The second stage regresses GDP at each horizon on fitted government purchases and the same controls:

$$Y_{t+h} = \hat{a}_h + \hat{b}_h \hat{G}_t + \hat{\gamma}'_h W_t + \hat{e}_{h,t}, \quad h = 1, 2. \quad (3)$$

The coefficients $\hat{b}_h$ are the two-stage least-squares estimates. Both stages include an intercept, and both horizons use the same estimation dates. Within each control specification, only the second-stage outcome changes with the horizon. The first-stage regression and fitted purchases are therefore identical across horizons.

The controls used in this example depend on variables determined before the current spending shock m_t is realized. They may reflect past spending shocks, but they do not respond to m_t . Let $V_t^{\perp W}$ denote the residual from the population linear regression of V_t on W_t and an intercept. Assume the regressors are not perfectly collinear in population and sample moments converge to their population values. The IV estimator consistently estimates the structural multiplier β_h if relevance and exclusion hold:

$$\sigma_{ZX}(W) = \text{Cov}(Z_t^{\perp W}, G_t^{\perp W}) \neq 0, \quad \text{relevance}, \quad (4)$$

$$\kappa_h(W) = \text{Cov}(Z_t^{\perp W}, (Y_{t+h} - \beta_h G_t)^{\perp W}) = 0, \quad \text{exclusion}. \quad (5)$$

Relevance requires a nonzero covariance between the instrument and current government purchases after removing the controls. The IV coefficient is the instrument's covariance with future GDP divided by its covariance with current purchases, with the same controls removed from all variables. For this ratio to equal the structural multiplier β_h , the numerator must equal β_h times the denominator. Exclusion expresses this requirement: the residualized instrument must have zero covariance with the residual of $Y_{t+h} - \beta_h G_t$.

In this example, the true multiplier is one at both horizons, so the two covariances must be equal. Subtracting current purchases from future GDP gives

$$Y_{t+1} - G_t = s_{2,t}, \quad Y_{t+2} - G_t = \varepsilon_t - s_{1,t}.$$

The current spending shock contributes equally to output and purchases and therefore disappears from these differences. Identification requires the instrument to be uncorrelated with the relevant difference after removing the controls. A control can predict GDP or purchases accurately while leaving this correlation nonzero. In the terms of Stock and Watson (2018), the instrument is invalid at both horizons because it is correlated with $s_{1,t}$, a disturbance other than the spending shock; the exclusion condition (5) is weaker, because it concerns the covariance with a particular combination of disturbances, and that

combination excludes $s_{1,t}$ at $h = 1$ but not at $h = 2$.

IV chooses its coefficient to satisfy the zero-covariance condition. The condition therefore holds at the population IV coefficient even when that coefficient differs from the true multiplier. Identification requires it to hold at the true multiplier, here one.

The population IV coefficient satisfies

$$b_h(W) = \frac{\text{Cov}(Z_t^{\perp W}, Y_{t+h}^{\perp W})}{\sigma_{ZX}(W)} = 1 + \frac{\kappa_h(W)}{\sigma_{ZX}(W)}. \quad (6)$$

The ratio $\kappa_h(W)/\sigma_{ZX}(W)$ is the difference between the population IV coefficient and the true multiplier. When exclusion fails, the IV estimator converges to a value different from the true multiplier as the sample grows.

2.2 WHICH CONDITION FAILS? I compare four specifications: no additional controls; $s_{1,t}$ observed without error; a noisy measure $s_{1,t} + \eta_t$; and an indicator $s_{1,t} + s_{2,t}$ that combines $s_{1,t}$ with another determinant of next-period GDP. The measurement error η_t has unit variance and is independent of all the variables in (1). It is determined before m_t is realized, so all three added controls precede the current spending shock.

Table 1: Relevance holds in every specification; exclusion need not

Controls	Relevance	Exclusion moments		Multipliers	
	$\sigma_{ZX}(W)$	$\kappa_1(W)$	$\kappa_2(W)$	$h = 1$	$h = 2$
None	2	0	-1	1	1/2
Exact state $s_{1,t}$	1	0	0	1	1
Noisy state $s_{1,t} + \eta_t$	3/2	0	-1/2	1	2/3
Mixed indicator $s_{1,t} + s_{2,t}$	3/2	-1/2	-1/2	2/3	2/3
True multipliers	—	—	—	1	1

Notes: Exact population calculations for (1). Each multiplier measures output one or two periods ahead per dollar of impact government purchases. The first four rows report population IV coefficients $b_h(W)$; the final row reports the structural multipliers $\beta_1 = \beta_2 = 1$. Every control specification includes an intercept and has full-rank regressors. Relevance requires $\sigma_{ZX}(W) \neq 0$; exclusion requires $\kappa_h(W) = 0$ at the horizon being estimated. The relevance column reports a residual covariance. The first-stage regression slope is $p = \sigma_{ZX}(W)/\text{Var}(Z_t^{\perp W})$. Appendix A.1 derives the results for all four control specifications.

Without additional controls, relevance holds: the instrument's covariance with current purchases is two. Its covariance with GDP is two at the first horizon and one at the second, giving IV coefficients $2/2 = 1$ and $1/2$. The variable $s_{1,t}$ raises current purchases and next-period GDP equally, so it causes no bias in the first ratio. At the second horizon, it contributes to the denominator but not the numerator. The IV coefficients therefore suggest that the effect on GDP falls by half, although the true multiplier is one at both horizons.

Controlling for $s_{1,t}$ removes its contribution from the instrument, current purchases, and next-period GDP. Both exclusion conditions then hold, and both IV coefficients equal one. With the noisy control $s_{1,t} + \eta_t$, residual variation in $s_{1,t}$ remains in those three variables. It contributes equally to the numerator and denominator of the first IV ratio, which therefore still equals one. At the second horizon, this

remaining component enters only the denominator, giving a biased coefficient of $1/(3/2) = 2/3$.

After projecting on $s_{1,t} + s_{2,t}$, the residual of next-period GDP is m_t . The residuals of the instrument and current purchases both contain the additional component $(s_{1,t} - s_{2,t})/2$. The covariance in the first-horizon numerator is therefore one, while the denominator is $3/2$. Adding this control changes the first IV coefficient from its correct value of one to $2/3$. The second-horizon coefficient is also $2/3$.

The noisy and mixed controls both explain half the variance of $s_{1,t}$ and give the same first-stage covariance. Both are determined before the current spending shock. These properties do not guarantee that adding a control preserves identification.

3 WHAT IDENTIFIES A POLICY RESPONSE?

The example shows why identification must be checked for each response and control specification. I now derive the restrictions under which the population IV coefficient equals the normalized response to the policy shock of interest. These restrictions depend on the outcome, horizon, and variable used to scale the shock. I examine when they hold jointly across responses, how they relate to proxy-SVAR assumptions, and how to construct confidence sets under departures from exclusion.

3.1 THE POLICY EXPERIMENT AND ITS NORMALIZATION Let m_t denote the policy shock of interest. The impulse response of a variable at horizon j is the change in that variable j periods after a one-unit increase in m_t , holding the history before the shock and all other shocks fixed. Private behavior and subsequent policy choices are allowed to respond to the shock. For a spending-news shock, for example, the GDP response includes the effects of any changes in purchases, taxes, and monetary policy induced by the news. The cumulative spending multiplier divides the cumulative GDP response by the cumulative purchases response over the same horizons. It therefore expresses the effect of the spending-news shock per unit of purchases induced by that shock.

The policy shock is generally unobserved. An external instrument can be correlated with that shock without measuring its magnitude. Under the exclusion conditions of Stock and Watson (2018), the instrument's covariance with the outcome equals the policy response multiplied by the instrument's unknown covariance with the shock. Its covariance with another variable affected by the shock contains the same unknown factor. Taking the ratio of these covariances eliminates that factor and identifies the outcome response relative to the response of the second variable. This second variable serves as the normalization variable and the endogenous regressor in the LP. The resulting coefficient expresses the policy effect in economic units, such as the price response per percentage-point increase in the interest rate or the GDP response per dollar of additional government purchases.

Assumption 1 (Additive and homogeneous policy-shock effects). *The shock m_t has mean zero and positive variance. Let $Y_{i,t}$ denote the value of the outcome variable i at date t , and let D_t denote the variable used to construct the normalization. Let $\mathcal{B}_t$ collect the prior history and all other shock realizations relevant to the variables under study.*

Setting $m_t = m$, while holding $\mathcal{B}_t$ fixed, produces potential outcomes satisfying

$$Y_{i,t+j}(m; \mathcal{B}_t) = Y_{i,t+j}(0; \mathcal{B}_t) + \theta_{i,j} m, \quad (7)$$

$$D_{t+s}(m; \mathcal{B}_t) = D_{t+s}(0; \mathcal{B}_t) + \theta_{D,s} m. \quad (8)$$

Subsequent behavior and policy choices may respond to m . The coefficients $\theta_{i,j}$ and $\theta_{D,s}$ are the impulse responses to a one-unit increase in the policy shock. They do not vary with the date, prior history, or other shock realizations.

Observed outcomes satisfy

$$Y_{i,t+j} = Y_{i,t+j}(m_t; \mathcal{B}_t), \quad D_{t+s} = D_{t+s}(m_t; \mathcal{B}_t).$$

Thus, all other shocks retain their realized values when the potential outcomes are evaluated at the realized policy shock. All variables used in the population analysis have finite second moments. The response used as the denominator of each reported ratio is nonzero.

Assumption 1 does not restrict the serial dependence of m_t . Correlation with policy disturbances at other dates can affect identification through the exclusion conditions derived below.

I suppress $\mathcal{B}_t$ in the potential-outcome notation below and omit the outcome index i when considering a single variable. This assumption restricts the effect of the current policy shock. It does not require all variation due to history and other shocks to enter linearly. I impose homogeneous effects to compare instruments and control specifications against a common causal response. This is an assumption about the policy experiment, not a consequence of estimating a linear LP. Under suitable identifying conditions, linear LPs can also identify weighted averages of nonlinear responses, with weights that can depend on the instrument and controls (Kolesár and Plagborg-Møller, 2025). In that setting, differences between estimates can reflect differences in weighting as well as failures of exclusion.

The linearity restriction in Assumption 1 is weaker than that in the structural moving-average model used for identification by Plagborg-Møller and Wolf (2021). Their model makes each observed variable a linear function of all current and past structural shocks; here, only the effects of the target policy shock are restricted to be additive and homogeneous. This comparison concerns their structural model: their LP-VAR equivalence result does not require a linear data-generating process.

For each policy effect of interest, indexed by r , we choose an outcome variable, the horizons over which to measure its response, and a normalization. The index t denotes the date of the policy shock. We construct the normalization variable as

$$X_{r,t} = \sum_{s=0}^{M_r} c_{r,s} D_{t+s}, \quad (9)$$

and the outcome as

$$Q_{r,t} = \sum_{j=0}^{J_r} d_{r,j} Y_{i(r),t+j}, \quad (10)$$

where $i(r)$ identifies the outcome used for response r , and the weights $c_{r,s}$ and $d_{r,j}$ are fixed before esti-

mation. Both variables are indexed by the shock date, even when they include observations from subsequent dates. The sums allow us to measure effects at individual horizons or cumulatively. For example, normalizing by the current interest rate uses only D_t , whereas a cumulative spending multiplier sums output and government purchases over the same horizons following the shock.

For a response at a single horizon h , set $d_{r,h} = 1$ and the other outcome weights to zero. For a cumulative response, set the weights to one over the horizons included in the sum and zero otherwise. The fiscal application sums normalized GDP over quarters 0–7 or 0–15. The Monte Carlo sums inflation over quarters 1–4 or 5–12. All weights and horizons are fixed before estimation.

Let $\theta_{Q,r}$ and $\theta_{X,r}$ denote the responses of the outcome $Q_{r,t}$ and the normalization variable $X_{r,t}$ to a one-unit increase in the policy shock. The normalized policy effect is

$$\beta_r = \frac{\theta_{Q,r}}{\theta_{X,r}}, \quad \theta_{Q,r} = \sum_j d_{r,j} \theta_{i(r),j}, \quad \theta_{X,r} = \sum_s c_{r,s} \theta_{D,s} \neq 0. \quad (11)$$

Thus, β_r expresses the policy shock's effect on $Q_{r,t}$ per unit change in $X_{r,t}$ induced by that same shock. Under Assumption 1, a policy shock of size a changes $Q_{r,t}$ by $a\theta_{Q,r}$ and $X_{r,t}$ by $a\theta_{X,r}$. For any nonzero a , the shock magnitude cancels from the ratio. Thus, β_r describes the relative response regardless of the size of the policy shock.

The condition $\theta_{X,r} \neq 0$ requires the policy shock to change the variable used to normalize its effect. Instrument relevance is a separate condition: the instrument and the normalization variable must remain correlated after removing the variation explained linearly by the controls. Spending news, for example, may leave current purchases unchanged while raising cumulative purchases. The response normalized by current purchases is then undefined, even if the instrument predicts current purchases because it is correlated with other shocks that affect them. Normalizing by cumulative purchases gives a well-defined ratio in this example, but instrument relevance must still be established for that normalization.

The normalized effect β_r remains specific to the policy shock under study. For a spending-news shock, changing the shock's size scales the purchases and output response paths together. A different intervention may instead change the timing of purchases or the accompanying tax and monetary-policy responses. Even if it generates the same cumulative increase in purchases over the same horizons, it can have a different effect on cumulative GDP. Applying the multiplier to such an intervention therefore requires additional assumptions.

3.2 THE IDENTIFYING RESTRICTION FOR AN INDIVIDUAL RESPONSE Consider the LP regression of $Q_{r,t}$ on $X_{r,t}$ and a finite vector of controls $W_{r,t}$, using Z_t as the instrument for $X_{r,t}$. The controls include an intercept. Write $\tilde{Z}_{r,t}$, $\tilde{Q}_{r,t}$, and $\tilde{X}_{r,t}$ for the residuals from their population linear projections on $W_{r,t}$. These residuals remove the variation explained linearly by the controls. The index r on $\tilde{Z}_{r,t}$ allows different equations to use different controls even when the original instrument is the same. When the control set needs to be explicit, write $A_t^{\perp W}$ for the residual of a variable A_t after projecting on W_t . For a fixed

specification, the population IV coefficient is

$$b_r = \frac{\mathbb{E}[\tilde{Z}_{r,t}\tilde{Q}_{r,t}]}{\sigma_{ZX,r}}, \quad \sigma_{ZX,r} = \mathbb{E}[\tilde{Z}_{r,t}\tilde{X}_{r,t}], \quad (12)$$

provided $\sigma_{ZX,r} \neq 0$. When comparing control sets, I write $b_r(W)$ and $\sigma_{ZX,r}(W)$ to show which controls are used. The denominator is the covariance between the residualized instrument and the residual normalization variable. A nonzero value establishes instrument relevance, but equality between b_r and the policy effect β_r also requires an exclusion restriction.

This exclusion restriction requires the residualized instrument to be uncorrelated with the residual of $Q_{r,t} - \beta_r X_{r,t}$, where β_r is the normalized policy effect defined above. This difference follows from the ratio defining β_r : the policy shock's effect on $Q_{r,t}$ equals β_r times its effect on $X_{r,t}$. Under Assumption 1, subtracting $\beta_r X_{r,t}$ therefore cancels the current policy shock's contribution. Define

$$u_{r,t} = \tilde{Q}_{r,t} - \beta_r \tilde{X}_{r,t}.$$

Decomposing $Q_{r,t} - \beta_r X_{r,t}$ into its population linear projection on the controls and its residual gives

$$Q_{r,t} = \beta_r X_{r,t} + \mathcal{P}_{W_r}(Q_{r,t} - \beta_r X_{r,t}) + u_{r,t}, \quad (13)$$

where $\mathcal{P}_{W_r}$ denotes that projection. Cancellation of the policy shock's contribution to $Q_{r,t} - \beta_r X_{r,t}$ does not establish exclusion: the residualized instrument can still covary with $u_{r,t}$. Identification requires both relevance and exclusion:

$$\sigma_{ZX,r} \neq 0, \quad (14)$$

$$\mathbb{E}[\tilde{Z}_{r,t}u_{r,t}] = 0. \quad (15)$$

Proposition 1 (Identification of a normalized policy response). *Under Assumption 1, suppose $\sigma_{ZX,r} \neq 0$. Then*

$$b_r - \beta_r = \frac{\mathbb{E}[\tilde{Z}_{r,t}u_{r,t}]}{\sigma_{ZX,r}}. \quad (16)$$

Hence the population IV coefficient b_r equals the normalized policy effect β_r if and only if the exclusion condition (15) holds.

Proposition 1 is the IV bias identity for the normalized response β_r . Its content lies in the error term: $u_{r,t}$ is the outcome net of the policy shock's contribution through the normalization, so the restriction is specific to the outcome, horizon, normalization, and controls that define r , and it is necessary as well as sufficient. Condition (15) differs in form from the exclusion restrictions of Stock and Watson (2018) and Plagborg-Møller and Wolf (2021), which require the instrument to be uncorrelated with each other structural shock at each date: under Assumption 1, $u_{r,t}$ is a linear combination of those shocks and of the past, so their shock-by-shock restrictions imply (15) for every response, whereas (15) requires only that the weighted sum of the corresponding covariances vanish for the response in question; Section 3.4

makes this comparison precise. Suppose, for illustration, that another shock $S_{j,t}$ enters the outcome and normalization variable additively, with constant coefficients $\theta_{Q,r}^j$ and $\theta_{X,r}^j$. Its contribution to $Q_{r,t} - \beta_r X_{r,t}$ is $\ell_{r,j} S_{j,t}$, where

$$\ell_{r,j} = \theta_{Q,r}^j - \beta_r \theta_{X,r}^j. \quad (17)$$

At a cumulative fiscal multiplier of one, $\ell_{r,j}$ is the difference between that shock's effects on cumulative GDP and cumulative government purchases. A shock that raises purchases more than GDP lowers consumption, investment, and net exports taken together, with all quantities measured over the same horizons and in the same units. Exclusion concerns the residualized instrument's covariance with this net contribution. The fiscal application in Section 4 uses this restriction to assess sensitivity to other shocks correlated with military news.

The shock $S_{j,t}$ introduced in the preceding paragraph contributes $\ell_{r,j} S_{j,t}$ to $Q_{r,t} - \beta_r X_{r,t}$. After removing the controls, this contribution becomes $\ell_{r,j} S_{j,t}^{\perp W_r}$. Its contribution to the exclusion moment is therefore

$$\ell_{r,j} \mathbb{E}[\tilde{Z}_{r,t} S_{j,t}^{\perp W_r}].$$

This contribution vanishes if the residualized instrument is uncorrelated with the shock, or if $\ell_{r,j} = 0$, so that the shock's effects on the outcome and normalization cancel in $Q_{r,t} - \beta_r X_{r,t}$. Nonzero contributions from different shocks can also offset. Because the denominator $\sigma_{ZX,r}$ in Proposition 1 is common to all responses that share the normalization and controls, the net loadings also govern the size of the bias across horizons: the same instrument contamination produces a small bias where $\ell_{r,j}$ is small and a large one where it is large, as the Monte Carlo experiment in Section 5 illustrates. Requiring zero covariance with each of these shocks is therefore stronger than requiring their contributions to sum to zero. Identification requires the entire exclusion moment to vanish, including contributions from any remaining sources of variation. A cancellation that holds for one outcome or horizon need not hold for another; changing the normalization or controls can also change whether exclusion holds.

3.3 FROM ONE RESPONSE TO A TRANSMISSION MECHANISM Describing policy transmission often requires responses of several variables at several horizons. Let $\mathcal{R}$ be a finite collection of responses estimated using the same instrument Z_t and controls W_t . Let S_t collect a finite set of other shocks and economic state variables, including relevant leads and lags. For each response $r \in \mathcal{R}$, suppose the residual difference $u_{r,t}(W) = (Q_{r,t} - \beta_r X_{r,t})^{\perp W}$ satisfies

$$u_{r,t}(W) = \ell_r' S_t^{\perp W} + \xi_{r,t}(W), \quad \mathbb{E}[Z_t^{\perp W} \xi_{r,t}(W)] = 0. \quad (18)$$

The vector ℓ_r gives the coefficients on the residualized shocks and state variables in this representation. The remainder $\xi_{r,t}(W)$ contains the variation not captured by these terms. The second condition requires this remainder to be uncorrelated with the residualized instrument, so it contributes nothing to the exclusion moment.

Let Δ_W collect the covariances between the instrument and the shocks and state variables in S_t , after

removing the controls:

$$\Delta_W = \mathbb{E}[Z_t^{\perp W} S_t^{\perp W}]. \quad (19)$$

Stack the rows ℓ'_r into a matrix $L_{\mathcal{R}}$, with one row for each response and one column for each variable in S_t . The exclusion moment for response r is

$$\mathbb{E}[Z_t^{\perp W} u_{r,t}(W)] = \ell'_r \Delta_W.$$

Exclusion for each response therefore restricts the same covariance vector Δ_W through that response's coefficients ℓ_r .

Proposition 2 (Joint identification of policy responses). *Under the representation of $u_{r,t}(W)$ and the zero-covariance restriction on its remainder in (18), exclusion holds for every response in $\mathcal{R}$ if and only if*

$$L_{\mathcal{R}} \Delta_W = 0. \quad (20)$$

Zero residual covariance with every included shock or state variable, $\Delta_W = 0$, is sufficient. It is also necessary if $L_{\mathcal{R}}$ has full column rank. Otherwise, this linear restriction alone does not rule out nonzero covariance vectors. Identification of all responses additionally requires instrument relevance in every equation:

$$\sigma_{ZX,r}(W) \neq 0 \quad \text{for every } r \in \mathcal{R}. \quad (21)$$

The condition $\Delta_W = 0$ is the shock-by-shock exclusion restriction of Stock and Watson (2018) and Plagborg-Møller and Wolf (2021), applied to the shocks and states in S_t after removing the controls. Proposition 2 therefore locates that restriction within the set of conditions that identify a given collection of responses: it is always sufficient, it is stronger than necessary when the net loadings of the responses of interest are linearly dependent, and it becomes necessary once they span the full column space.

Suppose two shocks remain correlated with the residualized instrument. Their contributions to the exclusion moment may cancel for the first-year price response. For a response at another horizon, their coefficients in $u_{r,t}(W)$ may differ, so the same covariances need not cancel again. Identifying both responses requires the common covariance vector Δ_W to satisfy both exclusion restrictions. An additional response adds an independent linear restriction on Δ_W exactly when its row ℓ'_r cannot be written as a linear combination of the existing rows. A proportional row therefore adds no new exclusion restriction.

The first-stage regression relates the normalization variable $X_{r,t}$ to the instrument Z_t and controls W_t . The outcome $Q_{r,t}$ does not enter this regression. Holding these variables fixed therefore leaves the population first stage unchanged when the outcome or its horizon changes, even though the exclusion restriction may change. Estimated first stages also coincide if the LP equations use the same observations. If the normalization changes with the horizon, as when purchases are accumulated over different numbers of quarters, the first-stage dependent variable changes, and its residual covariance with the instrument can change as well.

LP equations need not use identical controls to identify responses to the same policy shock. Under Assumption 1, for a fixed outcome, horizons, and normalization, two control sets that satisfy their respec-

tive relevance and exclusion conditions yield the same population IV coefficient, β_r . Different outcomes or horizons can therefore use different controls, provided relevance and exclusion hold in each equation. These conditions must be evaluated using the residuals from that equation's controls. Appendix A.3 compares the individual and joint identification requirements.

3.4 INTERNAL AND EXTERNAL PROXY-SVAR IDENTIFICATION An internal-instrument VAR includes the instrument among its variables and orders it first, as in Plagborg-Møller and Wolf (2021). Their identifying assumptions identify structural response paths up to a common scale. An external proxy-SVAR uses the instrument to identify a structural shock from the residuals of a VAR that excludes the instrument, as in Mertens and Ravn (2013). Under their structural VAR and proxy assumptions, responses to a unit-variance shock are identified up to sign. I compare each approach's exclusion restrictions with the conditions derived above for individual responses and selected collections of responses. For this comparison, I express the external VAR responses as ratios of responses to the same policy shock.

The restriction in this paper. Fix the controls W_t and suppose $\sigma_{ZX,r}(W) \neq 0$. Because the residualized instrument is orthogonal to the controls, the exclusion restriction (15) can equivalently be written as

$$\mathbb{E}[Z_t^{\perp W}(Q_{r,t} - \beta_r X_{r,t})] = 0. \quad (22)$$

Assumption 1 states that the policy shock's effects are proportional to its size, with response coefficients that do not vary across dates, histories, or realizations of other shocks. Since $\beta_r = \theta_{Q,r}/\theta_{X,r}$, the shock's effects on $Q_{r,t}$ and $\beta_r X_{r,t}$ are equal and therefore cancel in their difference. Exclusion concerns the residualized instrument's covariance with this entire difference. For example, if $\beta_r = 2$, another shock with constant additive effects of two units on $Q_{r,t}$ and one unit on $X_{r,t}$ contributes zero to the difference. Such cancellation can permit correlation with that shock while preserving exclusion. It may fail for another response because the shock's effects or the target ratio differ.

Plagborg-Møller and Wolf: exclusion at all dates. The structural moving-average model in Plagborg-Møller and Wolf (2021) expresses the observed macroeconomic variables as linear functions of current and past structural shocks. Let $\varepsilon_{j,t}$ denote these mean-zero shocks, which are uncorrelated across shock types and dates, with $\varepsilon_{1,t} = m_t$. Let $\mathcal{H}_{t-1}$ be the closed linear span of an intercept and all lags of the macroeconomic variables and the instrument. Write $\tilde{Z}_t^{\mathcal{H}} = Z_t - \mathcal{P}_{\mathcal{H}_{t-1}} Z_t$ for the residual from projection on this full history. Their identifying moment restrictions are

$$\mathbb{E}[\tilde{Z}_t^{\mathcal{H}} m_t] \neq 0, \quad (23)$$

$$\mathbb{E}[\tilde{Z}_t^{\mathcal{H}} \varepsilon_{j,t+k}] = 0 \quad \text{for every } (j,k) \neq (1,0), \quad k \in \mathbb{Z}. \quad (24)$$

After removing the variation explained by past macroeconomic variables and past values of the instrument, the residualized instrument must be correlated with the current target shock. It must be uncorrelated with every other structural shock at every date, and with past and future realizations of the tar-

get shock. These moment restrictions allow measurement noise that is uncorrelated with the structural shocks at all dates.

These restrictions identify every normalized response in their linear model when the LP uses the same full lag history as controls and the normalization variable responds to the target shock. The current target shock cancels from $Q_{r,t} - \beta_r X_{r,t}$, and the residualized instrument is uncorrelated with every remaining shock at every date. The exclusion condition for an individual response is weaker: it requires only that the shock coefficients in $Q_{r,t} - \beta_r X_{r,t}$, multiplied by their respective covariances with the residualized instrument, sum to zero. Some of these covariances can therefore be nonzero if the corresponding coefficient is zero or if contributions from different shocks cancel. Plagborg-Møller and Wolf’s restriction instead requires each covariance to be zero separately.

With unrestricted lag order and matching controls, Plagborg-Møller and Wolf (2021) show that LP-IV and internal-instrument VARs yield the same population response ratios. The LP must condition on the entire past of the variables included in the VAR, including the instrument. For cumulative outcomes or normalizations, the comparison uses the corresponding sums of VAR responses. The equivalence concerns the population ratios produced by the two estimation methods; it does not require those ratios to equal the policy effect. If exclusion fails for response r , both methods produce the same ratio, but that ratio differs from β_r . If the response-specific exclusion condition holds and the first stage is nonzero, both identify β_r , even if some of the stronger restrictions requiring zero covariance with individual shocks fail. No invertibility assumption is required: the target shock need not be recoverable from current and past macroeconomic observations. Appendix B.3 gives the precise conditions, including the treatment of current controls.

Mertens and Ravn: contemporaneous exclusion and a structural VAR. For the comparison with Mertens and Ravn (2013), let $\mathbf{Y}_t$ collect n macroeconomic variables, expressed as deviations from their means. Assume that these variables follow a stable causal structural VAR:

$$\mathbf{Y}_t = \sum_{\ell=1}^p A_\ell \mathbf{Y}_{t-\ell} + \mathbf{u}_t, \quad \mathbf{u}_t = B \varepsilon_t. \quad (25)$$

The vector $\mathbf{u}_t$ contains the one-period linear forecast errors based on past macroeconomic observations. The vector ε_t contains n structural shocks with zero means, unit variances, and zero covariances across different shocks and dates. Its first component is the target policy shock m_t . Column j of B gives the immediate effects of shock j on all variables. Assume that B is nonsingular, so that $\varepsilon_t = B^{-1} \mathbf{u}_t$. Thus all structural shocks can be recovered from current and past macroeconomic observations. This is the invertibility assumption. Miranda-Agrippino and Ricco (2023) show that identification requires only that the target shock be recoverable from the forecast errors, and they derive the resulting bias when the instrument is correlated with other shocks. I keep the nonsingular case to isolate the role of the exclusion restriction. The proxy restrictions stated next identify which combination of forecast errors corresponds to the target shock, up to sign.

For a single target shock, Mertens and Ravn impose

$$\gamma_1 := \mathbb{E}[Z_t m_t] \neq 0, \quad \gamma_j := \mathbb{E}[Z_t \varepsilon_{j,t}] = 0, \quad j = 2, \dots, n. \quad (26)$$

The instrument must be correlated with the current policy shock and uncorrelated with every other contemporaneous structural shock. These contemporaneous restrictions suffice here because the maintained structural VAR expresses current forecast errors as functions of current structural shocks, $\mathbf{u}_t = B\varepsilon_t$. Serial independence of the structural shocks alone would not imply that the instrument is uncorrelated with shocks at other dates. These proxy restrictions make the covariances between Z_t and the VAR forecast errors proportional to the target shock's column of B . These covariances therefore determine the impact responses up to a common scale. The unit-variance assumption for structural shocks, together with the forecast-error covariance matrix $\mathbb{E}[\mathbf{u}_t \mathbf{u}_t'] = BB'$, fixes that scale up to sign. The VAR coefficients then determine the responses at every later horizon. To compare these responses with the estimands in this paper, I divide the response of $Q_{r,t}$ by the response of $X_{r,t}$ to the same shock, assuming the latter is nonzero.

The conditions above identify the entire response path. Identifying a particular response ratio need not require every contemporaneous exclusion restriction. Let $\theta_{Q,r}^j$ and $\theta_{X,r}^j$ denote the responses of $Q_{r,t}$ and $X_{r,t}$ to a one-unit increase in shock j at date t . For the target shock, $\theta_{Q,r}^1 = \theta_{Q,r}$ and $\theta_{X,r}^1 = \theta_{X,r}$. Suppose the target shock changes $X_{r,t}$ and the external proxy-SVAR's population ratio has a nonzero denominator. That ratio equals β_r if and only if

$$\sum_{j=2}^n \underbrace{(\theta_{Q,r}^j - \beta_r \theta_{X,r}^j)}_{\text{net effect of shock } j} \gamma_j = 0. \quad (27)$$

Each term multiplies the instrument's covariance with another current shock by that shock's effect on $Q_{r,t} - \beta_r X_{r,t}$. Mertens and Ravn's exclusion restrictions set each γ_j in this sum to zero. For a particular ratio, a nonzero covariance need not invalidate identification if the corresponding net effect is zero or if its contribution is offset by those of other shocks. This is the analogous cancellation condition for the external estimator under the maintained structural VAR. Appendix B.1 derives its population ratio and bias.

Proposition 3 (Identification of all normalized impact responses). *Maintain (25) and its stated assumptions. Let $D_t = d'\mathbf{Y}_t$, for a fixed known vector d , and let $u_t^D = d'\mathbf{u}_t$ denote its VAR forecast error. Suppose the target shock has a nonzero impact on D_t and $\mathbb{E}[Z_t u_t^D] \neq 0$. Normalize every response by the impact response of D_t . The external proxy-SVAR identifies the normalized impact responses of all n variables if and only if*

$$\gamma_1 \neq 0, \quad \gamma_j = 0 \quad \text{for } j = 2, \dots, n.$$

These conditions also identify the responses at every later horizon, with each response divided by the same impact response of D_t .

An individual impact ratio can be correct because contributions from other shocks cancel. With

nonsingular B , such cancellation cannot make all n impact ratios correct while the instrument remains correlated with another contemporaneous structural shock. Requiring all impact ratios to be correct therefore rules out those correlations. Once the complete normalized impact vector is identified, the VAR coefficients determine the responses at every later horizon. One correct impact ratio alone need not suffice, because a variable's later response can depend on the shock's initial effects on other variables. The necessity of contemporaneous exclusion is the full-column-rank case of Proposition 2. Appendix B.1 gives the proof.

The two approaches differ in their assumptions about both invertibility and the instrument. Plagborg-Møller and Wolf's conditions for LP-IV and the internal-instrument VAR do not require invertibility. They require the residualized instrument to be uncorrelated with all past and future structural shocks and with every current shock except the target shock. Mertens and Ravn's external approach assumes that the structural shocks can be recovered from macroeconomic forecast errors. Under that structural VAR, contemporaneous proxy restrictions identify the target shock's entire response path. Correlation with future shocks can enter an LP's exclusion moment without entering the external estimator's impact moment; Appendix B.2 derives this difference. Neither complete set of assumptions implies the other. Within each maintained model, however, the distinction between conventional and response-specific exclusion is the same: the conventional restrictions set individual shock covariances to zero, while identification of a selected ratio requires only their weighted sum to vanish.

3.5 WHAT CONTROLS MUST ACCOMPLISH To examine the role of controls, let $\mathcal{F}_{t-}$ denote information available to policymakers or markets before the policy decision. For response r , consider a finite vector $R_{r,t}$ of variables drawn from this information. These variables may reflect past policy shocks, but they remain fixed under an intervention that changes the current policy shock m_t . Write

$$Q_{r,t} = \alpha_r + \beta_r X_{r,t} + \Gamma'_r R_{r,t} + \zeta_{r,t}. \quad (28)$$

Here $\zeta_{r,t}$ is what remains of $Q_{r,t} - \beta_r X_{r,t}$ after subtracting $\alpha_r + \Gamma'_r R_{r,t}$. Controlling for $R_{r,t}$ identifies β_r when

$$\mathbb{E}[Z_t^{\perp R_r} X_{r,t}^{\perp R_r}] \neq 0, \quad \mathbb{E}[Z_t^{\perp R_r} \zeta_{r,t}] = 0, \quad (29)$$

where residualization includes an intercept. The instrument must remain correlated with $X_{r,t}$ after controlling for $R_{r,t}$, while being uncorrelated with the remainder $\zeta_{r,t}$. The vector $R_{r,t}$ describes controls that would identify β_r if they were observed and included in the IV regression. The researcher instead uses observed controls W_t , which may omit some components of $R_{r,t}$ or measure them with error. The following result derives the difference between the resulting population IV coefficient and β_r .

Proposition 4 (Identification with feasible controls). *Under (28), let W_t be a finite vector of controls, including an intercept, and suppose the instrument remains relevant, $\sigma_{ZX,r}(W) \neq 0$. Then*

$$b_r(W) - \beta_r = \frac{\Gamma'_r \mathbb{E}[Z_t^{\perp W} R_{r,t}^{\perp W}] + \mathbb{E}[Z_t^{\perp W} \zeta_{r,t}]}{\sigma_{ZX,r}(W)}. \quad (30)$$

The population IV coefficient equals the normalized policy effect β_r if and only if

$$\Gamma_r' \mathbb{E}[Z_t^{\perp W} R_{r,t}^{\perp W}] + \mathbb{E}[Z_t^{\perp W} \zeta_{r,t}] = 0. \quad (31)$$

Sufficient conditions, valid for any coefficient vector Γ_r , are

$$\mathbb{E}[Z_t^{\perp W} R_{r,t}^{\perp W}] = 0, \quad \mathbb{E}[Z_t^{\perp W} \zeta_{r,t}] = 0. \quad (32)$$

If every component of $R_{r,t}$ is a linear combination of W_t , then $R_{r,t}^{\perp W} = 0$, so the first condition holds. Additional controls can nevertheless change the residualized instrument's covariance with $\zeta_{r,t}$. Even if the exclusion restriction holds when the regression controls for $R_{r,t}$, it need not hold after additional controls are included. The example in Appendix A.6, discussed below, shows how adding a control determined before the policy shock can invalidate an otherwise valid IV regression.

The vector $R_{r,t}$ contains the predecision variables that would provide valid controls for response r . The regression instead uses the observed controls W_t . The residuals $R_{r,t}^{\perp W}$ are the variation in those predecision variables that W_t does not explain. The coefficients Γ_r describe how these variables enter $Q_{r,t} - \beta_r X_{r,t}$, the outcome minus the target ratio times the normalization variable. Thus $\Gamma_r' R_{r,t}^{\perp W}$ is their contribution to this difference that remains after controlling for W_t . The first numerator term measures its covariance with the residualized instrument.

The remainder $\zeta_{r,t}$ contains the variation in $Q_{r,t} - \beta_r X_{r,t}$ not accounted for by the intercept and $\Gamma_r' R_{r,t}$. The second numerator term measures its covariance with the residualized instrument. Changing the controls changes the instrument residual and can therefore change this covariance, even though $\zeta_{r,t}$ remains fixed. Identification requires the two numerator terms to sum to zero; they need not vanish separately. Their sum, divided by the residual covariance $\sigma_{ZX,r}(W)$ between the instrument and $X_{r,t}$, gives the difference between the population IV coefficient and β_r .

The coefficients Γ_r can differ across outcomes, horizons, and normalizations. The same omitted predecision variable can therefore contribute differently to bias in two LP equations, even if its residual covariance with the instrument is unchanged. Its contribution is zero if either this covariance or its coefficient in $Q_{r,t} - \beta_r X_{r,t}$ is zero. A small covariance can nevertheless produce a large contribution if the coefficient is large or the residual first-stage covariance $\sigma_{ZX,r}(W)$ is small.

Suppose one component of $R_{r,t}$ is a forecast formed before the policy decision, but the researcher observes only a noisy measure of it. Macroeconomic lags may explain most of the true forecast's variation but little of the measurement error. After removing the variation explained by those lags, the observed forecast can therefore consist largely of measurement error, even if it closely tracks the true forecast overall. Including the noisy forecast does not ensure that the exclusion restriction holds; the resulting IV coefficient may be biased. The Monte Carlo in Section 5 quantifies this case and its consequences for LP-IV bias.

Adding a control can cause the exclusion restriction to fail in an otherwise valid IV regression, even when the added variable is determined before the current policy shock.

Proposition 5 (Adding a predecision control can violate exclusion). *There exist models satisfying Assumption 1 in which the population IV coefficient equals β_r both with only an intercept and with a particular predecision variable included as a control. Starting from the intercept-only specification, adding a different predecision variable violates exclusion even though the instrument remains relevant and the added variable predicts $X_{r,t}$.*

Recall the first-horizon fiscal example in Section 2. The variable $s_{1,t}$ enters the instrument and government purchases, while $s_{2,t}$ adds to GDP without changing purchases. Thus $Y_{t+1} - G_t = s_{2,t}$. Because the instrument is uncorrelated with $s_{2,t}$, IV identifies the unit multiplier without additional controls. Controlling for $s_{1,t}$ alone also preserves identification.

Now include $s_{1,t} + s_{2,t}$ as a control. The regression compares observations holding this sum fixed: a higher value of $s_{1,t}$ must then accompany a lower value of $s_{2,t}$. Although the two variables were initially independent, they are negatively associated within this comparison. Since $s_{1,t}$ enters the instrument, the residualized instrument becomes negatively correlated with $s_{2,t}$, the component of GDP that remains after subtracting purchases. Exclusion therefore fails. The IV coefficient becomes $2/3$, although the true multiplier remains one. Forecast controls have this structure: a GDP forecast aggregates the information that moves the instrument with other determinants of GDP, so conditioning on it can induce the same association between the instrument and the part of the outcome that the normalization does not account for. Appendix A.6 gives a formal construction.

A forecast formed after the policy decision may respond to the current policy shock. Controlling for it can remove variation caused by the shock and alter the residualized instrument's covariance with $Q_{r,t} - \beta_r X_{r,t}$, so the resulting IV coefficient need not equal β_r . The Monte Carlo in Section 5 illustrates this possibility.

Predecision controls cannot remove the instrument's covariance with news that is unpredictable from the predecision information set. Let n_t denote a component of the announcement other than the target policy shock, and suppose $\mathbb{E}[n_t | \mathcal{F}_{t-}] = 0$. This news is uncorrelated with any controls W_t drawn from $\mathcal{F}_{t-}$, so removing those controls from the instrument leaves

$$\mathbb{E}[Z_t^{\perp W} n_t] = \mathbb{E}[Z_t n_t].$$

To see why this matters for identification, recall that β_r measures the outcome's response per unit change in the normalization variable induced by the target policy shock. That shock therefore changes $Q_{r,t}$ and $\beta_r X_{r,t}$ by equal amounts. Its contribution cancels from the difference $Q_{r,t} - \beta_r X_{r,t}$. Other shocks or news can change this difference if they affect the outcome and normalization in different proportions. Suppose n_t enters this difference additively with a nonzero coefficient and is correlated with the instrument. Its contribution to the exclusion moment remains nonzero after controlling for W_t . Exclusion can hold only if covariance contributions from other components offset this term. Lemma 1 establishes the covariance identity.

3.6 CONFIDENCE SETS FOR POLICY RESPONSES When relevance and exclusion hold, the population IV coefficient equals the policy response ratio β_r . I now turn to confidence sets that describe sampling

uncertainty about this ratio. Their construction must account for dependence over time. For example, cumulative outcomes starting in adjacent periods share outcome observations. This overlap can make the regression errors correlated across dates.

For a candidate value β_0 of the policy response ratio, define

$$\psi_{r,t}(\beta_0) = \tilde{Z}_{r,t}(\tilde{Q}_{r,t} - \beta_0 \tilde{X}_{r,t}). \quad (33)$$

This is the residualized instrument multiplied by the part of $Q_{r,t} - \beta_0 X_{r,t}$ that the controls do not explain. Under exclusion, its expectation is zero at the true policy response ratio: $\mathbb{E}[\psi_{r,t}(\beta_r)] = 0$. Replacing the population projections with sample OLS regressions gives $\hat{\psi}_{r,t}(\beta_0)$, the product of the two sample residuals. Its sample average is $\bar{\psi}_r(\beta_0) = T^{-1} \sum_{t=1}^T \hat{\psi}_{r,t}(\beta_0)$. The following assumption states the conditions for a normal approximation to the sample moment and consistent estimation of its variance, accounting for dependence over time.

Assumption 2 (Sampling and long-run variance). *Consider K policy response ratios, $\beta_1, \dots, \beta_K$, specified before estimation. Each ratio is defined by an outcome variable, the horizons and weights used to construct its response, and the normalization. For example, the two-year and four-year cumulative spending multipliers constitute two ratios. The number of ratios, their definitions, the choice of instruments, and the control dimensions remain fixed as the sample size grows.*

- (S1) *The observed process, including the leads and lags used in the regressions, is strictly stationary and ergodic, with finite $4 + \delta$ moments for some $\delta > 0$. Redundant controls are removed, and $\mathbb{E}[W_{r,t} W_{r,t}']$ is positive definite for each response.*
- (S2) *Sample second moments converge in probability to their population values. The OLS coefficients in the projections of Z_t and $Q_{r,t} - \beta_r X_{r,t}$ on the controls have estimation errors of order $O_p(T^{-1/2})$.*
- (S3) *Exclusion holds for every response. Let $\beta = (\beta_1, \dots, \beta_K)'$ and collect the residual products in*

$$\psi_t(\beta) = (\psi_{1,t}(\beta_1), \dots, \psi_{K,t}(\beta_K))'.$$

Their sample sum satisfies

$$T^{-1/2} \sum_{t=1}^T \psi_t(\beta) \xrightarrow{d} N(0, \Omega), \quad \Omega = \sum_{k=-\infty}^{\infty} \mathbb{E}[\psi_t(\beta) \psi_{t-k}(\beta)'].$$

The long-run covariance matrix Ω accounts for dependence across dates and across responses. The covariance series converges absolutely, and every diagonal element of Ω is strictly positive.

- (S4) *The chosen heteroskedasticity and autocorrelation consistent (HAC) estimator, computed from the sample residual products evaluated at the true response ratios, consistently estimates Ω . In particular, for each response,*

$$\hat{\Omega}_r(\beta_r) \xrightarrow{p} \Omega_{rr}.$$

Stationarity and finite moments alone do not establish the normal approximation or consistent variance estimation. The normal approximation requires restrictions on dependence over time, while consistent HAC estimation also requires an appropriate bandwidth. For the Bartlett estimator used here, the bandwidth determines how many lagged covariances enter the variance estimate. Appendix A.7 gives a stationary Gaussian example with finite dependence and a growing bandwidth for which the sampling conditions hold under exclusion.

Proposition 6 (IV and Anderson–Rubin inference). *Suppose Assumption 2 (Sampling and long-run variance) holds. For each response, let*

$$\sigma_{ZX,r} = \mathbb{E}[\tilde{Z}_{r,t}\tilde{X}_{r,t}]$$

denote the covariance between the instrument and the normalization variable after removing the controls.

If each $\sigma_{ZX,r}$ is fixed and nonzero as the sample size grows, the vector of sample IV coefficients, $\hat{b} = (\hat{b}_1, \dots, \hat{b}_K)'$, converges in probability to β , and

$$\sqrt{T}(\hat{b} - \beta) \xrightarrow{d} N(0, \Sigma_{ZX}^{-1} \Omega \Sigma_{ZX}^{-1}), \quad \Sigma_{ZX} = \text{diag}(\sigma_{ZX,1}, \dots, \sigma_{ZX,K}). \quad (34)$$

For a candidate value β_0 , consider the null hypothesis $H_0 : \beta_r = \beta_0$. Under this null, the Anderson–Rubin statistic satisfies

$$AR_r(\beta_0) = \frac{T\tilde{\psi}_r(\beta_0)^2}{\hat{\Omega}_r(\beta_0)} \xrightarrow{d} \chi_1^2. \quad (35)$$

This null distribution does not require $\sigma_{ZX,r} \neq 0$.

The two results use instrument relevance differently. After removing the controls, IV estimation divides the sample covariance between the instrument and the outcome by its sample covariance with the normalization variable. The stated normal approximation requires the denominator to converge to a nonzero value. The Anderson–Rubin test instead fixes a candidate response ratio and tests whether the corresponding exclusion moment is zero. It therefore does not require division by the first-stage covariance. This preserves the null distribution when relevance fails, but it does not establish identification. Under exclusion,

$$\mathbb{E}[\psi_{r,t}(\beta_0)] = (\beta_r - \beta_0)\sigma_{ZX,r}.$$

When $\sigma_{ZX,r} \neq 0$, only the true response ratio makes this moment zero. When $\sigma_{ZX,r} = 0$, every candidate value does so, and the population moment cannot distinguish among them.

An Anderson–Rubin confidence set contains all candidate response ratios that the test does not reject at the chosen significance level. The set can be unbounded or consist of separate intervals; I report the full set in either case.

Residualizing both variables makes the variance calculation account for the controls used in estimation. To see why this matters, suppose a constant is added to every observation of $Q_{r,t}$. Because the controls include an intercept, this change leaves the IV estimate and its sampling uncertainty unchanged. It also leaves the sample exclusion moment unchanged: the residualized instrument sums to zero. But the individual products of the residualized instrument and $Q_{r,t} - \beta_0 X_{r,t}$ change, so a HAC calculation

based on those products can report a different variance. Removing the controls from $Q_{r,t} - \beta_0 X_{r,t}$ before forming the products avoids this problem. The same reasoning applies to variation explained by the other controls. Appendix A.7 gives the formal argument.

For joint inference about K policy response ratios specified before estimation, I use the Bonferroni correction (Romano et al., 2010). I construct a separate Anderson–Rubin confidence set for each ratio using a test with significance level α/K . The joint set contains all vectors whose entries belong to their corresponding individual confidence sets. Under the exclusion and sampling conditions in Assumption 2, the probability that any true ratio is excluded is asymptotically at most the sum of the individual error probabilities, $K(\alpha/K) = \alpha$. The joint set therefore covers all true ratios with asymptotic probability at least $1 - \alpha$, without requiring independence across estimates. Both applications report two response ratios, so I use two 97.5% marginal confidence sets for nominal 95% joint coverage.

3.7 ALLOWING DEPARTURES FROM EXCLUSION I ask which conclusions about policy effects remain supported when the instrument is correlated with other shocks. In the fiscal application, for example, the question is whether spending multipliers remain below one when other shocks correlated with military news also affect output and government purchases. I place bounds on these shocks' contributions to the coefficients on the instrument in regressions of the outcome and normalization variable, and construct confidence sets for the policy response ratios under those bounds. For a candidate response ratio β_0 , consider the population regression of $Q_{r,t} - \beta_0 X_{r,t}$ on the controls $W_{r,t}$, including the intercept, and the instrument Z_t . Let $\sigma_{Z,r}^2 = \mathbb{E}[\tilde{Z}_{r,t}^2] > 0$ denote the variance of the residualized instrument. The coefficient on the instrument, $a_r(\beta_0)$, and the regression residual, $e_{r,t}(\beta_0)$, are

$$\begin{aligned} a_r(\beta_0) &= \sigma_{Z,r}^{-2} \mathbb{E}[\tilde{Z}_{r,t}(\tilde{Q}_{r,t} - \beta_0 \tilde{X}_{r,t})], \\ e_{r,t}(\beta_0) &= (\tilde{Q}_{r,t} - \beta_0 \tilde{X}_{r,t}) - a_r(\beta_0) \tilde{Z}_{r,t}. \end{aligned} \tag{36}$$

Under exclusion, $a_r(\beta_r) = 0$. Testing this zero-coefficient restriction with a HAC standard error gives the regression form of the Anderson–Rubin test. Its squared t -statistic has the same χ_1^2 null limit as the moment-based statistic when their respective variance estimators are consistent, although the statistics can differ in finite samples. The sensitivity analysis replaces the zero restriction with values implied by the bounds on departures from exclusion.

The sensitivity analysis must distinguish a nonzero population coefficient from sampling error in its estimate. Let $\hat{a}_r(\beta_0)$ denote the estimated coefficient on the instrument in the regression above, and let $\hat{\text{se}}_r(\beta_0)$ denote its HAC standard error. Assume that the regressor vector $(W'_{r,t}, Z_t)'$ has fixed dimension and a positive definite population second-moment matrix, after removing redundant controls, and that sample second moments converge to their population values. For response r , assume that the following approximation holds at the true policy response ratio $\beta_0 = \beta_r$:

$$\sqrt{T}\{\hat{a}_r(\beta_0) - a_r(\beta_0)\} = \sigma_{Z,r}^{-2} T^{-1/2} \sum_t \tilde{Z}_{r,t} e_{r,t}(\beta_0) + o_p(1) \xrightarrow{d} N(0, V_r(\beta_0)). \tag{37}$$

Here $V_r(\beta_0)$ is the limiting variance of the scaled estimation error. Require $0 < V_r(\beta_0) < \infty$ and

$$T\widehat{\text{se}}_r(\beta_0)^2 \xrightarrow{P} V_r(\beta_0).$$

These conditions describe sampling uncertainty around $a_r(\beta_0)$, including when it is nonzero. The variance calculation uses $e_{r,t}(\beta_0)$, the residual after removing the contributions of both the controls and the instrument.

The sensitivity analysis places bounds on contamination of the coefficients on the instrument in the outcome and normalization regressions. By contamination, I mean the contributions to these coefficients from sources of variation other than the current policy shock. Under Assumption 1,

$$Q_{r,t} = \theta_{Q,r} m_t + Q_{r,t}(0), \quad X_{r,t} = \theta_{X,r} m_t + X_{r,t}(0).$$

The first term in each equation is the current policy shock's contribution to that variable. The second contains variation due to the prior history and other shocks. The coefficients on the instrument in regressions of these second components define the contamination that the sensitivity analysis bounds. Let $\delta_{Q,r}$ and $\delta_{X,r}$ denote the population coefficients on Z_t in regressions of $Q_{r,t}(0)$ and $X_{r,t}(0)$, respectively, on the instrument and controls:

$$\delta_{Q,r} = \sigma_{Z,r}^{-2} \mathbb{E}[\tilde{Z}_{r,t} \tilde{Q}_{r,t}(0)], \quad \delta_{X,r} = \sigma_{Z,r}^{-2} \mathbb{E}[\tilde{Z}_{r,t} \tilde{X}_{r,t}(0)].$$

Here $\tilde{Q}_{r,t}(0) = [Q_{r,t}(0)]^{\perp W_r}$ and $\tilde{X}_{r,t}(0) = [X_{r,t}(0)]^{\perp W_r}$ are the residuals from projecting the zero-shock outcomes on the original controls. The instrument and controls are the same random variables used in the original regressions. The expectations average over their joint distribution with the counterfactual dependent variables.

Linearity of regression separates each instrument coefficient into the contribution of the current policy shock and the corresponding δ . At the true response ratio, subtracting β_r times the normalization coefficient from the outcome coefficient cancels the policy shock's contributions, leaving

$$a_r(\beta_r) = \delta_{Q,r} - \beta_r \delta_{X,r}. \tag{38}$$

This equation shows how the two contributions determine the departure from exclusion.

In the fiscal application, the estimated multiplier divides the GDP coefficient on military news by the purchases coefficient. Other shocks correlated with news can therefore change the estimated multiplier through either its numerator or its denominator. Bounding the GDP contribution while setting the purchases contribution to zero would assume that other sources of variation make no net contribution to the purchases coefficient. I instead bound both contributions and examine whether the below-one conclusion survives the changes in both coefficients allowed by those bounds.

Beyond the data and regression specification, computing the confidence sets requires a joint set $\mathcal{D}$ of contamination coefficients and a chosen joint confidence level $1 - \alpha$. The set $\mathcal{D}$ collects all combinations of contamination coefficients that satisfy the assumed bounds. One element specifies an outcome

contribution and a normalization contribution for every response ratio under consideration. Each contribution is in the units of a coefficient the researcher already estimates: $\delta_{Q,r}$ is a shift in the coefficient on the instrument in the outcome regression, and $\delta_{X,r}$ is a shift in the first-stage coefficient. A bound can therefore be stated as a fraction of the estimated coefficient, as in Section 4, or elicited directly as the largest effect that other shocks correlated with the instrument could plausibly have on the outcome or normalization variable per unit of the instrument. Dividing a contamination of the outcome coefficient by the first-stage coefficient expresses it in units of the response ratio itself, so bounds can also be chosen as the largest bias in the policy effect that one is willing to entertain. Varying the bounds shows how the conclusions depend on the assumed contamination.

One way to specify $\mathcal{D}$ is to place a separate bound on contamination of each coefficient. For illustration, consider one response ratio and suppose that other sources of variation contribute at most 0.05, in absolute value, to the outcome coefficient on the instrument and at most 0.02 to the normalization coefficient. These assumptions give

$$|\delta_Q| \leq 0.05, \quad |\delta_X| \leq 0.02,$$

so

$$\mathcal{D} = [-0.05, 0.05] \times [-0.02, 0.02].$$

The set includes every combination within these two ranges. When testing a response ratio of one, for example, the contamination term $\delta_Q - \delta_X$ can range from -0.07 to 0.07 . The test retains one if the estimated instrument coefficient is statistically compatible with at least one value in this range.

For the two fiscal multipliers, this construction requires four bounds: one for the GDP coefficient and one for the purchases coefficient at each horizon. No additional restrictions linking the four contributions are needed. I vary the bounds to determine when the conclusion that multipliers are below one changes.

Given these inputs and a specified HAC variance estimator and bandwidth, the calculation uses the observed regressions to obtain $\hat{a}_r(\beta_{0,r})$ and its standard error for each candidate response ratio. The vector $\beta_0 = (\beta_{0,1}, \dots, \beta_{0,K})'$ contains candidate values of the policy response ratios. The confidence set $\mathcal{C}(\mathcal{D})$ collects the candidate vectors that the data do not reject under at least one combination of contamination coefficients in $\mathcal{D}$. The same combination must account for all K regressions simultaneously. Computing the set requires no estimates of the policy shock, the counterfactual outcomes, or the responses to individual contaminating shocks.

Proposition 7 (Confidence sets under bounded exclusion errors). *Consider K policy response ratios and let $\mathcal{D} \subseteq \mathbb{R}^{2K}$ be the set of contamination coefficient vectors satisfying the joint sensitivity bounds. Assume that $\mathcal{D}$ is nonrandom, specified before estimation, and contains the true coefficient vector.*

Suppose Assumption 1 holds. For $r = 1, \dots, K$, assume that the regression normal approximation in (37) and its variance-consistency condition hold at the true ratio β_r . Define the two-sided Bonferroni critical value

$$c_{\alpha,K} = \Phi^{-1}\left(1 - \frac{\alpha}{2K}\right),$$

where Φ is the standard normal distribution function.

Include a candidate vector $\beta_0 = (\beta_{0,1}, \dots, \beta_{0,K})'$ in the confidence set $\mathcal{C}(\mathcal{D})$ if at least one joint contamination vector $(\delta_{Q,r}, \delta_{X,r})_{r=1}^K \in \mathcal{D}$ satisfies

$$\frac{|\hat{a}_r(\beta_{0,r}) - \delta_{Q,r} + \beta_{0,r} \delta_{X,r}|}{\hat{\text{se}}_r(\beta_{0,r})} \leq c_{\alpha,K}, \quad r = 1, \dots, K. \quad (39)$$

The resulting set contains the true vector of policy response ratios with asymptotic probability at least $1 - \alpha$. This result does not require the residual first-stage covariances $\sigma_{ZX,r}$ to be nonzero.

Suppose first that the true contamination coefficients were known. At the true response ratios, each test subtracts the population instrument coefficient from its estimate. Each test therefore rejects its true response ratio with asymptotic probability α/K . The Bonferroni correction ensures that all K true ratios are retained with asymptotic probability at least $1 - \alpha$, without requiring independence across tests. When the contamination coefficients are unknown, the procedure considers every combination in $\mathcal{D}$. If this set contains the true coefficients, the resulting confidence set includes every response vector retained using those coefficients. Considering additional combinations can enlarge the confidence set but cannot remove those vectors. Appendix A.8 gives the formal proof.

The confidence sets describe what the data imply under the assumed contamination bounds. Economic evidence about other sources of variation associated with the instrument can help assess whether these bounds are plausible. The coverage result applies when the bounds are specified before estimation and contain the true contamination coefficients. Bounds estimated or selected using the data require a separate argument that accounts for that estimation or selection. This construction follows the relaxed-exclusion approach of Conley et al. (2012), applied jointly to contamination of the outcome and normalization coefficients.

4 FISCAL MULTIPLIERS AND THE BELOW-ONE CONCLUSION

This section asks how much contamination of the military-news instrument would overturn the conclusion that cumulative spending multipliers are below one. Using the data and specification of Ramey and Zubairy (2018), the confidence sets for the two- and four-year multipliers lie below one when exclusion holds. Other shocks correlated with military news contribute to both the output and the purchases coefficients. A multiplier of one enters the two-year set once their net contribution biases the multiplier by 0.03, and the four-year set at 0.045. This section derives these thresholds and the combinations of output and purchases bounds that preserve the conclusion at each horizon.

4.1 HISTORICAL DESIGN AND ESTIMATION Military news measures revisions to the expected present value of government purchases associated with political and military events. Because this news can precede actual purchases, the impact response of purchases need not capture the size of the fiscal intervention (Ramey, 2011). Following Ramey and Zubairy (2018), I define the multiplier as the cumulative GDP response divided by the cumulative government-purchases response over the same window, without discounting either sum. I use their linear specification and public replication data. The multiplier includes

the effects of any subsequent spending, tax, and monetary-policy responses induced by the news shock.

Following Ramey and Zubairy, I divide real GDP and real government purchases by current potential real GDP and nominal military news by lagged nominal potential GDP. Let Y_t denote real GDP, G_t nominal government purchases, P_t the GDP deflator, and $\tilde{Y}_t$ the potential real GDP series supplied in their replication data. Let News_t denote nominal military spending news. The regression variables are

$$y_t = \frac{Y_t}{\tilde{Y}_t}, \quad g_t = \frac{G_t/P_t}{\tilde{Y}_t}, \quad Z_t = \frac{\text{News}_t}{P_{t-1} \tilde{Y}_{t-1}}. \quad (40)$$

Government purchases include federal, state, and local purchases and exclude transfers. For windows of $H = 8$ or 16 quarters, let $Q_{H,t}$ and $X_{H,t}$ denote cumulative normalized GDP and purchases. The structural multiplier M_H is the ratio of their responses to the spending-news shock:

$$Q_{H,t} = \sum_{j=0}^{H-1} y_{t+j}, \quad X_{H,t} = \sum_{j=0}^{H-1} g_{t+j}, \quad M_H = \frac{\sum_{j=0}^{H-1} \theta_{y,j}}{\sum_{j=0}^{H-1} \theta_{g,j}}. \quad (41)$$

The coefficients $\theta_{y,j}$ and $\theta_{g,j}$ are the responses of normalized GDP and purchases, respectively, j quarters after the spending-news shock. Both windows start in the news quarter, $j = 0$: $H = 8$ covers quarters 0–7, and $H = 16$ covers quarters 0–15. Ramey and Zubairy index the last included quarter by h , so $h = H - 1$. I define a positive shock as an upward revision to the expected present value of government purchases and assume that it raises cumulative purchases over both windows.

The control vector W_t contains an intercept and four lags each of y_t , g_t , and Z_t . For each window H , I estimate the cumulative LP-IV regression of Ramey and Zubairy (2018):

$$Q_{H,t} = b_H X_{H,t} + \gamma'_H W_t + e_{H,t}. \quad (42)$$

I use two-stage least squares, instrumenting cumulative purchases $X_{H,t}$ with military news Z_t . The first stage regresses cumulative purchases $X_{H,t}$ on military news Z_t and the controls W_t . The reduced form regresses cumulative GDP $Q_{H,t}$ on the same variables, using the same observations:

$$X_{H,t} = a_{G,H} Z_t + \gamma'_{G,H} W_t + e_{G,H,t}, \quad (43)$$

$$Q_{H,t} = a_{Y,H} Z_t + \gamma'_{Y,H} W_t + e_{Y,H,t}. \quad (44)$$

With one instrument, the IV estimate is the ratio of the estimated coefficients on news,

$$\hat{b}_H = \frac{\hat{a}_{Y,H}}{\hat{a}_{G,H}},$$

provided $\hat{a}_{G,H} \neq 0$.

I estimate cumulative LP-IV equations for the two- and four-year horizons using the same 485 starting quarters, from 1891Q1 through 2012Q1. The four-year equation uses outcome data through 2015Q4. Both equations use the same instrument and controls. Their first stages differ because the variable being instrumented is cumulative purchases over two years in one equation and four years in the other.

Appendix C documents the data construction and the small differences from the estimates saved in the replication archive.

Let $\tilde{Z}_t$, $\tilde{Q}_{H,t}$, and $\tilde{X}_{H,t}$ denote the residuals from the population linear projections of the instrument, cumulative GDP, and cumulative purchases on W_t . The instrument residual is common to both horizons because they use the same controls and estimation dates. Relevance requires the instrument to be correlated with cumulative purchases after removing the controls, $\mathbb{E}[\tilde{Z}_t \tilde{X}_{H,t}] \neq 0$. Given this relevance condition, the population IV coefficient b_H equals the structural multiplier M_H if and only if exclusion holds:

$$\mathbb{E}[\tilde{Z}_t(\tilde{Q}_{H,t} - M_H \tilde{X}_{H,t})] = 0. \quad (45)$$

After removing the controls, the instrument's covariance with cumulative GDP must equal M_H times its covariance with cumulative purchases. Other shocks correlated with the residualized instrument can contribute to both covariances. An exclusion argument must therefore consider their contributions to both output and purchases. This restriction can hold in the LP equation for one horizon and fail in the other.

Because the spending-news shock raises cumulative g_t , a multiplier below one means that the cumulative response of y_t is smaller than that of g_t . Government purchases are part of GDP. Thus, $y_t - g_t$ equals private consumption plus private investment plus net exports, with all components measured using the same deflator and normalization. A multiplier below one therefore implies a negative cumulative response of these other components taken together. It does not establish which component falls.

4.2 ESTIMATES UNDER THE EXCLUSION RESTRICTION I construct confidence sets by inverting Anderson–Rubin tests (Anderson and Rubin, 1949), which do not require a strong first stage (Andrews et al., 2019). Ramey and Zubairy (2018) use these tests to compare fiscal multipliers across economic states. For each candidate multiplier β_0 , I regress $Q_{H,t} - \beta_0 X_{H,t}$ on W_t and Z_t . The estimated coefficient on news is $\hat{a}_H(\beta_0) = \hat{a}_{Y,H} - \beta_0 \hat{a}_{G,H}$. Under exclusion, the population coefficient on news is zero when β_0 equals the true multiplier. The confidence set contains all values of β_0 for which the test does not reject this restriction. For joint inference about the two- and four-year multipliers, I use a Bonferroni correction (Romano et al., 2010) and test each horizon at the nominal 2.5% significance level.² With a two-sided test, the confidence set includes β_0 when

$$|\hat{a}_H(\beta_0)| \leq c \hat{se}_H(\beta_0), \quad c = \Phi^{-1} \left(1 - \frac{0.05}{2 \times 2} \right) = \Phi^{-1}(0.9875),$$

where Φ is the standard-normal distribution function. The standard error $\hat{se}_H(\beta_0)$ of the estimated news coefficient allows for heteroskedasticity and serial correlation. I compute it from the regression residuals using Bartlett weights with 16 quarterly lags. I multiply the estimated covariance matrix by $T/(T-14)$, where $T = 485$ and the 14 regressors are an intercept, 12 lagged controls, and the instrument.

²Two separate 95% confidence sets need not cover both true multipliers with probability 95%. Using two 97.5% sets, the Bonferroni bound limits the probability that at least one set misses its multiplier to at most $2.5\% + 2.5\% = 5\%$. The argument does not require the two estimates to be independent. Here the bound is asymptotic and rests on the identifying and sampling assumptions.

Table 2: Cumulative spending multipliers under the exclusion restriction

Window	IV estimate	First-stage F	97.5% AR set
Two years	0.665	17.94	[0.545, 0.939]
Four years	0.713	11.66	[0.619, 0.890]

Notes: Two-stage least squares regresses cumulative output on cumulative government purchases, instrumented with military spending news. All variables use the potential-GDP normalization in (40). The windows include quarters 0–7 and 0–15. Both regressions use 485 origins, 1891Q1–2012Q1, an intercept, and four lags each of normalized output, purchases, and news. Bartlett HAC uses 16 quarterly lags and the covariance correction $T/(T - 14)$. The product of the two Anderson–Rubin sets has asymptotic coverage of at least 95% under the maintained exclusion and sampling conditions.

Table 2 reports cumulative multiplier estimates of 0.665 over two years and 0.713 over four years. Under exclusion, both confidence sets lie below one. Both remain below one when I replace the baseline HAC bandwidth of 16 quarters with eight or 24 quarters.

Most of the instrument’s variation after removing the controls comes from a few wartime quarters. Five quarters account for 66.8% of the sum of squared military-news residuals: 1917Q2 during World War I; 1941Q4 and 1942Q3 during World War II; and 1950Q3 and 1950Q4 during the Korean War. These dates warrant close attention when assessing exclusion. Other wartime shocks correlated with military news can contribute to the output and purchases coefficients. An exclusion argument must explain why these contributions do not bias the ratio of those coefficients. Controlling for lagged GDP, purchases, and news does not by itself establish this restriction. The sample also spans different fiscal and monetary arrangements. A common multiplier across these regimes, and its applicability to modern peacetime purchases, are assumptions about policy transmission that the regression does not establish.

4.3 SENSITIVITY TO VIOLATIONS OF EXCLUSION I assess how violations of exclusion affect the below-one conclusion. Let $Q_{H,t}(0)$ and $X_{H,t}(0)$ denote cumulative normalized GDP and purchases if the current spending-news shock were set to zero, holding the prior history and all other innovations fixed. Let $\tilde{Q}_{H,t}(0)$ and $\tilde{X}_{H,t}(0)$ denote the residuals from their population linear projections on the observed controls W_t . Keeping the observed instrument and controls, define

$$\delta_{Y,H} = \frac{\mathbb{E}[\tilde{Z}_t \tilde{Q}_{H,t}(0)]}{\mathbb{E}[\tilde{Z}_t^2]}, \quad \delta_{G,H} = \frac{\mathbb{E}[\tilde{Z}_t \tilde{X}_{H,t}(0)]}{\mathbb{E}[\tilde{Z}_t^2]}.$$

These are the population coefficients on military news in regressions of the counterfactual GDP and purchases variables on news and the controls. They measure the association with news that remains after removing the effects of the current spending-news shock. Their units are cumulative normalized GDP or purchases per unit of normalized military news. Because the counterfactual outcomes are unobserved, the sensitivity analysis places bounds on these coefficients. At the true multiplier M_H , subtracting M_H times the purchases coefficient from the GDP coefficient cancels the target shock’s contribution, leaving

$$a_H(M_H) = a_{Y,H} - M_H a_{G,H} = \delta_{Y,H} - M_H \delta_{G,H}.$$

Exclusion requires $\delta_{Y,H} = M_H \delta_{G,H}$. At a multiplier of one, this restriction requires equal contributions to the two news coefficients; both contributions may be nonzero. The difference $\delta_{Y,H} - \delta_{G,H}$ is the coefficient on observed military news in the population regression of $Q_{H,t}(0) - X_{H,t}(0)$ on news and W_t . By GDP accounting, this counterfactual difference equals cumulative private consumption plus private investment plus net exports when the current spending-news shock is set to zero, measured in the same normalized units.

I use sensitivity bounds to assess whether the below-one conclusion survives departures from exclusion. For each specified set of bounds, I construct adjusted confidence sets. A multiplier is included if the Anderson–Rubin test does not reject it for at least one allowed combination of contributions from sources of variation other than the current spending-news shock. If both confidence sets lie entirely below one, the conclusion survives every combination allowed by the bounds. Specifically, I impose fixed, nonnegative bounds

$$|\delta_{Y,H}| \leq d_{Y,H}, \quad |\delta_{G,H}| \leq d_{G,H}, \quad H \in \{8, 16\}. \quad (46)$$

The bound $d_{Y,H}$ limits the absolute contribution of sources of variation other than the current spending-news shock to $a_{Y,H}$, the population coefficient on military news in the regression of cumulative GDP $Q_{H,t}$ on Z_t and W_t . The bound $d_{G,H}$ imposes the corresponding limit on $a_{G,H}$, the population coefficient on military news in the regression of cumulative purchases $X_{H,t}$ on the same instrument and controls. These contributions may be positive or negative. Larger bounds permit a wider range of contributions and may therefore admit more multiplier values. I vary the bounds to determine which choices preserve the below-one conclusion. Within the four bounds, I allow any combination of contributions across the two LP equations. Setting all four bounds to zero gives the baseline Anderson–Rubin confidence sets, although exclusion itself also permits nonzero contributions that cancel.

Suppose the true multiplier is one and the population coefficient on news in the purchases regression, $a_{G,H}$, is positive. A negative difference $\delta_{Y,H} - \delta_{G,H}$ then makes the population IV coefficient smaller than one. For example, another shock could be positively correlated with residual military news and raise cumulative purchases without changing cumulative GDP. This shock increases the denominator of the IV ratio without increasing its numerator. By GDP accounting, the increase in purchases is offset by a decline in the combined cumulative value of private consumption, private investment, and net exports. The bias comes from the instrument's correlation with this other shock. Any crowding out caused by the target spending-news shock itself is part of its multiplier.

I ask whether the conclusion that the two- and four-year spending multipliers are below one survives specified departures from exclusion. To answer this question, I construct confidence sets that allow the contributions of sources of variation other than the current spending-news shock to range within the sensitivity bounds. I test each candidate multiplier for all combinations of $\delta_{Y,H}$ and $\delta_{G,H}$ that satisfy the sensitivity bounds, retaining it if the test does not reject it for at least one combination.

Consider the test of a multiplier of one. For a given pair of assumed contributions, I use the regression of cumulative GDP less cumulative purchases, $Q_{H,t} - X_{H,t}$, on military news Z_t and controls W_t . Under a multiplier of one and the assumed contributions, the population coefficient on news equals $\delta_{Y,H} - \delta_{G,H}$.

The Anderson–Rubin test therefore does not reject a multiplier of one for this pair when

$$|\hat{a}_H(1) - (\delta_{Y,H} - \delta_{G,H})| \leq c \hat{se}_H(1), \quad (47)$$

where c and $\hat{se}_H(1)$ are the same critical value and standard error used in the baseline test of a multiplier of one. The standard error uses the residuals from regressing $Q_{H,t} - X_{H,t}$ on military news and the controls, with the news coefficient estimated by OLS. Changing $\delta_{Y,H}$ and $\delta_{G,H}$ changes the hypothesized news coefficient, $\delta_{Y,H} - \delta_{G,H}$, while leaving the regression estimates and standard error unchanged.

The joint confidence set includes a pair of two- and four-year multipliers if there is a combination of the four contributions satisfying the sensitivity bounds for which neither test rejects. Suppose these bounds are fixed and contain the true contributions. The joint set then includes every multiplier pair that would be retained if the true contributions were known. The Bonferroni correction therefore gives at least 95% asymptotic coverage of the true multiplier pair under the stated regression sampling assumptions. Proposition 7 establishes this result.

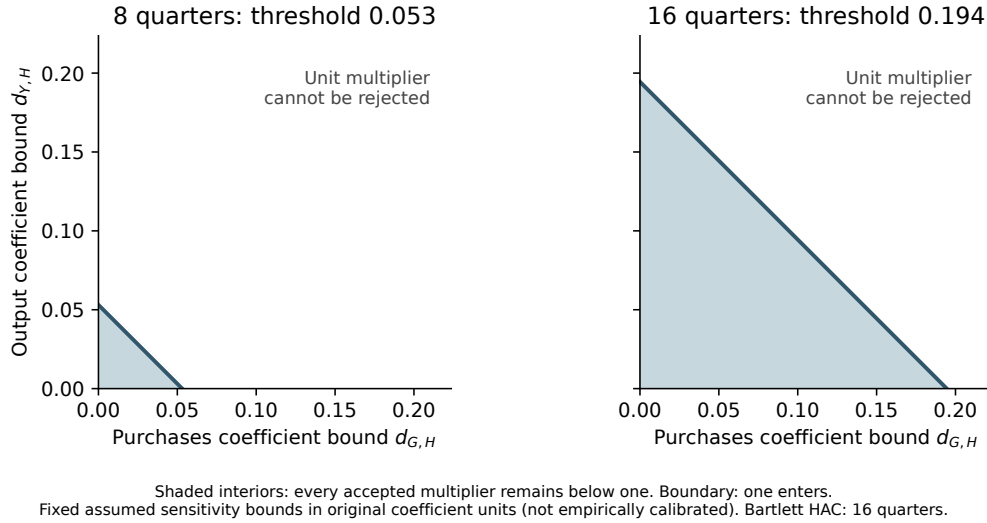
I next calculate how large the sensitivity bounds must be for the confidence set to include a multiplier of one. At this multiplier, the difference $\delta_{Y,H} - \delta_{G,H}$ can take any value between $-(d_{Y,H} + d_{G,H})$ and $d_{Y,H} + d_{G,H}$. The test retains one if this interval contains a value within $c \hat{se}_H(1)$ of the estimated news coefficient $\hat{a}_H(1)$. The smallest sum of bounds that satisfies this condition is

$$d_H^* = \max\{0, |\hat{a}_H(1)| - c \hat{se}_H(1)\}. \quad (48)$$

I call d_H^* the *sensitivity threshold*. The confidence set includes one if and only if $d_{Y,H} + d_{G,H} \geq d_H^*$. The line $d_{Y,H} + d_{G,H} = d_H^*$ is the *sensitivity frontier*. For these estimates and the stated bounds, Appendix C also establishes that every multiplier greater than one is excluded whenever the sum of bounds lies below this threshold. The frontier therefore marks where the confidence set ceases to lie entirely below one.

Figure 1 shows the combinations of output and purchases bounds for which each confidence set remains below one. The thresholds are $d_8^* = 0.0531$ over two years and $d_{16}^* = 0.1945$ over four years. Along each frontier, an increase in the output bound $d_{Y,H}$ requires an equal reduction in the purchases bound $d_{G,H}$. Dividing each threshold by the estimated coefficient on military news in the corresponding cumulative purchases regression gives 3.12% and 4.50%. This ratio is in multiplier units: at a true multiplier of one, the bias in the IV coefficient equals $(\delta_{Y,H} - \delta_{G,H})$ divided by the population first-stage coefficient, so the ratio measures the bias in the multiplier that contamination at the threshold would induce. The below-one conclusion therefore survives only if other shocks correlated with military news bias the two-year multiplier by less than 0.03 and the four-year multiplier by less than 0.045. These margins are small relative to the point estimates of 0.665 and 0.713 and to the upper confidence bounds of 0.939 and 0.890: sampling uncertainty already accounts for most of the distance between the point estimates and one. If the sum of bounds increases as the same percentage of the first-stage coefficient at both horizons, the two-year confidence set reaches one first. This comparison does not establish that exclusion is more plausible over four years. Other shocks correlated with military news can have different cumulative effects on GDP and purchases at the two horizons.

Figure 1: Sensitivity bounds that preserve a multiplier below one



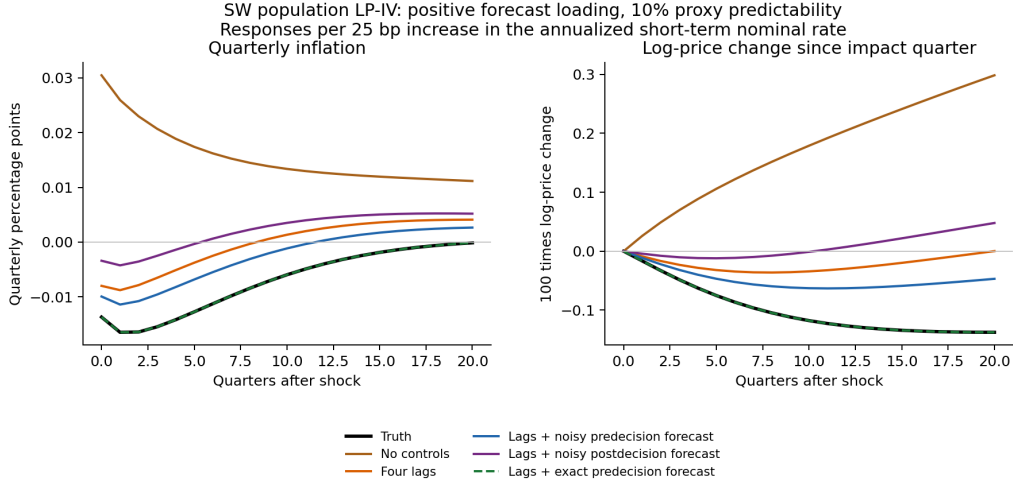
Notes: Each panel varies assumed upper bounds on the absolute nuisance contributions to the cumulative output and purchases coefficients on military news, conditional on the controls in Table 2. Below the boundary, every multiplier in the projected marginal confidence set is less than one; on the boundary, a unit multiplier enters. The calculation uses the common 485-origin sample, a 16-quarter Bartlett bandwidth, and 97.5% marginal tests for the two-window 95% procedure. The axes describe assumed bounds. They do not locate actual exclusion violations.

To defend the below-one conclusion under departures from exclusion, a researcher must justify the bounds used in the sensitivity analysis. These bounds limit contributions from sources of variation other than the current spending-news shock to the population coefficients on military news in the cumulative GDP and purchases regressions. The historical estimates determine which combinations of bounds preserve the conclusion, but do not establish the size or sign of those contributions. The reported percentages express the thresholds relative to estimated purchases coefficients. Using those estimates to choose the population bounds would require additional inference to account for their uncertainty. The economic argument for the bounds must address both LP equations: a bound justified for the two-year equation need not be justified for the four-year equation, even though both use the same instrument and controls.

5 MONETARY TRANSMISSION AND THE CHOICE OF CONTROLS

This section asks whether the controls used in practice, lags and forecasts, remove contamination from past shocks. In the model of Smets and Wouters (2007) they do so only in part. Four lags leave the first-year price decline understated by 55% and the later decline by 98%. A noisy forecast formed before the policy decision removes about a third of each bias, and the exact forecast removes all of it. A forecast formed after the decision introduces bias even with a valid instrument. Across these specifications the first-stage covariance moves by about 3%. I construct an instrument that combines the current monetary policy innovation, an inflation forecast formed before the policy decision, and independent measurement noise. The instrument is uncorrelated with every other current structural innovation. Its

Figure 2: Incomplete controls distort the persistence of the log-price response



Notes: Structural responses and exact population IV projections, normalized to a 25-basis-point increase in the annualized short-term nominal interest rate. The left panel reports quarterly inflation in percentage points; the right reports changes in 100 times the log price level since the impact quarter, excluding impact-quarter inflation. All LP-IV curves use the same contaminated instrument, with forecast loading $\lambda = \sqrt{2/9}$ and independent noise standard deviation one. The forecast accounts for 10% of unconditional instrument variance. The curves differ only in their controls. With an uncontaminated instrument, $\lambda = 0$, LP-IV identifies the structural responses without controls. Observable controls contain four lags each of output, inflation, and the annualized short-term nominal interest rate. The exact forecast control recovers the structural response. Forecast measurements and their timing are defined below.

contamination arises from past shocks reflected in the forecast. These shocks continue to affect both future prices and the current nominal interest rate used to normalize the policy response. They therefore contribute to the coefficients on the instrument in both the outcome and normalization regressions.

Four lags of output, inflation, and nominal interest rates explain 86.6% of the forecast's variance. These controls do not fully capture the model state used to form the forecast, so its unexplained component remains in the residualized instrument.

Figure 2 compares the model's true inflation and log-price responses with population LP-IV responses under different controls. All LP-IV responses in the figure use the same contaminated instrument, with forecast loading $\lambda = \sqrt{2/9}$. The curves differ only in their controls. With an uncontaminated instrument, $\lambda = 0$, LP-IV identifies the structural responses without controls. The true log-price response continues to decline after the first year. Without controls, the instrument's correlation with past shocks is strong enough that LP-IV incorrectly implies price increases during both the first year and years two and three following a monetary tightening. With lagged controls, the population LP-IV coefficients incorrectly imply that the decline ends in the second year and is then reversed. Adding a forecast measured with error and formed before the policy decision reduces these errors, but the response still understates the decline and reverses too early. Adding the exact forecast recovers the true response. A noisy forecast formed after the policy decision produces a brief decline followed by an increase in the log price level relative to its impact-quarter response. These comparisons hold the model and policy shock fixed. The discrepancies from the true responses reflect identification bias rather than sampling error.

5.1 THE MONETARY EXPERIMENT AND THE ECONOMIC CLAIMS I use a first-order approximation to the Smets–Wouters model at the reported posterior mode for 1966Q1–2004Q4, retaining the original values of the calibrated parameters.³ All parameters remain fixed throughout the experiment. The simulated data include all seven structural shocks, with Gaussian innovations. To define the policy response, I increase the current monetary policy innovation m_t by one standard deviation. I calculate the true responses and population LP-IV coefficients from the model solution, then use repeated simulated samples to study the sampling distribution of the LP-IV estimator.

I report responses scaled to a 25-basis-point increase in the model's short-term nominal interest rate. Let i_t denote this rate in quarterly percentage points. The normalization variable expresses it in annualized percentage points:

$$X_t = 4i_t. \quad (49)$$

The two outcomes measure the change in the log price level during the first year and the additional change during years two and three. Let π_t denote the model's quarterly inflation deviation from steady state, and define p_t by $p_t - p_{t-1} = \pi_t$. In the first-order model, cumulating these inflation deviations gives changes in 100 times the log price level relative to the path implied by steady-state inflation. The outcomes are

$$Q_{E,t} = p_{t+4} - p_t = \sum_{j=1}^4 \pi_{t+j}, \quad Q_{L,t} = p_{t+12} - p_{t+4} = \sum_{j=5}^{12} \pi_{t+j}. \quad (50)$$

Both exclude inflation in the impact quarter. These changes in 100 times log prices are approximately percentage price changes.

I multiply each response ratio by 0.25 to express it per 25-basis-point increase in the annualized short-term nominal interest rate. The true responses imply approximate price declines of 0.0625% during the first year and a further 0.0642% during years two and three. The additional decline is therefore approximately as large as the initial decline. Both LP equations use the same current normalization variable, instrument, controls, and estimation dates. Their first stages are therefore identical within each control specification.

I construct the instrument as

$$Z_t = m_t + \lambda f_t + v_t, \quad (51)$$

where m_t is the current monetary policy innovation, standardized to have unit variance, and v_t is independent Gaussian measurement noise. The variable f_t is the model's forecast of the first-year log-price change $Q_{E,t}$, formed before any date- t innovations and scaled to have unit unconditional variance.⁴ It therefore reflects past shocks. A positive λ raises the instrument when this forecast predicts higher inflation over the following year. The past shocks reflected in the forecast also affect future log prices and the

³The model code and parameter file come from Johannes Pfeifer's [DSGE_mod repository](#).

⁴The forecast uses the full model state at $t-1$, including variables absent from the LP controls. Write the model solution as $s_t = As_{t-1} + B\varepsilon_t$, where s_t is the full 40-variable model vector and ε_t is an i.i.d. vector of seven independent, mean-zero Gaussian innovations. The model parameters are fixed at the values described above. Define $\mathcal{I}_{t-1} = \sigma(s_\tau : \tau \leq t-1)$, the information generated by the state history through the previous quarter. The latest state s_{t-1} is sufficient for the forecast. Let e_π select quarterly inflation from the model state, so that $\pi_t = e'_\pi s_t$. Define the row vector $F = \sum_{j=1}^4 e'_\pi A^j$. Then Fs_t is the forecast of first-year cumulative inflation given the current state. Because the instrument uses information available before the

current short-term nominal interest rate. They can therefore contribute to the coefficients on the instrument in both the outcome and interest-rate regressions. As in Proposition 1, identification depends on their net contribution to the exclusion restriction after conditioning on the controls.

I choose the loading λ to study what happens when the instrument also contains information about future inflation. This concern motivates work that distinguishes monetary policy innovations from information about economic fundamentals (Nakamura and Steinsson, 2018; Jarociński and Karadi, 2020). Forecast controls provide one way to separate these sources of variation (Miranda-Agrippino and Ricco, 2021). The experiment examines how the choice of controls affects LP-IV estimates.

The benchmark instrument sets $\lambda = 0$. I introduce contamination by setting $\lambda = \pm\sqrt{2/9}$. When $\lambda = \pm\sqrt{2/9}$ and the independent noise has standard deviation one, the forecast component accounts for 10% of the instrument's unconditional variance. I also increase the noise standard deviation to three while holding λ fixed. The forecast component then accounts for only 2.2% of instrument variance, but the population LP-IV coefficients remain unchanged. For any given control specification, independent noise changes neither the instrument's covariance with the outcome nor its covariance with the normalization variable. Adding this noise therefore reduces instrument strength without changing identification bias.

5.2 WHY FORECAST RELIABILITY MUST BE ASSESSED AFTER CONTROLLING FOR LAGS I compare five sets of controls. The first contains only an intercept. The second adds four lags each of output, inflation, and the annualized short-term nominal interest rate. The remaining three retain these controls and add the exact predecision forecast f_t , a noisy predecision forecast $f_t + \eta_t$, or a noisy postdecision forecast $f_t^+ + \eta_t$. The measurement error η_t is independent Gaussian noise with standard deviation 0.5. The postdecision forecast f_t^+ incorporates the current monetary policy innovation and the other date- t innovations. Both forecasts predict the same first-year log-price change, $Q_{E,t}$, covering quarters $t + 1$ through $t + 4$. Both use the same scale, chosen so that the predecision forecast has unit unconditional variance.

The lagged controls already explain most of the true inflation forecast. Adding a forecast control helps remove instrument contamination if it captures the information those lags miss. Here, measurement error makes that difficult. Although the observed forecast closely tracks the true forecast overall, much of their common variation is already explained by the lags. After controlling for those lags, roughly two-thirds of the observed forecast's remaining variance is measurement error. The noisy forecast therefore removes about a third of the bias from instrument contamination, from 0.034 to 0.023 in the first year and from 0.063 to 0.041 in years two and three, but does not eliminate it.

Adding the exact inflation forecast changes the covariance between the instrument and the current nominal interest rate by only about 3%. This covariance is calculated after removing the variation explained by the controls from both variables. Yet adding the forecast changes the conclusion about how

current innovations, its forecast instead starts from s_{t-1} and equals FAs_{t-1} :

$$\mathbb{E}[Q_{E,t} | \mathcal{J}_{t-1}] = \sum_{j=1}^4 e'_\pi A^{j+1} s_{t-1} = FAs_{t-1}, \quad f_t = \frac{FAs_{t-1}}{\sqrt{FAPA'F}},$$

where $P = \text{Var}(s_t)$ solves $P = APA' + B\Sigma B'$ and $\Sigma = \text{Var}(\varepsilon_t)$. I compute the expectation by starting from s_{t-1} and setting all innovations from date t onward to their conditional mean of zero. This averages over the current policy innovation, the other current innovations, and future innovations; the forecast follows the continuing effects of past shocks.

long prices continue to fall. Here, $\lambda = \sqrt{2/9}$, so higher forecast inflation raises the instrument even when the current policy innovation is unchanged. With lagged controls alone, LP-IV implies that prices end the third year close to their level at the end of the first year. Including the exact forecast recovers the true response: prices fall about as much over the following two years as they do during the first year. Judging the controls by the stability of the first-stage covariance would therefore miss a substantial error in the persistence of the price response.

Adding the noisy forecast reduces population bias relative to lagged controls alone, but the implied price decline remains too small and too concentrated in the first year. With $f_t + \eta_t$ added to the lagged controls, the population LP-IV coefficients imply approximate price declines of 0.0399% during the first year and 0.0229% during years two and three, per 25-basis-point increase in the annualized short-term nominal interest rate. These declines are 36% and 64% smaller than the true responses, respectively.

A forecast formed after the policy decision can introduce bias because it responds to the policy shock itself. In the model, a monetary tightening lowers forecasted inflation. Holding the forecast fixed can therefore associate a tightening with other shocks that raise forecasted inflation. Those shocks also affect future prices, so the comparison need not recover the monetary policy response. With $\lambda = \sqrt{2/9}$, adding the noisy postdecision forecast to the lagged controls implies a price decline of approximately 0.0117% during the first year followed by an increase of 0.0189% during years two and three, per 25-basis-point increase in the annualized short-term nominal interest rate. The true price response continues to decline. This control also produces bias when $\lambda = 0$: even an initially valid instrument can fail to identify the policy response after conditioning on a forecast that responds to the policy shock.

5.3 SAMPLING UNCERTAINTY CAN OBSCURE OR SHARPEN THE DISTORTION I simulate samples of 120 and 240 quarters to examine how often LP-IV leads to incorrect conclusions about the price response. I check whether the confidence sets contain the true responses and whether they incorrectly imply that prices rise after a monetary tightening. The specifications that identify the true responses provide a separate check: does the confidence procedure attain its stated coverage when population bias is zero? Each of 2,000 replications covers 60 combinations: two sample lengths, five sets of controls, three values of λ , and two standard deviations of the instrument noise. Each simulation starts with a draw of the model's state from its stationary distribution and includes additional observations to construct the lagged controls and future outcomes. Within each replication, the specifications share the same structural innovations and underlying measurement-error draws. The shorter sample uses the first 120 estimation dates of the longer sample. These choices reduce simulation noise when comparing specifications. Appendix C.2 describes the model solution, initialization, and implementation checks.

I construct confidence sets as if the instrument satisfies exclusion after conditioning on the chosen controls, including in specifications where this restriction fails. I use the Anderson–Rubin procedure in Section 3.7. For each candidate response, I estimate the variance allowing for serial correlation up to twelve quarters, using Bartlett weights. I also report results using 24 quarters. Individual response sets have a nominal 95% confidence level. For inference on both responses together, I combine two 97.5% sets using the Bonferroni correction in Proposition 7. If each set achieves its stated coverage, both contain the true responses with probability at least 95%. This joint coverage concerns the two responses

Table 3: Population bias and confidence-set coverage in the monetary experiment

Controls	Population bias		Coverage of true policy responses (%)		Coverage of population LP-IV coefficients (%)	
	Year 1	Years 2–3	$n = 120$	$n = 240$	$n = 120$	$n = 240$
No controls	0.151	0.180	50.45	16.95	73.65	78.85
Four lags	0.034	0.063	82.70	85.15	87.20	92.45
Exact forecast	0.000	0.000	85.75	92.20	85.75	92.20
Noisy predecision forecast	0.023	0.041	83.60	87.80	86.85	92.90
Noisy postdecision forecast	0.051	0.083	76.90	73.85	87.25	92.45

Notes: $\lambda = \sqrt{2/9}$, independent instrument-noise standard deviation one, 2,000 paired replications, and twelve lags in the Bartlett HAC variance estimate. A population LP-IV coefficient is the value to which the estimate converges as the sample size grows, for the specified instrument and controls. Population bias is this coefficient minus the true policy response, in units of 100 times the log price level per 25-basis-point increase in the annualized short-term nominal interest rate. Coverage is the fraction of replications in which both 97.5% Anderson–Rubin confidence sets contain their respective targets, for the first-year and years-two–three price changes.

within each specification. The Monte Carlo measures actual coverage with the two fixed bandwidths.

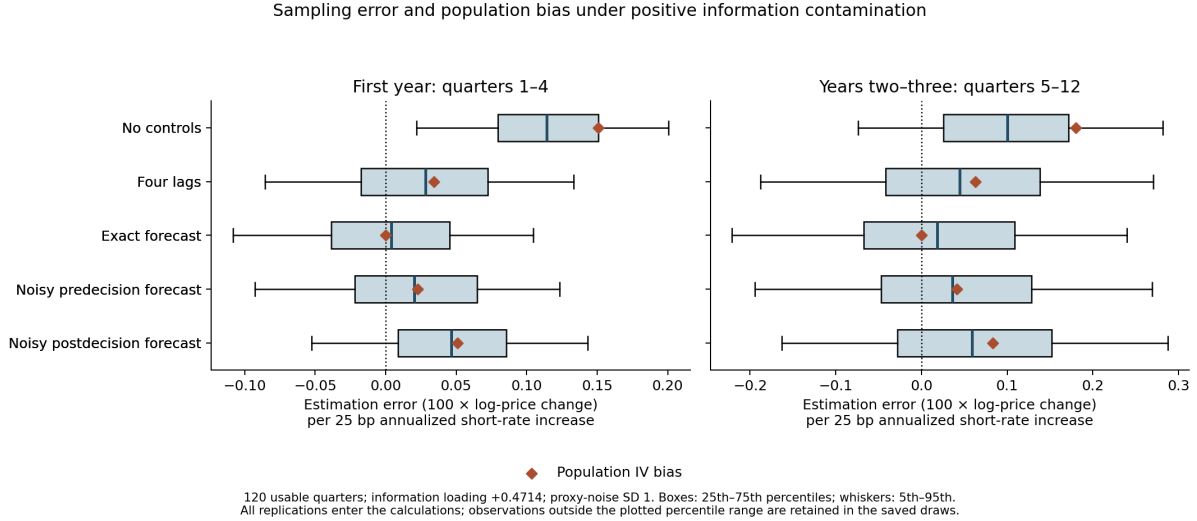
With the exact forecast included among the controls, the population IV coefficients equal the true policy responses. Yet the confidence sets fall short of their joint 95% coverage target. In Table 3, both true responses are included in their respective sets in only 85.75% of samples with 120 observations and 92.20% with 240 observations. The Monte Carlo standard errors are 0.78 and 0.60 percentage points. The same problem occurs with an uncontaminated instrument, $\lambda = 0$, and lagged controls alone: coverage is 86.05% and 92.30%. These specifications identify the policy responses, so their coverage shortfalls reflect the finite-sample performance of the confidence procedure.

Using 24 rather than 12 lags in the HAC variance estimate makes coverage worse in these samples. With the exact forecast control, joint coverage falls to 80.15% with 120 observations and 89.20% with 240 observations. Coverage improves when the sample is substantially longer. In a separate experiment with 2,000 independent replications of 960 quarterly observations, the confidence sets cover both true responses in 94.70% of samples, with a Monte Carlo standard error of 0.50 percentage points. This experiment uses an uncontaminated instrument, $\lambda = 0$, the exact forecast control, and twelve lags in the HAC variance estimate. Appendix C reports the additional results for alternative instruments, controls, and bandwidths.

With the contaminated instrument used in Table 3 and no controls, larger samples more often produce confidence sets that incorrectly imply price increases after a monetary tightening. At least one of the two 97.5% confidence sets lies entirely above zero in 18.85% of samples with 120 observations and 38.95% with 240 observations, although both true responses are negative. The latter frequency has a Monte Carlo standard error of 1.09 percentage points. With the exact forecast control, this false conclusion occurs in only 0.45% of samples with 240 observations. As the sample grows, coverage of the true responses falls from 50.45% to 16.95%, even as coverage of the biased population LP-IV coefficients improves. The confidence sets increasingly support the wrong policy responses because their construction assumes exclusion holds.

In this experiment, adding a noisy forecast formed before the policy decision improves coverage

Figure 3: Sampling dispersion can conceal different biases in early and later log-price changes



Notes: Estimation errors relative to structural first-year and additional years-two-three log-price responses, in units of 100 times the log price level per 25-basis-point increase in the annualized short-term nominal interest rate. The experiment uses 2,000 samples of 120 quarterly origins, the positive 10% reference information loading, and independent proxy-noise standard deviation one. Boxes contain the middle 50% of estimates, interior lines mark medians, and whiskers mark the fifth and 95th percentiles. Diamonds mark exact population IV biases. Every replication enters the calculations, including estimates outside the displayed percentile range.

slightly, while adding one formed afterwards makes it worse. With 240 observations, the confidence sets contain both true policy responses in 85.15% of samples with lagged controls alone, 87.80% with the noisy predecision forecast, and 73.85% with the noisy postdecision forecast. With the postdecision forecast, the sets contain both population LP-IV coefficients in 92.45% of samples. That higher coverage concerns coefficients distorted by conditioning on a forecast that responds to the policy shock. It does not imply that the confidence sets reliably describe the true policy responses.

With 120 observations, the median LP-IV estimate understates the price decline even when the controls eliminate population bias. Figure 3 reports the estimated response minus the true response across the simulated samples. With the exact forecast control, the median errors are 0.004 for the first-year response and 0.018 for the additional response during years two and three. With lagged controls alone, they are 0.028 and 0.044. These errors are measured in 100 times log prices per 25-basis-point increase in the annualized short-term nominal interest rate. Positive errors mean that the estimated price decline is too small. The LP-IV estimate divides the instrument's sample covariance with the log-price change by its sample covariance with the current nominal interest rate. Both covariances are calculated after removing the variation explained by the chosen controls from the instrument, the outcome, and the interest rate. For the specification with lagged controls, these are four lags each of output, inflation, and the nominal interest rate, together with an intercept. The exact-forecast specification also includes f_t . If the instrument and the interest rate have little remaining covariance in a particular sample, the denominator can be close to zero and the estimated response can be very large. I therefore summarize the estimates with medians and quantiles, which are less sensitive to extreme estimates; see Lewis and Mertens (2026).

A stable first stage does not establish identification, and identification does not ensure that confidence sets attain their stated coverage in short samples. In this experiment, adding the exact forecast eliminates population bias while barely changing the first-stage covariance. Yet the confidence sets still fall short of their coverage target. In an empirical application, the argument for the controls must explain why exclusion is credible for the particular response being estimated. The researcher also needs to assess whether the confidence sets achieve their stated coverage, given the sample size and the persistence in the data.

6 CONCLUSION

An instrument correlated with other shocks can still identify some effects of a policy change. I derive the restrictions under which LP-IV identifies individual responses, cumulative effects, and differences between responses at different horizons. Identification depends on how those other shocks affect both the outcome and the variable used to scale the policy shock, after accounting for the controls. Their contributions can cancel for one response but not another. An instrument can therefore identify the initial effect of a policy change without identifying how long that effect lasts.

When these restrictions fail, bounds on the contributions of other shocks can still support conclusions about policy effects. I construct joint confidence sets that allow contamination of the coefficients on the instrument in both the outcome and normalization regressions. The procedure allows weak instruments and has valid asymptotic coverage under the stated sampling conditions when the bounds, specified before estimation, contain the true contributions. In the fiscal application, I calculate how far these bounds can be widened before the confidence sets include spending multipliers of one or more.

The monetary experiment shows why a control cannot be defended by its timing. Lagged controls leave most of the bias from past shocks in place, a noisy forecast removes a third of it, and a forecast formed after the policy decision introduces bias even when the instrument is valid, while the first stage barely moves across these cases. Even when the controls eliminate population bias, confidence sets fall short of their stated coverage in short samples.

For applied work, four practices follow. Argue exclusion for each reported response: the instrument must be uncorrelated with the outcome net of the normalization, after the controls used in that regression. Report sensitivity thresholds in the units of the response, as the bias in the multiplier or the price response that would overturn the conclusion. Do not read a stable first stage across control sets as evidence about exclusion. Judge a forecast control by what it adds conditional on the lags, and date it before the policy decision.

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. doi:10.1214/aoms/1177730090.
- Andrews, Isaiah, James H. Stock, and Liyang Sun. 2019. “Weak Instruments in Instrumental Variables Regression: Theory and Practice.” *Annual Review of Economics* 11: 727–753. doi:10.1146/annurev-economics-080218-025643.
- Briganti, Edoardo, Gillian Brunet, and Victor Sellemi. 2026. “Why Does GDP Move Before Government Spending? It’s All in the Measurement.” Working paper, June. https://edoardobriganti.github.io/research/measurement_delays.pdf.
- Conley, Timothy G., Christian B. Hansen, and Peter E. Rossi. 2012. “Plausibly Exogenous.” *Review of Economics and Statistics* 94 (1): 260–272. doi:10.1162/REST_a_00139.
- Deuchert, Eva, and Martin Huber. 2017. “A Cautionary Tale About Control Variables in IV Estimation.” *Oxford Bulletin of Economics and Statistics* 79 (3): 411–425.
- Glynn, Adam N., Miguel R. Rueda, and Julian Schuessler. 2024. “Post-Instrument Bias in Linear Models.” *Sociological Methods & Research* 53 (4): 1829–1845.
- González-Páramo, José Manuel. 2005. “Taxation, Tax Reform and Monetary Policy.” Speech at Universidad Complutense, Madrid, May 13. European Central Bank. <https://www.ecb.europa.eu/press/key/date/2005/html/sp050513.en.html>.
- Inoue, Atsushi, Òscar Jordà, and Guido M. Kuersteiner. 2026. “Inference for Local Projections.” *The Econometrics Journal* 29 (1): 2–26.
- Jarociński, Marek, and Peter Karadi. 2020. “Deconstructing Monetary Policy Surprises – The Role of Information Shocks.” *American Economic Journal: Macroeconomics* 12 (2): 1–43.
- Jordà, Òscar. 2005. “Estimation and Inference of Impulse Responses by Local Projections.” *American Economic Review* 95 (1): 161–182.
- Jordà, Òscar, and Alan M. Taylor. 2025. “Local Projections.” *Journal of Economic Literature* 63 (1): 59–110.
- Khalaf, Lynda, Zhenjiang Lin, and Haowei Tang. 2026. “Monetary Policy Surprises: Robust Dynamic Causal Effects.” *Journal of Economic Dynamics and Control* 186: 105309. doi:10.1016/j.jedc.2026.105309.
- Kolesár, Michal, and Mikkel Plagborg-Møller. 2025. “Dynamic Causal Effects in a Nonlinear World: the Good, the Bad, and the Ugly.” *Journal of Business & Economic Statistics* 43 (4): 737–754. doi:10.1080/07350015.2025.2539478.

- Leeper, Eric M., Todd B. Walker, and Shu-Chun Susan Yang. 2013. “Fiscal Foresight and Information Flows.” *Econometrica* 81 (3): 1115–1145.
- Lewis, Daniel J., and Karel Mertens. 2026. “Weak Instrument Bias in Impulse Response Estimators.” Federal Reserve Bank of Dallas Working Paper 2601. [doi:10.24149/wp2601](https://doi.org/10.24149/wp2601).
- Lloyd, Simon, and Ed Manuel. 2024. “Controls, Not Shocks: Estimating Dynamic Causal Effects in Macroeconomics.” Bank of England Staff Working Paper 1079.
- Mertens, Karel, and Morten O. Ravn. 2013. “The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States.” *American Economic Review* 103 (4): 1212–1247.
- Miranda-Agrippino, Silvia, and Giovanni Ricco. 2021. “The Transmission of Monetary Policy Shocks.” *American Economic Journal: Macroeconomics* 13 (3): 74–107.
- Miranda-Agrippino, Silvia, and Giovanni Ricco. 2023. “Identification with External Instruments in Structural VARs.” *Journal of Monetary Economics* 135: 1–19.
- Montiel Olea, José Luis, Mikkel Plagborg-Møller, Eric Qian, and Christian K. Wolf. 2026. “Local Projections or Vector Autoregressions? A Primer for Macroeconomists.” In *NBER Macroeconomics Annual 2025*, volume 40. Chicago: University of Chicago Press.
- Nakamura, Emi, and Jón Steinsson. 2018. “High-Frequency Identification of Monetary Non-Neutrality: The Information Effect.” *Quarterly Journal of Economics* 133 (3): 1283–1330.
- Newey, Whitney K., and Kenneth D. West. 1987. “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix.” *Econometrica* 55 (3): 703–708.
- Plagborg-Møller, Mikkel, and Christian K. Wolf. 2021. “Local Projections and VARs Estimate the Same Impulse Responses.” *Econometrica* 89 (2): 955–980.
- Ramey, Valerie A. 2011. “Identifying Government Spending Shocks: It’s All in the Timing.” *Quarterly Journal of Economics* 126 (1): 1–50.
- 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.
- Rambachan, Ashesh, and Neil Shephard. 2019. “When Do Common Time Series Estimands Have Nonparametric Causal Meaning?” arXiv:1903.01637, revised January 2025.
- Romano, Joseph P., Azeem M. Shaikh, and Michael Wolf. 2010. “Hypothesis Testing in Econometrics.” *Annual Review of Economics* 2: 75–104. [doi:10.1146/annurev.economics.102308.124342](https://doi.org/10.1146/annurev.economics.102308.124342).
- Smets, Frank, and Rafael Wouters. 2007. “[Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach.](#)” *American Economic Review* 97 (3): 586–606.

Stock, James H., and Mark W. Watson. 2012. "Disentangling the Channels of the 2007–09 Recession." *Brookings Papers on Economic Activity*, Spring: 81–135.

Stock, James H., and Mark W. Watson. 2018. "Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments." *Economic Journal* 128 (610): 917–948.

A FORMAL DERIVATIONS

Write $Q_{r,t}$ for the specified outcome or finite weighted sum of outcomes and $X_{r,t}$ for the normalization variable. All projections include an intercept. The response index, weights, horizons, and projection spaces are fixed. The proofs therefore cover a single outcome at a given horizon, a cumulative outcome, and a prespecified difference between responses at different horizons.

A.1 THE GOVERNMENT-SPENDING MULTIPLIER EXAMPLE One stationary construction for (1) takes mutually independent sequences of independent standard normal innovations m_t, u_t, v_t, η_t and sets

$$G_t = m_t + u_{t-1}, \quad Y_t = m_{t-1} + m_{t-2} + u_{t-2}, \quad Z_t = m_t + u_{t-1} + v_t.$$

The predecision influences in the main text are $s_{1,t} = u_{t-1}$ and $s_{2,t} = m_{t-1}$, and the additional later-output disturbance is $\varepsilon_t = m_{t+1} + u_t$. Thus ε_t is independent of $m_t, s_{1,t}, s_{2,t}, v_t, \eta_t$ and has variance two. The measurement error η_t is realized before m_t . Outcomes for different estimation dates overlap, but the finite moving-average construction is stationary and ergodic with finite moments. Sample covariances therefore converge to their population values.

The expenditure identity completes the construction: $Y_t - G_t$ is consumption plus investment plus net exports in the same units. A unit increase in m_t raises current government purchases by one and leaves current GDP unchanged, so these other components fall by one in total. GDP then rises by one at each of the next two dates, while government purchases return to their original path. A unit increase in $s_{1,t} = u_{t-1}$ likewise raises current government purchases with a complete contemporaneous offset, followed by a one-unit GDP increase at $t+1$ and no change at $t+2$. The construction specifies these expenditure offsets without modeling household or firm behavior. The comparisons in the main text use future GDP relative to current government purchases, so their discrepancies are not same-date measures of expenditure excluding government purchases.

For the calculations, suppress the date subscript and write $Y_1 = Y_{t+1}$ and $Y_2 = Y_{t+2}$. Then

$$G = m + s_1, \quad Z = m + s_1 + v, \quad Y_1 = m + s_1 + s_2, \quad Y_2 = m + \varepsilon.$$

The variables m, s_1, s_2, v, η are independent with unit variance, and ε is independent of them. All variables are centered. With no additional controls, $\text{Cov}(Z, G) = 2$, $\text{Cov}(Z, Y_1) = 2$, and $\text{Cov}(Z, Y_2) = 1$. At the true multiplier of one, the exclusion moments are

$$\kappa_1 = \text{Cov}(Z, Y_1 - G) = \text{Cov}(Z, s_2) = 0, \quad \kappa_2 = \text{Cov}(Z, Y_2 - G) = \text{Cov}(Z, \varepsilon - s_1) = -1.$$

Equation (6) gives multipliers one and 1/2.

After projecting on s_1 , the residuals are

$$G^{\perp s_1} = m, \quad Z^{\perp s_1} = m + v, \quad Y_1^{\perp s_1} = m + s_2, \quad Y_2^{\perp s_1} = m + \varepsilon.$$

The first-stage covariance is one. The two residual discrepancies are s_2 and ε , both orthogonal to $Z^{\perp s_1}$.

Thus both exclusion moments are zero and both multipliers equal one. Although $\text{Cov}(s_1, Y_2 - G) = -1$, this does not invalidate the control: the augmented equation is $Y_2 = G - s_1 + \varepsilon$. Exclusion concerns the residual discrepancy after including the control.

For the noisy control $W = s_1 + \eta$, define $a = (s_1 - \eta)/2$. Projection gives

$$G^{\perp W} = m + a, \quad Z^{\perp W} = m + v + a, \quad Y_1^{\perp W} = m + a + s_2, \quad Y_2^{\perp W} = m + \varepsilon.$$

Since $\text{Var}(a) = 1/2$, the first-stage covariance is $3/2$. The residual discrepancies are s_2 and $\varepsilon - a$, giving $\kappa_1 = 0$ and $\kappa_2 = -1/2$. The IV multipliers are one and $2/3$.

For the mixed predecision control $W = s_1 + s_2$, define $c = (s_1 - s_2)/2$. The corresponding residuals are

$$\begin{aligned} G^{\perp W} &= m + c, & Z^{\perp W} &= m + v + c, \\ Y_1^{\perp W} &= m, & Y_2^{\perp W} &= m + \varepsilon, & s_2^{\perp W} &= -c. \end{aligned} \tag{52}$$

The first-stage covariance is again $3/2$. The residual discrepancies are $-c$ and $\varepsilon - c$, so both exclusion moments equal $-1/2$. Hence both IV multipliers equal $2/3$. In particular, adding the mixed indicator changes the first-horizon exclusion moment from zero to $-1/2$.

In the order of the four control specifications in Table 1, the residualized instrument variances are 3, 2, $5/2$, and $5/2$. Dividing each first-stage covariance by its residualized instrument variance gives the first-stage regression slopes in (2):

$$p = \frac{2}{3}, \quad \frac{1}{2}, \quad \frac{3}{5}, \quad \frac{3}{5}.$$

These slopes differ from the covariances reported in the table. Finally,

$$\text{Corr}(s_1, s_1 + \eta)^2 = \text{Corr}(s_1, s_1 + s_2)^2 = \frac{1}{2}.$$

The indicators have equal predictive accuracy for s_1 and equal residual first-stage covariances, but their other components have different roles in the output discrepancy. All controls remain fixed under the intervention in m_t . The design matrices have full rank, and residual relevance holds in every control specification. The failures in the table are therefore failures of exclusion at the structural multiplier.

A.2 SINGLE-EQUATION IDENTIFICATION

Proof of Proposition 1. Assumption 1 and the definition of β_r imply

$$Q_{r,t}(m) - \beta_r X_{r,t}(m) = Q_{r,t}(0) - \beta_r X_{r,t}(0).$$

Thus the target discrepancy $Q_{r,t} - \beta_r X_{r,t}$ is invariant to the target intervention. This property defines the discrepancy independently of the instrument; it does not establish statistical exclusion. Linearity of projection gives

$$\tilde{Q}_{r,t} = \beta_r \tilde{X}_{r,t} + u_{r,t}.$$

Multiply by $\tilde{Z}_{r,t}$, take expectations, and divide by the nonzero residual first stage. The result is (16). The

IV coefficient equals the structural target exactly when its numerator is zero. $\square$

A.3 COMMON DESIGNS AND THE RANK OF EXCLUSION RESTRICTIONS For a fixed W_t , common validity is the intersection of the response-specific relevance and exclusion requirements. It is therefore weakly more restrictive than validity for any one response. Under (18), define

$$\mathcal{J}_r = \{j : \ell_{r,j} \neq 0\}, \quad \mathcal{J}_{\mathcal{R}} = \bigcup_{r \in \mathcal{R}} \mathcal{J}_r.$$

Balancing the residualized instrument against every component in $\mathcal{J}_r$ suffices for response r ; balancing every component in $\mathcal{J}_{\mathcal{R}}$ suffices for the family. These are sufficient conditions because the exact moments can also vanish through cancellation. Adding a response imposes a stricter identification requirement only if its restriction is not implied by the existing restrictions and maintained model. Repeated component labels or proportional loading rows need not do so. Allowing W_r to differ across responses changes the projections at which these conditions are evaluated, while leaving each response's relevance and exclusion requirements intact.

Proof of Proposition 2. The maintained orthogonality of $\xi_{r,t}(W)$ in (18) gives

$$\mathbb{E}[Z_t^{\perp W} u_{r,t}(W)] = \ell'_r \mathbb{E}[Z_t^{\perp W} S_t^{\perp W}] = \ell'_r \Delta_W.$$

Stacking these equalities gives $L_{\mathcal{R}} \Delta_W$. Common exclusion holds exactly when this vector is zero. Componentwise balance sets the entire vector Δ_W to zero and is sufficient. If $L_{\mathcal{R}}$ has full column rank, its null space is trivial and componentwise balance is necessary. A rank-deficient matrix has a nontrivial null space; exclusion alone then permits nonzero imbalance in that space. Whether a particular imbalance vector is feasible also depends on the maintained joint distribution. Every residual first stage must separately be nonzero for common validity. $\square$

A.4 FRESH INFORMATION AND PREDECISION CONDITIONING Let $\mathcal{F}_{t-}$ be the information determined before the intervention and instrument realization. Its measurable variables remain fixed under that intervention.

Lemma 1 (Predecision controls and fresh information). *Suppose $\mathbb{E}[n_t | \mathcal{F}_{t-}] = 0$. If Z_t , n_t , and W_t are square integrable and W_t is $\mathcal{F}_{t-}$ -measurable, then*

$$n_t^{\perp W} = n_t, \quad \mathbb{E}[Z_t^{\perp W} n_t^{\perp W}] = \mathbb{E}[Z_t n_t]. \quad (53)$$

Proof. For any square-integrable $\mathcal{F}_{t-}$ -measurable H_t , $\mathbb{E}[H_t n_t] = \mathbb{E}[H_t \mathbb{E}[n_t | \mathcal{F}_{t-}]] = 0$. Apply this equality to the constant and each component of W_t to obtain $\mathcal{P}_W n_t = 0$. Apply it to $\mathcal{P}_W Z_t$ to obtain

$$\mathbb{E}[Z_t^{\perp W} n_t] = \mathbb{E}[Z_t n_t] - \mathbb{E}[(\mathcal{P}_W Z_t) n_t] = \mathbb{E}[Z_t n_t].$$

$\square$

If fresh news is the only nonorthogonal component of a target discrepancy and has nonzero loading, predecision controls cannot restore exclusion. An offsetting covariance with another component would constitute cancellation; it would not remove the covariance with fresh news.

A.5 OMITTED-INDEX BIAS

Proof of Proposition 4. Residualizing (28) on $(1, W_t)$ gives

$$Q_{r,t}^{\perp W} = \beta_r X_{r,t}^{\perp W} + \Gamma_r' R_{r,t}^{\perp W} + \zeta_{r,t}^{\perp W}.$$

Orthogonality of the residualized instrument to the projection span implies $\mathbb{E}[Z_t^{\perp W} \zeta_{r,t}^{\perp W}] = \mathbb{E}[Z_t^{\perp W} \zeta_{r,t}]$. Multiply the displayed equation by $Z_t^{\perp W}$, take expectations, and divide by $\sigma_{ZX,r}(W) \neq 0$. This proves (30) without assuming that controlling for $R_{r,t}$ yields a valid IV regression. The coefficient equals the target exactly when the two numerator terms sum to zero. The conditions in (32) set each term to zero separately. Exact spanning implies $R_{r,t}^{\perp W} = 0$ but places no restriction on the remaining instrument– $\zeta_{r,t}$ covariance. $\square$

The identity applies separately to each response and its controls. With common controls, every moment uses W_t ; with response-specific controls, each moment uses its assigned $W_{r,t}$. Under the homogeneous intervention, neither choice changes the structural targets. Balancing a stacked index and satisfying every remainder moment is sufficient for common exclusion, but can impose redundant restrictions.

A.6 A PREDECISION CONTROL THAT CREATES BIAS

Proof of Proposition 5. Let $A_t, B_t, e_{Z,t}, v_t, u_t$ be mutually independent standard normal variables. The states A_t, B_t are realized before the intervention, and the target shock is $m_t = e_{Z,t}$. Define

$$\begin{aligned} W_t &= A_t + B_t, & Z_t &= A_t + e_{Z,t}, \\ D_t &= Z_t + v_t, & Y_t &= \beta D_t + \gamma B_t + u_t, \end{aligned} \quad \gamma \neq 0. \quad (54)$$

The shock effects on D_t and Y_t are one and β , so the normalized target is β . Without controls, IV identifies the target because $\text{Cov}(Z_t, B_t) = 0$ and $\text{Cov}(Z_t, D_t) = 2$. Controlling for the exact state B_t removes γB_t , leaves Z_t and D_t unchanged by projection, and preserves orthogonality to u_t .

The alternative predecision control W_t has variance two and covariance one with both Z_t and B_t . Hence

$$Z_t^{\perp W} = \frac{1}{2}(A_t - B_t) + e_{Z,t}, \quad B_t^{\perp W} = \frac{1}{2}(B_t - A_t), \quad D_t^{\perp W} = Z_t^{\perp W} + v_t.$$

The residualized instrument–state covariance is $-1/2$, and the residual first stage is $3/2$. The remaining error stays orthogonal to the residualized instrument. Proposition 4 therefore yields $b(W) - \beta = -\gamma/3$. Although the control precedes the shock and contains the state B_t , conditioning on it induces dependence between the instrument and that state. $\square$

A.7 RESIDUALIZATION, IV, AND ANDERSON–RUBIN INFERENCE

Proof of Proposition 6. For a fixed response, denote the population projection coefficients of Z_t and $Q_{r,t} - \beta_r X_{r,t}$ on $W_{r,t}$, including the intercept, by γ_Z and γ_β , and write hats for their sample OLS counterparts. Let $\widehat{G} = T^{-1} \sum_t W_{r,t} W'_{r,t}$ be the sample control second moment. The sample normal equations give the exact identity

$$\bar{\psi}_r(\beta_r) = T^{-1} \sum_t \psi_{r,t}(\beta_r) - (\widehat{\gamma}_Z - \gamma_Z)' \widehat{G} (\widehat{\gamma}_\beta - \gamma_\beta). \quad (55)$$

The projection rates and second-moment convergence in Assumption 2 make the last term $O_p(T^{-1})$. Estimated residualization therefore leaves the first-order score limit unchanged.

Let $\widehat{\sigma}_{ZX,r} = T^{-1} \sum_t \widehat{\widehat{Z}}_{r,t} \widehat{\widehat{X}}_{r,t}$. The estimating equation gives

$$\widehat{b}_r - \beta_r = \frac{\bar{\psi}_r(\beta_r)}{\widehat{\sigma}_{ZX,r}}, \quad \widehat{\sigma}_{ZX,r} \rightarrow_p \sigma_{ZX,r}.$$

For fixed nonzero first stages, the stacked score CLT and Slutsky's theorem imply consistency and (34). Under a scalar AR null, (55) instead gives $\sqrt{T} \bar{\psi}_r(\beta_r) \xrightarrow{d} N(0, \Omega_{rr})$. Consistency and positivity of the estimated long-run variance imply the χ^2_1 limit in (35). This argument does not divide by a first stage.

For a fixed family of K responses, let $\mathcal{C}_r$ invert a level- α/K test. Each true target is excluded with limiting probability at most α/K . Thus

$$\limsup_{T \rightarrow \infty} \Pr\{\text{some } \beta_r \notin \mathcal{C}_r\} \leq \alpha.$$

Cross-response independence is unnecessary. Every member still requires its null exclusion and sampling conditions. Along weak-first-stage sequences these same conditions must hold along the sequence; a uniform conclusion requires uniform assumptions. $\square$

Sample orthogonality makes $\bar{\psi}_r(\beta_0)$ equal to the mean of the residualized instrument times the unresidualized discrepancy. It does not make their individual products equal. A long-run variance computed from the latter products can omit the first-order effect of estimating the instrument projection. Equation (55) explains why the score in (33) residualizes both variables.

For one response, a Bartlett estimator of the long-run score variance is

$$\widehat{\Omega}_r(\beta_0) = \widehat{\gamma}_{r,0}(\beta_0) + 2 \sum_{k=1}^{\ell_T} \left(1 - \frac{k}{\ell_T + 1}\right) \widehat{\gamma}_{r,k}(\beta_0),$$

where

$$\widehat{\gamma}_{r,k}(\beta_0) = T^{-1} \sum_{t=k+1}^T \{\widehat{\psi}_{r,t}(\beta_0) - \bar{\psi}_r(\beta_0)\} \{\widehat{\psi}_{r,t-k}(\beta_0) - \bar{\psi}_r(\beta_0)\}.$$

Stationarity and finite moments alone do not imply projection rates, a score CLT, or covariance consistency. Under exclusion, a stationary Gaussian process with finite memory, nonsingular retained control moments, and positive score long-run variances provides a sufficient setting. Here finite memory means that observations separated by more than a fixed number of periods are independent. A conservative

Bartlett bandwidth in this setting satisfies $\ell_T \rightarrow \infty$ and $\ell_T = o(T^{1/4})$. At the true response ratios, the residual products and their derivatives with respect to the projection coefficients are polynomials in a finite collection of Gaussian variables. They therefore have moments of every order and finite dependence, while the sample projection coefficients converge at rate $T^{-1/2}$. These properties and the stated bandwidth satisfy the sufficient conditions for covariance consistency in Newey and West (1987). Centering the sample residual products does not change the probability limit. The Smets–Wouters process instead has an infinite moving-average representation; Appendix C.2 describes the simulation.

A.8 PROJECTION OVER A FIXED NUISANCE CLASS

Proof of Proposition 7. Under the homogeneous intervention, $Q_{r,t} = \theta_{Q,r} m_t + Q_{r,t}(0)$ and $X_{r,t} = \theta_{X,r} m_t + X_{r,t}(0)$. Subtracting $\beta_r X_{r,t}$ from $Q_{r,t}$ cancels the policy shock's contribution, so linearity of projection gives $a_r(\beta_r) = \delta_{Q,r} - \beta_r \delta_{X,r}$. At that target and the true nuisance pair, the numerator of the test in (39) is $|\hat{a}_r(\beta_r) - a_r(\beta_r)|$. The regression CLT in (37) and its variance consistency condition imply

$$\frac{\hat{a}_r(\beta_r) - \delta_{Q,r} + \beta_r \delta_{X,r}}{\hat{\text{se}}_r(\beta_r)} \xrightarrow{d} N(0, 1).$$

Each absolute-value inequality with critical value $c_{\alpha,K}$ fails with limiting probability α/K . The union bound controls the probability that any true-target inequality fails by α . Because the fixed joint class $\mathcal{D}$ contains the true nuisance vector, projection over that class preserves the event on which all these inequalities hold. The target vector consequently belongs to $\mathcal{C}(\mathcal{D})$ with asymptotic probability at least $1 - \alpha$.

The proof does not require the first-stage covariance to be bounded away from zero. It does require the stated regression limits and consistent estimates of positive variances, including along any weak-first-stage sequence under consideration. A nuisance class estimated or selected using the data requires a separate coverage argument that accounts for that estimation or selection. $\square$

B PROXY-SVAR IDENTIFICATION

B.1 EXTERNAL PROXY-SVAR IDENTIFICATION Maintain the structural VAR in (25). Let C_h be its reduced-form propagation matrices, with $C_0 = I_n$, so that $\Theta_h = C_h B$ contains structural responses. Define $\gamma = \mathbb{E}[\varepsilon_t Z_t] = (\gamma_1, \gamma'_{-1})'$, where γ_{-1} collects covariances with shocks $2, \dots, n$. The external proxy-SVAR propagates the impact covariance $\mathbb{E}[\mathbf{u}_t Z_t] = B\gamma$. For the weighted outcome and exposure defined in the main text, it therefore reports

$$b_r^{\text{ext}} = \frac{\sum_{j=1}^n \theta_{Q,r}^j \gamma_j}{\sum_{j=1}^n \theta_{X,r}^j \gamma_j}, \quad (56)$$

when its denominator is nonzero. Subtracting the well-defined target β_r gives

$$b_r^{\text{ext}} - \beta_r = \frac{\sum_{j=2}^n (\theta_{Q,r}^j - \beta_r \theta_{X,r}^j) \gamma_j}{\sum_{j=1}^n \theta_{X,r}^j \gamma_j}. \quad (57)$$

The target shock's term vanishes because $\theta_{Q,r}^1 = \beta_r \theta_{X,r}^1$. This proves the equivalence between response-specific identification and (27). If $\gamma_1 \neq 0$ and $\gamma_j = 0$ for $j \geq 2$, the numerator is zero and the denominator is $\theta_{X,r}^1 \gamma_1 \neq 0$ for every maintained normalization. Stacking the numerators gives $L_{\mathcal{R}} \gamma_{-1} = 0$. Componentwise exclusion is necessary for a selected family when these net-loading rows have full column rank, as in Proposition 2.

The full innovation representation $\mathbf{u}_t = B\epsilon_t$ suffices for this comparison. Under their maintained dynamic proxy restrictions, Plagborg-Møller and Wolf (2021, Section 3.3) also establish external-proxy identification with invertibility of the target shock alone. Here, the full representation justifies identification from the contemporaneous moments in (26) and establishes necessity for the complete impact vector.

Proof of Proposition 3. Write the common current normalization as $D_t = d'Y_t$ and let e_1 select the target shock. The maintained nonzero impact and covariance conditions are $d'Be_1 \neq 0$ and $d'B\gamma \neq 0$. Equality of all normalized impact responses is the vector equality

$$\frac{B\gamma}{d'B\gamma} = \frac{Be_1}{d'Be_1}. \quad (58)$$

Multiplying by B^{-1} gives $\gamma = \lambda e_1$, with $\lambda = (d'B\gamma)/(d'Be_1) \neq 0$. Conversely, this covariance restriction implies (58). Multiplying the common normalized impact vector by C_h then gives the correct external proxy-SVAR response at every horizon. Nonsingularity of B establishes necessity: without it, distinct structural-shock covariance vectors can produce the same impact covariance direction. $\square$

B.2 EXTERNAL PROXY-SVARs AND LP-IV WITH MACROECONOMIC HISTORY LP-IV and the external proxy-SVAR can differ if the instrument predicts future innovations. To isolate this difference, maintain (25) and let $\mathcal{H}_{t-1}^Y$ be the closed linear span of the intercept and all lags of Y_t . Define $Z_t^{\perp \mathcal{H}_{t-1}^Y} = Z_t - \mathcal{P}_{\mathcal{H}_{t-1}^Y} Z_t$ and the dated covariance vectors $\kappa_s = E[\epsilon_{t+s} Z_t^{\perp \mathcal{H}_{t-1}^Y}]$ for $s \geq 0$. Causality and white structural shocks imply that current innovations are orthogonal to macroeconomic history, so $\kappa_0 = \gamma$. These vectors differ from γ_j , which denotes the covariance with contemporaneous shock j .

For $k \geq 1$, $\epsilon_{t-k} = B^{-1} \mathbf{u}_{t-k}$ belongs to $\mathcal{H}_{t-1}^Y$, so residualization removes covariance with every past structural innovation. The moving-average representation therefore gives

$$E[Y_{t+h} Z_t^{\perp \mathcal{H}_{t-1}^Y}] = C_h B \gamma + g_h, \quad (59)$$

$$g_h = \sum_{s=1}^h C_{h-s} B \kappa_s, \quad g_0 = 0. \quad (60)$$

The first term propagates the impact covariance used by the external proxy-SVAR. The second collects future innovations associated with the residualized instrument. For a current normalization, both methods divide by $d'B\gamma$, assumed nonzero. Their population responses for variable i at horizon h therefore differ by $e_i' g_h / (d'B\gamma)$, where e_i selects that variable. Orthogonality to future structural shocks is sufficient for equality at every horizon; equality for one response only requires its projection of g_h to vanish.

For weighted outcomes and normalization variables, covariances with future shocks can enter both

the numerator and denominator. Let $Y_{i,t} = e'_i \mathbf{Y}_t$ and $D_t = d' \mathbf{Y}_t$, with $i = i(r)$ for response r . Whenever its denominator is nonzero, the LP-IV coefficient is

$$b_r(\mathcal{H}_{t-1}^Y) = \frac{\sum_{j=1}^n \theta_{Q,r}^j \gamma_j + \sum_{h=0}^{J_r} d_{r,h} e'_{i(r)} g_h}{\sum_{j=1}^n \theta_{X,r}^j \gamma_j + \sum_{s=0}^{M_r} c_{r,s} d' g_s}. \quad (61)$$

Thus contemporaneous proxy exclusion and relevance identify the ratio implied by external proxy-SVAR propagation under the maintained VAR and a nonzero structural normalization. LP-IV also requires restrictions on the contributions of future shocks to $Q_{r,t} - \beta_r X_{r,t}$.

For the common current normalization, suppose $d' B e_1 \neq 0$ and $d' B \gamma \neq 0$. The entire LP-IV response vector is correct at every horizon $0, \dots, H$ if and only if $\gamma = \lambda e_1$ with $\lambda \neq 0$ and $\kappa_s = 0$ for $s = 1, \dots, H$. Equality at impact gives the first condition by (58). At horizon one, equality then requires $B \kappa_1 = 0$. Nonsingularity of B gives $\kappa_1 = 0$; applying the same argument successively establishes the remaining restrictions. Substitution proves sufficiency. Selected responses can instead satisfy the weaker cancellations in Proposition 2.

These comparisons use macroeconomic history. A VAR that includes Z_t has a different innovations representation and must be compared with an LP conditioning on the same expanded history.

B.3 INTERNAL-INSTRUMENT VARs AND MATCHED INFORMATION The population LP-VAR equivalence requires matching the conditioning information (Plagborg-Møller and Wolf, 2021). Under covariance stationarity, pure nondeterminism, and the regularity conditions supporting an unrestricted linear innovations representation, a recursive VAR containing the instrument, treatment, and outcomes, with the instrument ordered first, reproduces LP-IV when the LP conditions on the entire past of that same expanded vector, including instrument lags. This result uses a closed span of infinitely many lags and is distinct from the finite-control sampling argument.

If the LP also conditions on current pre-announcement variables C_t , the recursive ordering must put C_t before Z_t and retain the same past. The innovation assigned to Z_t then matches the LP's residualized instrument. More general controls require a construction reproducing their projection space; placing an instrument in an arbitrary VAR is insufficient.

Let $\theta_{i,j}^{\text{rec}}$ and $\theta_{D,s}^{\text{rec}}$ denote the matched recursive responses of outcome i and the policy variable, with the same shock scaling. For finite weighted sums defining the outcome and normalization variable,

$$b_r(W) = \frac{\sum_{j=0}^{J_r} d_{r,j} \theta_{i(r),j}^{\text{rec}}}{\sum_{s=0}^{M_r} c_{r,s} \theta_{D,s}^{\text{rec}}}, \quad (62)$$

provided the denominator is nonzero. This equality follows from linearity of the response sums. A cumulative normalization uses the corresponding weighted treatment response in the denominator. These equalities do not establish exclusion. Finite-order VARs also impose extrapolation restrictions, so finite-lag comparisons need not yield exact agreement (Plagborg-Møller and Wolf, 2021; Montiel Olea et al., 2026).

C EMPIRICAL AND SIMULATION DETAILS

C.1 FISCAL ESTIMATES AND THE SENSITIVITY FRONTIER The fiscal data come from the November 2016 workbook in the Ramey–Zubairy replication archive (Ramey and Zubairy, 2018). Its potential-GDP series is based on a sixth-degree polynomial fitted to log real GDP, excluding 1930–1946. I treat the supplied potential-GDP values as fixed. The regression includes four lags each of normalized output, purchases, and news, together with an intercept. The coefficients do not vary across economic states. I include World War II observations and omit additional tax controls and a time trend.

Military news is missing throughout 1889 in both archived data formats. With four news lags, the first usable estimation date is therefore 1891Q1; the archived replication results report 1890Q2. The eight-quarter estimate using all available estimation dates through 2014Q1 is 0.6637143. Restricting the sample to dates with all observations needed for the sixteen-quarter outcome gives 0.6653506. The sixteen-quarter estimate is 0.7133606. The archived replication reports coefficients of 0.6619107 and 0.7119396. The replication materials record an April 2016 update, while the workbook is dated November 2016. I have not reconciled the remaining differences in the estimates. I do not impute missing news observations.

The original replication selects the HAC bandwidth automatically and evaluates weak-IV tests over a grid. I use Bartlett bandwidths of eight, sixteen, and 24 quarters and solve the Anderson–Rubin inequalities analytically. Both LP equations use the same 485 estimation dates, 1891Q1–2012Q1, with future observations through 2015Q4. Joint inference covers the two prespecified horizons within this application.

Every regression uses the same controls. To estimate the variance at a candidate multiplier β_0 , I regress $Q_H - \beta_0 X_H$ on the controls and instrument without restricting the instrument coefficient. I use the resulting residuals in the HAC calculation. The degrees-of-freedom adjustment multiplies the covariance estimate by $T/(T - \text{rank}(W) - 1)$, counting the intercept in W . Lags in the Bartlett estimator follow calendar quarters. Solving the test inequalities retains empty, unbounded, and disconnected confidence sets. The fixed bandwidths define the procedures used in the application; the asymptotic coverage results require the stated consistency conditions for the variance estimator.

To show why the confidence sets exclude every multiplier above one when the sensitivity bounds lie below the frontier, write

$$v_H(\beta_0) = \widehat{\text{se}}_H(\beta_0)^2 = v_{YY,H} - 2\beta_0 v_{YG,H} + \beta_0^2 v_{GG,H}.$$

This variance is nonnegative because its covariance matrix is positive semidefinite. Let $c = \Phi^{-1}(0.9875)$, as in the Bonferroni procedure for the two horizons. In the data, $\widehat{a}_{G,H} > 0$ and $\widehat{a}_{Y,H} < \widehat{a}_{G,H}$. For $\beta_0 \geq 1$, the test rejects every combination of contamination coefficients within the bounds whenever

$$\text{gap}_H(\beta_0) = \beta_0(\widehat{a}_{G,H} - d_{G,H}) - \widehat{a}_{Y,H} - d_{Y,H} - c\sqrt{v_H(\beta_0)} > 0.$$

Table 4: Fiscal exclusion frontiers in original projection units

H	HAC lags	$\hat{a}_H(1)$	SE	d_H^*	Relevance margin	Scaled d_H^* (%)
8	8	-0.569	0.236	0.040	0.734	2.35
8	16	-0.569	0.230	0.053	0.801	3.12
8	24	-0.569	0.229	0.055	0.839	3.25
16	8	-1.239	0.447	0.237	1.474	5.47
16	16	-1.239	0.466	0.194	1.485	4.50
16	24	-1.239	0.470	0.186	1.464	4.31

Notes: H is the horizon in quarters. $\hat{a}_H(1)$ is the estimated coefficient on normalized military news in a regression of cumulative GDP minus cumulative government purchases on news and the common controls. SE is its HAC standard error, calculated from unrestricted regression residuals. $d_H^* = \max\{0, |\hat{a}_H(1)| - c\text{SE}\}$, with $c = \Phi^{-1}(0.9875)$. The relevance margin is $\hat{a}_{G,H} - c\text{SE}(\hat{a}_{G,H})$, where $\hat{a}_{G,H}$ is the estimated coefficient on news in the cumulative-purchases regression. Scaled d_H^* expresses d_H^* as a percentage of $|\hat{a}_{G,H}|$. In each row, d_H^* is positive and smaller than the relevance margin. Joint coverage concerns only the two horizons at a bandwidth specified before estimation. The contamination bounds are assumptions rather than estimates of the actual contamination.

Since $\sqrt{v_H(\beta_0)}$ is a Euclidean norm, for $\beta_{0,2} > \beta_{0,1}$,

$$\sqrt{v_H(\beta_{0,2})} - \sqrt{v_H(\beta_{0,1})} \leq (\beta_{0,2} - \beta_{0,1})\sqrt{v_{GG,H}}.$$

It follows that

$$\text{gap}_H(\beta_{0,2}) - \text{gap}_H(\beta_{0,1}) \geq (\beta_{0,2} - \beta_{0,1})\{\hat{a}_{G,H} - d_{G,H} - c\hat{\text{se}}_{G,H}\}.$$

The quantities $\hat{a}_{G,H} - c\hat{\text{se}}_{G,H}$ are 0.8005 and 1.4853 at the two horizons. They exceed the sensitivity thresholds $d_8^* = 0.0531$ and $d_{16}^* = 0.1945$. Any pair of bounds satisfying $d_{Y,H} + d_{G,H} < d_H^*$ therefore gives $\text{gap}_H(1) > 0$ and $\text{gap}_H(\beta_0) > 0$ for all $\beta_0 \geq 1$. On or above the frontier, a combination with $\delta_{Y,H} \leq 0$ and $\delta_{G,H} \geq 0$ prevents rejection of a multiplier of one. The frontier thus marks where the confidence set ceases to lie entirely below one, including when the set becomes unbounded.

Together, the bounds for the two horizons define a fixed set of four contamination coefficients. Applying the two Bonferroni tests to every combination in this set gives the confidence set in Proposition 7. Dividing the displayed bounds by the estimated purchases coefficients expresses the frontier as percentages of those estimates. If the estimates were instead used to choose the population bounds, the coverage argument would need to account for their uncertainty.

C.2 SMETS–WOUTERS SOLUTION AND MEASUREMENT SYSTEM I solve the Smets–Wouters model at the reported posterior mode for 1966Q1–2004Q4, using the archived values of seven innovation standard deviations and 29 structural parameters (Smets and Wouters, 2007). I use a first-order approximation in Dynare 7.1 with MATLAB R2026b and do not re-estimate the parameters. The risk-premium and investment-efficiency equations follow the replication specification, including its documented differences from the printed equations.

The solution expresses the model variables after the date- t innovations as

$$s_t = A s_{t-1} + B \varepsilon_t, \quad \text{Var}(\varepsilon_t) = \Sigma. \quad (63)$$

The largest eigenvalue of A in modulus is 0.9762. The target shock m_t is the current innovation to the monetary disturbance, standardized to have unit variance. The disturbance itself is persistent. A one-standard-deviation innovation raises the annualized short-term nominal interest rate by 0.7215 percentage points on impact. Quarterly inflation and the quarterly short rate are measured in percentage points; multiplying the short rate by four expresses it in annualized units. Summing quarterly inflation gives changes in 100 times the log price level without a further annualization factor.

The matrix B maps the innovations into model variables, and Σ contains the innovation variances. The stationary covariance P solves $P = APA' + B\Sigma B'$. The full 40-variable vector contains redundant coordinates, so P can be singular. I draw the initial state and compute control projections using the directions with positive variance. Before computing the population projections, I standardize the control covariance matrix and discard eigenvalues below 10^{-11} times its largest eigenvalue.

Let e'_π select quarterly inflation and define $F = \sum_{j=1}^4 e'_\pi A^j$ and $\sigma_f^2 = \text{Var}(FAs_{t-1})$. The two forecast measurements are

$$f_t = \frac{FAs_{t-1}}{\sigma_f}, \quad f_t^+ = \frac{Fs_t}{\sigma_f}. \quad (64)$$

Both are forecasts of the sum of inflation from $t+1$ through $t+4$, divided by σ_f . This sum equals the change in 100 times log prices over the first year. The first forecast conditions on the state before date- t innovations; the second incorporates those innovations. They therefore predict the same outcome using different information. The predecision forecast has unit variance. With independent measurement error of standard deviation 0.5, the true forecast accounts for 80% of the variance of $f_t + \eta_t$. After removing the variation explained by the lagged controls from f_t and $f_t + \eta_t$, the true forecast accounts for 34.8% of the measured forecast's residual variance.

Impulse responses computed by iterating the state equation agree with those from the model solution within 1.2×10^{-15} . An independent calculation agrees within 5.6×10^{-15} . The stationary covariance equation holds with relative error 1.3×10^{-15} , and the computed population bias with the exact forecast control is below 1.1×10^{-14} .

Increasing the independent instrument-noise standard deviation from one to three, while holding the forecast loading fixed, reduces the forecast's share of instrument variance from 10% to approximately 2.17%. The true policy responses and population IV coefficients remain unchanged. I also calculate population coefficients for positive and negative forecast loadings that give variance shares of 5%, 10%, and 20% when the noise standard deviation is one. Without controls, the covariance between the instrument and the current annualized short-term nominal interest rate is zero at $\lambda = -0.4226$. The population IV ratio is undefined at this loading, which is outside the values used in the calculations.

C.2.1 SAMPLING AND NUMERICAL VERIFICATION The simulations include all seven Gaussian innovations. The main experiment uses 2,000 replications with 120 or 240 quarterly estimation dates, plus four preceding and twelve subsequent quarters to construct the controls and outcomes. Each replication starts with a draw from the stationary distribution. Innovations are independent across replications. Within each replication, the specifications share the same innovations and measurement-error draws, and the shorter sample uses the first 120 estimation dates of the longer sample.

Table 5: Coverage under valid exclusion in the Smets–Wouters experiment

Controls	$n = 120$			$n = 240$		
	Year 1	Years 2–3	Joint	Year 1	Years 2–3	Joint
No controls	91.55	91.10	90.40	92.65	93.40	93.20
Four lags	90.00	87.10	86.05	92.75	91.70	92.30
Exact forecast	89.15	87.50	85.75	92.85	91.40	92.20
Noisy predecision forecast	90.00	87.75	86.00	93.25	91.50	92.50

Notes: An uncontaminated instrument, independent instrument-noise standard deviation one, 2,000 replications, and twelve lags in the Bartlett HAC variance estimate. The year-one and years-two–three columns report coverage of the corresponding true policy responses by nominal 95% confidence sets. Joint coverage is the frequency with which two 97.5% sets contain both true responses. All displayed specifications satisfy population exclusion. Entries are percentages.

The stable linear model has an infinite moving-average representation. Its stationary Gaussian dynamics imply a central limit theorem for the regression moments when the retained regressor covariance matrices are nonsingular and the relevant long-run variances are positive. Consistent HAC estimation also requires the bandwidth to grow at a suitable rate with sample size. The fixed 12- and 24-quarter bandwidths are the choices evaluated in the simulations. Reporting results separately for specifications that satisfy exclusion isolates failures of the sampling approximation from identification bias.

In an independent check using 40,000 simulated histories, the simulated covariances match the analytical covariances within 1.75 Monte Carlo standard errors. The checks cover the timing of innovations and forecasts, the common estimation dates, the short-rate normalization, and the twelve lagged controls. All replications enter the reported summaries, including those with unbounded or disconnected confidence sets.

Table 5 reports coverage when exclusion holds. A separate experiment uses 2,000 independent replications with 960 quarterly estimation dates, an uncontaminated instrument with noise standard deviation one, and either lagged controls alone or lagged controls plus the exact forecast. With twelve HAC lags and the exact forecast, the individual 95% confidence sets cover the first-year and years-two–three responses in 94.70% and 93.65% of replications. Combining two 97.5% sets gives joint coverage of 94.70%, with Monte Carlo standard error 0.50 percentage points. With 24 HAC lags, joint coverage is 93.85% with lagged controls alone and 94.10% with the exact forecast. None of the 97.5% sets is unbounded in these 2,000 replications, although unbounded sets remain possible.

Table 6 reports the effects of reversing the forecast loading, increasing instrument noise, and increasing the HAC bandwidth. Independent noise leaves population bias unchanged but makes the estimates less precise. Confidence sets can then include the true policy responses more often, even though exclusion still fails. Increasing the bandwidth does not eliminate the coverage shortfall in specifications that satisfy exclusion.

Table 6: Joint coverage of the true policy responses under alternative specifications

Controls	$\lambda = -\sqrt{2/9}$		Noise SD 3		HAC lags 24	
	$n = 120$	$n = 240$	$n = 120$	$n = 240$	$n = 120$	$n = 240$
No controls	63.75	33.00	82.40	69.55	42.85	15.90
Four lags	86.15	90.10	89.05	91.30	76.75	81.55
Exact forecast	85.75	92.20	88.80	92.20	80.15	89.20
Noisy predecision forecast	86.80	91.75	89.35	91.90	78.85	85.10
Noisy postdecision forecast	84.50	87.90	88.20	88.55	71.00	69.65

Notes: Coverage percentages across 2,000 replications. Each pair of columns changes one feature of the benchmark with $\lambda = \sqrt{2/9}$, instrument-noise standard deviation one, and twelve HAC lags. Reversing the sign of λ preserves the forecast's 10% share of instrument variance. Increasing noise holds λ fixed. Joint coverage is the frequency with which two 97.5% confidence sets contain both true policy responses within a specification; it does not cover all specifications jointly.