

Local Linear Local Projections: Estimating Shock-Size Heterogeneity*

Tom Pessó

September 2026

*Preliminary and incomplete.*¹
[\[Latest version here\]](#)

Abstract

Do large shocks propagate differently from small ones, and do positive and negative shocks have the same effects? A nonlinear response to the initial shock can reflect either a direct nonlinear effect of the shock or propagation through economic conditions changed by the shock. I show that these mechanisms are generally observationally equivalent from the response curve alone: distinguishing them requires additional information on the propagation mechanism. Rather than seeking to identify which mechanism generates the nonlinearity, I take a local perspective, asking how the response to a marginal change in the shock varies across shock values and when this response has a causal interpretation. I propose Local Linear Local Projections (LL-LP) to trace this variation without imposing bins, thresholds, or a global nonlinear functional form. I show how conventional linear projections and specifications that split shocks by sign or size provide different summaries of the underlying slope curve. The framework also provides a basis for testing sign and size nonlinearities separately while allowing the shock distribution to be asymmetric.

Keywords: Local Linear Local Projections; shock-size heterogeneity; nonlinear shocks; state dependence; local derivatives.

JEL Codes: C14; C22; C32; E32.

*Earlier versions were presented or circulated as *Fiscal Multiplier: the Size of the Shock Matters*.

¹This preliminary version develops the population framework and first-stage asymptotics. Inference for the local-slope estimator and the scalar sign and magnitude summaries under time-series dependence remains to be developed.

Introduction

Do large shocks propagate differently from small ones? Do positive and negative shocks have the same effects? These questions arise whenever the response to a shock may vary across the range of shock values observed in the data. A conventional linear local projection (Jordà, 2005) summarizes this response with a single coefficient. When the response is nonlinear, that coefficient can average slopes from different parts of the shock distribution, obscuring where and how transmission changes.

A growing empirical literature allows the effects of structural shocks to vary with their sign or magnitude. Such nonlinearities have been documented in the transmission of fiscal shocks (Bai et al., 2026; Barnichon, Debortoli, & Matthes, 2022), monetary-policy shocks (Ascari & Haber, 2022), and financial shocks (Barnichon, Matthes, & Ziegenbein, 2022). This evidence suggests that a single linear coefficient can conceal economically meaningful heterogeneity. Existing approaches, however, typically characterize this heterogeneity through transformations, thresholds, or categories specified in advance. This raises a more general question: how can the response be characterized without specifying in advance where or how it varies with shock size?

Shock-size dependence also raises an interpretation problem: it need not originate in a nonlinear effect of the initial shock itself. A shock may change unemployment, fiscal space, financial stress, inflation pressure, or monetary-policy accommodation, and these conditions may in turn shape its subsequent transmission. Even if the impact response is linear, larger and smaller shocks can therefore induce different state paths and generate nonlinear responses at later horizons. More generally, nonlinearity in the response curve can reflect either a direct nonlinear effect of the shock or nonlinear propagation through economic conditions changed by the shock.

These two mechanisms raise an identification problem. I show that they are generally observationally equivalent from the response curve alone: the same shock-size dependence can be represented as originating directly in the shock or through its subsequent propagation. Distinguishing between these interpretations therefore requires additional information on the propagation mechanism. Rather than imposing such structure, I take a local perspective, asking how the response to a marginal change in the shock varies across shock values and when this response has a causal interpretation.

I propose Local Linear Local Projections (LL-LP) to estimate this local response. At each shock value, LL-LP uses local-linear smoothing to estimate the slope of the response curve in a neighborhood of that value. Tracing these local slopes across the shock distribution produces a slope curve that reveals where and how transmission changes, without imposing bins, thresholds, or a global nonlinear functional form. Under appropriate local assignment and support conditions, the slope at a given shock value has a causal interpretation: nearby differences in the shock can be interpreted as local interventions after conditioning on pre-shock information.

The slope curve also clarifies how more familiar projection coefficients summarize shock-size heterogeneity. Conventional linear local projections aggregate slopes across shock values, while specifications that split shocks by sign or size summarize different regions of the same underlying curve. LL-LP retains the starting shock value rather than aggregating it away, making it possible to see which parts of the shock distribution drive these scalar summaries.

Recent work provides related interpretations of projection coefficients as averages of heterogeneous effects. Rambachan and Shephard (2025) and Kolesár and Plagborg-Møller (2025) show that conventional time-series projection coefficients can be interpreted as weighted averages of marginal effects

under appropriate conditions, while Gonçalves et al. (2026) study average effects of increasing a structural shock by a fixed amount. These objects average slopes or finite changes across starting shock values. LL-LP instead retains the starting value and reports the local slope there.

The framework also complements work on state-dependent local projections. David et al. (2026) give a causal interpretation to state-dependent coefficients when the conditional mean remains linear in the shock, while Gonçalves et al. (2024) distinguish coefficients around realized state–shock configurations from responses to finite interventions. Here the response may instead be nonlinear in the shock itself, including because the shock changes states that shape its subsequent propagation. A local approach allows the response to be nonlinear in the shock while retaining a causal interpretation under the conditions described above. Without local comparability of nearby shocks, the slope remains descriptive.

Estimation combines local-linear smoothing in the shock with Robinson partialling out of finite-dimensional controls (Robinson, 1988). This keeps the nonparametric dimension one while allowing the controls used in macroeconomic specifications to enter parametrically. Because residualization changes the conditioning implicit in the pointwise response, linking the empirical one-dimensional slope to the state-conditional causal object requires the composition condition developed in Section 2.5. The appendix establishes the population residualization identities and the asymptotic properties of the finite-dimensional first step.²

The pointwise perspective is particularly useful for separating sign from magnitude nonlinearities. At a given absolute magnitude, comparing the slopes at equally sized positive and negative shocks isolates sign dependence, while comparing these quantities across magnitudes reveals size dependence. These pointwise comparisons can also be aggregated into scalar summaries, connecting the framework to the sign and magnitude components studied by Caravello and Martínez Bruera (2024). With an asymmetric shock distribution, however, an unweighted scalar comparison can mix sign dependence with differences in the magnitudes observed on either side of zero. I therefore define aggregation relative to a symmetric reference distribution. A natural choice symmetrizes the observed density, preserving the distribution of absolute shock magnitudes while assigning equal weight to the two signs. This makes explicit the distribution relative to which sign and size nonlinearities are summarized and provides a basis for testing them separately.

Simulations illustrate how LL-LP can reveal nonlinear response patterns obscured by more restrictive specifications, how differences in the composition of pre-shock states can generate apparent shock-size heterogeneity, and how scalar summaries vary with the reference distribution.

Section 1 develops the dynamic-state mechanism and the observational-equivalence result. Section 2 defines the local response objects and studies their causal interpretation and relation to conventional projection summaries. Section 3 develops the LL-LP estimator and the accompanying theoretical results. Section 4 presents the simulations, and Section 5 concludes.

²Inference for the second-stage local-slope estimator under time-series dependence will be developed in a subsequent version.

1 Endogenous Propagation and the Nonlinear Response Function

State-dependent macroeconomic models condition on variables observed before the shock. These variables are predetermined when they enter each period's law of motion, although the initial shock changes their later path. If transmission varies along that path, the horizon response can be nonlinear in the initial shock even when the impact equation is linear conditional on the initial state. The linear benchmark obtains when the shock-induced state movement leaves subsequent response loadings unchanged.

1.1 Linear benchmark

Let X_t be an aggregate shock, let Z_{t-1} be a predetermined state, and let Y_{t+h} be the horizon- h outcome. Write $R_h(x, z)$ for the corresponding response normalized relative to a zero shock, so that $R_h(0, z) = 0$. A coefficient-based state-dependent LP is built around a conditional mean of the form

$$\mathbb{E}[Y_{t+h} \mid X_t = x, Z_{t-1} = z] = a_h(z) + b_h(z)x. \quad (1)$$

The coefficient $b_h(z)$ is the slope of the observed conditional mean. When the conditional mean in (1) identifies the potential-response curve, every supported finite perturbation satisfies

$$R_h(x + \delta, z) - R_h(x, z) = b_h(z)\delta. \quad (2)$$

Thus one coefficient represents all starting shock values and perturbation sizes. David et al. (2026) allow a richer predetermined component and condition on the full pre-shock information set, but retain linearity in the aggregate shock. Under their sufficient condition, heterogeneity is indexed by pre-shock information, not by the starting shock value conditional on that information.

This coefficient representation remains valid in first-order linearized systems with stable propagation, even if the shock moves future states. It generally fails when the response is directly nonlinear in the shock or when the induced state path changes later transmission. In those cases, the finite effect depends on the starting value x and perturbation δ , so a coefficient indexed only by z no longer represents every causal response. LL-LP targets a different object: the derivative with respect to x . Under the linear benchmark that derivative is constant and equals $b_h(z)$; outside the benchmark it need not be constant across starting shock values.

1.2 A dynamic unemployment-threshold example

Fiscal applications often allow the response to depend on unemployment measured before the shock. Following the unemployment-threshold specification in Ramey and Zubairy (2018), consider an impact relation

$$Y_t = B(U_{t-1})X_t, \quad (3)$$

and suppose unemployment evolves according to

$$U_t = \rho U_{t-1} - \lambda X_t, \quad \lambda > 0. \quad (4)$$

Suppose propagation of the impact response into the next period is governed by the unemployment state inherited from period t ,

$$Y_{t+1} = A(U_t)Y_t + \varepsilon_{t+1}. \quad (5)$$

The state is predetermined period by period: U_{t-1} governs the impact relation, and U_t is predetermined for the transition from t to $t + 1$. The LP is not conditioned on U_t ; that state appears only when the model is iterated from the initial value U_{t-1} . Define the structural response relative to a zero shock while holding ε_{t+1} fixed across the two interventions. Conditional on the initial state $U_{t-1} = u$, substitution gives

$$R_1(x, u) = A(\rho u - \lambda x)B(u)x. \quad (6)$$

If $\mathbb{E}[\varepsilon_{t+1} \mid X_t, U_{t-1}] = 0$, (6) also equals the observed conditional-mean response normalized relative to zero. The impact response $R_0(x, u) = B(u)x$ is linear in x , but the one-period response is generally nonlinear because the shock changes the unemployment state through which the impact effect propagates.

If propagation changes discretely across unemployment regimes,

$$A(v) = A^H \mathbf{1}\{v \geq \bar{U}\} + A^L \mathbf{1}\{v < \bar{U}\}, \quad (7)$$

then

$$R_1(x, u) = \left[A^H \mathbf{1}\left\{x \leq \frac{\rho u - \bar{U}}{\lambda}\right\} + A^L \mathbf{1}\left\{x > \frac{\rho u - \bar{U}}{\lambda}\right\} \right] B(u)x. \quad (8)$$

A threshold in the future unemployment state appears, from the perspective of the original pre-shock information set, as a threshold in shock size. The threshold need not be zero; it depends on the initial unemployment state and on how strongly the fiscal shock moves unemployment.

At the induced threshold, the response in (8) is generally discontinuous when $A^H \neq A^L$; even under a continuity restriction it can have a kink. The derivative statements below therefore apply to interior points of the two regimes. At the threshold itself, report finite contrasts and, separately, the limiting slopes of the two branches. Without additional joining conditions, those limiting slopes need not be one-sided derivatives of the discontinuous response at the threshold.

If $(\rho u - \bar{U})/\lambda = 0$, the marginal response differs by sign but is constant within each sign. If the threshold lies away from zero, small shocks of either sign may remain in one regime while larger shocks in one direction cross into the other. Away from the threshold, the derivative can therefore contain both sign and magnitude components in the terminology of Caravello and Martínez Bruera (2024).

1.3 Dynamic propagation and shock-size dependence

Continue to let $R_h(x, z)$ denote the normalized horizon- h response to a shock x , conditional on the pre-shock state z . Let

$$G_h(z, x) := (G_0(z, x), G_1(z, x), \dots, G_{h-1}(z, x)) \quad (9)$$

collect the relevant state values generated between the initial shock and horizon h by iterating the law of motion from z . A dynamic state-mediated representation writes

$$R_h(x, z) = B_h(G_h(z, x), z)x. \quad (10)$$

The shock enters linearly given the state path generated by the dynamic system, but the response conditional on the initial state can be nonlinear in x because the state path itself depends on x .

Proposition 1.1 (Dynamic propagation and shock-size dependence). *Suppose R_h has representation (10), and suppose B_h and G_h are differentiable in the relevant arguments. For $x \neq 0$, define*

$$Q_h(x, z) = \frac{R_h(x, z)}{x} = B_h(G_h(z, x), z). \quad (11)$$

Then

$$\partial_x Q_h(x, z) = \nabla_G B_h(G_h(z, x), z)^\top \partial_x G_h(z, x). \quad (12)$$

Consequently, shock-size irrelevance at the point (x, z) , defined as $\partial_x Q_h(x, z) = 0$, holds when the shock does not move the relevant state path or when transmission is locally invariant in the direction in which the shock moves that path. Constancy of $Q_h(\cdot, z)$ over an interval not containing zero requires (12) to vanish throughout that interval; an interval crossing zero additionally requires a compatible continuous extension of Q_h at zero.

Linearity in the initial shock is therefore an additional restriction, not a consequence of using predetermined states. It is retained over a supported interval when propagation is stable over the induced path or when the orthogonality condition in Proposition 1.1 holds at every point in that interval.

1.4 What the response curve cannot distinguish

At a fixed initial state z and horizon h , the curve $R_h(x, z)$ need not reveal the source of its curvature. Direct nonlinearity in the shock and linear exposure followed by state-dependent propagation can generate the same response curve. The following result concerns this curve, rather than equivalence of complete dynamic data-generating processes.

Proposition 1.2 (Auxiliary representation). *Fix z and let $R_h(\cdot, z)$ satisfy $R_h(0, z) = 0$. Let $G_h(z, \cdot)$ be a state-path map and let $\mathcal{S}_z$ denote its range over nonzero shocks. There exists a function $B_h : \mathcal{S}_z \times \{z\} \rightarrow \mathbb{R}$ satisfying*

$$R_h(x, z) = B_h(G_h(z, x), z)x \quad (13)$$

for every $x \neq 0$ if and only if

$$G_h(z, x_1) = G_h(z, x_2) \implies \frac{R_h(x_1, z)}{x_1} = \frac{R_h(x_2, z)}{x_2} \quad (14)$$

for all nonzero x_1, x_2 . In particular, if $G_h(z, \cdot)$ is one-to-one, such a representation exists for any response curve $R_h(\cdot, z)$. If R_h is differentiable at zero, the representation can be extended to $x = 0$ by setting $B_h(G_h(z, 0), z) = \partial_x R_h(0, z)$, provided this definition is compatible with every nonzero shock attaining the same state-path value.

For example, if $A(v) = a_0 + a_1 v$ in (6), then

$$R_1(x, u) = \{a_0 + a_1(\rho u - \lambda x)\}B(u)x = \kappa_1(u)x + \kappa_2(u)x^2, \quad (15)$$

where $\kappa_1(u) = (a_0 + a_1\rho u)B(u)$ and $\kappa_2(u) = -a_1\lambda B(u)$. This quadratic curve is consistent either with a primitive nonlinear shock effect or with a linear impact equation whose propagation varies with $U_t = \rho u - \lambda x$. Since $x \mapsto U_t(x)$ is one-to-one when $\lambda \neq 0$, the state-mediated representation satisfies Proposition 1.2.

The response curve identifies the combined effect of direct shock nonlinearity and nonlinear propagation along the induced state path. It does not distinguish the two when they imply the same conditional response. Doing so requires restrictions or direct information on the state transition and propagation mechanism.

1.5 Pointwise derivatives under shock-dependent propagation

When (10) is differentiable,

$$\partial_x R_h(x, z) = B_h(G_h(z, x), z) + x \nabla_G B_h(G_h(z, x), z)^\top \partial_x G_h(z, x). \quad (16)$$

The first term is the response loading evaluated along the state path generated by the shock. The second appears because a change in the initial shock also changes that path. Under the linear benchmark, the derivative is constant in x conditional on the initial state. When transmission changes along the induced state path, the derivative varies with shock size and contains both terms in (16). When the state-conditional mean identifies the potential-response curve, its derivative is this total local derivative; normalization relative to zero does not change the slope. The relationship between this state-conditional derivative and the one-dimensional empirical LL-LP target is established in Section 2.5. A decomposition requires additional structural information.

When state propagation generates shock-size dependence. A nonlinear dynamic system need not produce a nonlinear response to X_t . The response can remain linear if the shock does not move the relevant state, if propagation is stable over the supported shock range, or if transmission is locally flat along the induced path. When shock-size dependence is present, it is still a statement about the total reduced-form response because G_h and B_h are not separately identified from the response curve alone.

2 Local Derivatives of the Response Function

A nonlinear response curve can be summarized in several ways. A global LP reports a projection slope under the historical shock distribution. A transformed-shock regression compresses sign or magnitude variation into a scalar projection. A finite-shift estimand asks for an average effect of moving each realized structural shock by the same amount. LL-LP asks a local question: how steep is the response curve around a given shock value? These estimands are functionals of related curves, but they average over different dimensions. This distinction is connected to the broader nonlinear impulse-response literature, in which responses can depend on the shock, the history, and the counterfactual under consideration (Gouriéroux & Lee, 2025; Kilian & Vigfusson, 2011; Koop et al., 1996; Potter, 2000).

2.1 Scalar summaries and local derivatives

Let

$$m_h(x, z) = \mathbb{E}[Y_{t+h} \mid X_t = x, Z_{t-1} = z] \quad (17)$$

be a selected version of the observed conditional-mean curve. Whenever zero belongs to the relevant support, connect this object with Section 1 by defining the normalized descriptive response

$$R_h^{\text{obs}}(x, z) := m_h(x, z) - m_h(0, z). \quad (18)$$

More generally, subtracting any baseline that is constant in x leaves the derivative unchanged. Hence $\partial_x R_h^{\text{obs}}(x, z) = \partial_x m_h(x, z)$ wherever the derivative exists. Under the causal conditions below, the same statement applies to $R_h^*(x, z) = m_h^*(x, z) - m_h^*(0, z)$ whenever that normalization is defined. Thus the response derivatives studied in Section 1 and the conditional-mean derivatives studied here coincide under the corresponding normalization.

Proposition 2.1 (Simple conditional LP coefficient as an average derivative). *Fix z . Suppose $X_t \mid Z_{t-1} = z$ has density f_z on a compact interval $[\underline{x}_z, \bar{x}_z]$, conditional mean μ_z , and variance $0 < \sigma_z^2 < \infty$. Suppose $\mathbb{E}[Y_{t+h}^2 \mid Z_{t-1} = z] < \infty$ and $m_h(\cdot, z)$ is absolutely continuous. Define*

$$\omega_z(u) = \frac{1}{\sigma_z^2} \int_u^{\bar{x}_z} (x - \mu_z) f_z(x) dx. \quad (19)$$

Assume additionally that $\int_{\underline{x}_z}^{\bar{x}_z} |\theta_h(u, z)| \omega_z(u) du < \infty$. Then $\omega_z(u) \geq 0$ on the support and

$$\int_{\underline{x}_z}^{\bar{x}_z} \omega_z(u) du = 1. \quad (20)$$

The population slope from the simple linear projection of Y_{t+h} on 1 and X_t , conditional on $Z_{t-1} = z$, satisfies

$$\alpha_h^{\text{LP}}(z) = \int_{\underline{x}_z}^{\bar{x}_z} \theta_h(u, z) \omega_z(u) du, \quad \theta_h(u, z) := \partial_u m_h(u, z). \quad (21)$$

Proposition 2.1 concerns the simple conditional projection just defined. Controlled LPs have analogous but generally different weighting representations because they depend on the residualized shock and the maintained conditioning structure. The proposition nevertheless makes the central point: a projection slope can average local slopes and need not equal the marginal response at any specified shock value.

The reduced-form local derivative response is

$$\theta_h(x, z) = \partial_x m_h(x, z). \quad (22)$$

It is the marginal change in the horizon- h response around (x, z) , not the finite effect of moving the shock from zero to x . If $m_h(\cdot, z)$ is absolutely continuous on the interval between two supported points x_0 and x_1 , then

$$m_h(x_1, z) - m_h(x_0, z) = \int_{x_0}^{x_1} \theta_h(u, z) du. \quad (23)$$

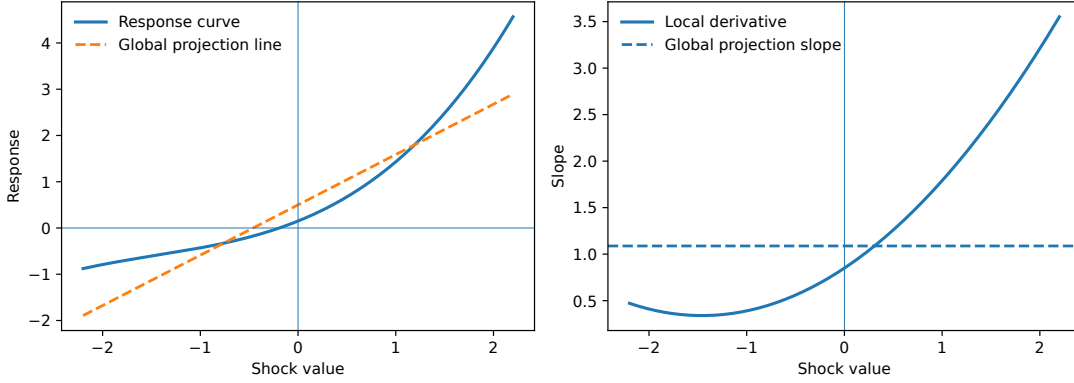


Figure 1: A nonlinear conditional-mean curve, its simple global projection, and its pointwise derivative. The projection line reports one distribution-dependent slope; the derivative records the slope at each shock value.

One coefficient represents every such contrast only when $\theta_h(u, z)$ is constant in u .

2.2 Sign and magnitude components

For a magnitude $r > 0$ such that both r and $-r$ lie in the supported shock region, define

$$\theta_h^{\text{mag}}(r, z) = \frac{\theta_h(r, z) + \theta_h(-r, z)}{2}, \quad \theta_h^{\text{sign}}(r, z) = \frac{\theta_h(r, z) - \theta_h(-r, z)}{2}. \quad (24)$$

For $x \neq 0$,

$$\theta_h(x, z) = \theta_h^{\text{mag}}(|x|, z) + \text{sgn}(x)\theta_h^{\text{sign}}(|x|, z). \quad (25)$$

At zero the two component functions may be defined by continuous limits when those limits exist.

The sign component compares marginal responses to positive and negative shocks of the same magnitude. Variation of the sign-symmetric component $\theta_h^{\text{mag}}(r, z)$ with r measures magnitude dependence after averaging over signs. A linear response has a constant magnitude component and a zero sign component. Pure magnitude dependence has $\theta_h^{\text{sign}} = 0$ and a nonconstant θ_h^{mag} , whereas pure sign asymmetry has a nonzero sign component and no variation in the magnitude component.

The decomposition itself does not assume a symmetric shock distribution. Pointwise comparisons require support near both r and $-r$. Symmetry becomes important when the components are summarized by transformed-shock regressions. Under symmetry, even transformations isolate weighted sign components and odd transformations isolate weighted magnitude components (Caravello & Martínez Bruera, 2024). Under asymmetry these components generally mix. LL-LP instead forms the pointwise population components directly from matched slopes.

Reading LPs with indicators of the shock. An indicator constructed from the shock is not a predetermined state. For example,

$$Y_{t+h} = a_h + \beta_h^- X_t \mathbf{1}\{X_t < 0\} + \beta_h^+ X_t \mathbf{1}\{X_t \geq 0\} + u_{t+h} \quad (26)$$

is a piecewise-linear approximation to the dose-response curve in X_t . It is not a state-dependent LP in the sense of a coefficient indexed by information fixed before intervention. At a fixed conditioning state, if the observed and causal curves coincide throughout the fitted sign regions and satisfy the displayed common-intercept piecewise-linear specification, the two coefficients equal the causal derivatives within those regions. Otherwise they are projection summaries of a curved response. With sign-specific intercepts, Proposition 2.1 applies to each truncated shock distribution. With a common intercept, the two regional fits are additionally tied by a cross-region restriction.

A small-large specification similarly combines a central region with two disconnected tails. A single large-shock coefficient can mix sign asymmetry, magnitude dependence, and unequal frequencies of positive and negative tail observations. LL-LP instead compares derivatives at matched values and across selected magnitudes.

Scalar projections under a symmetric target. Fix a state z , suppress it below, and let $I = [-\bar{r}, \bar{r}]$ be a symmetric common-support interval. Let ϖ be a symmetric probability density on I , specified before estimation, and let $f \in L^2(\varpi)$. Define

$$z_f(x) = (1, x, f(x))^\top, \quad e_3 = (0, 0, 1)^\top, \quad (27)$$

and the target projection

$$a_{h,f}^\varpi = \arg \min_{a \in \mathbb{R}^3} \int_I \{m_h(x) - z_f(x)^\top a\}^2 \varpi(x) dx, \quad \tau_{h,f}^\varpi = e_3^\top a_{h,f}^\varpi. \quad (28)$$

The target density determines how magnitudes are weighted. The response curve itself remains the observed conditional mean.

Proposition 2.2 (Symmetry-balanced component projections). *Let $e_f^\varpi(x)$ be the residual from projecting $f(x)$ on 1 and x under ϖ , and suppose*

$$0 < d_f^\varpi := \int_I \{e_f^\varpi(x)\}^2 \varpi(x) dx < \infty. \quad (29)$$

Define

$$H_f^\varpi(x) = \int_{-\bar{r}}^x e_f^\varpi(u) \varpi(u) du, \quad \omega_f^\varpi(x) = -\frac{H_f^\varpi(x)}{d_f^\varpi}. \quad (30)$$

If m_h is absolutely continuous on I and the displayed integrals are finite, then

$$\tau_{h,f}^\varpi = \int_I \omega_f^\varpi(x) \theta_h(x) dx. \quad (31)$$

If f is even, ω_f^ϖ is odd and

$$\tau_{h,f}^\varpi = 2 \int_0^{\bar{r}} \omega_f^\varpi(r) \theta_h^{\text{sign}}(r) dr. \quad (32)$$

If f is odd, ω_f^ϖ is even, integrates to zero, and

$$\tau_{h,f}^\varpi = 2 \int_0^{\bar{r}} \omega_f^\varpi(r) \theta_h^{\text{mag}}(r) dr. \quad (33)$$

Thus an even transformation yields a directional scalar summary of sign asymmetry, while an odd transformation yields a directional scalar summary of magnitude dependence, without requiring the observed shock distribution to be symmetric.

Choosing the symmetric target. The target density is part of the estimand. Under a global null in which the relevant component is zero at every magnitude, every admissible symmetric target gives $\tau_{h,f}^\varpi = 0$. Away from that null, different targets can give different coefficients because they weight magnitudes differently; a zero projection need not imply a zero component curve.

For this aggregation discussion, write ν_t for the same scalar shock previously denoted X_t , and let f_ν denote its marginal density. A natural reference on I is the symmetrized observed density

$$\varpi^{\text{sym}}(x) = \frac{f_\nu(x) + f_\nu(-x)}{2p_I} \mathbf{1}\{|x| \leq \bar{r}\}, \quad p_I = \Pr(|\nu_t| \leq \bar{r}). \quad (34)$$

It preserves the observed conditional distribution of $|\nu_t|$ on I and assigns equal target probability to the two signs at each magnitude. Appendix A records the elementary identity.

The symmetrized marginal density is a permissible reference for a state-conditional curve, but equality with an observed transformed-shock projection also requires matching the response curve and conditioning law. For the unconditional comparison with Caravello and Martínez Bruera (2024), apply Proposition 2.2 to $\mu_h(x) = \mathbb{E}[Y_{t+h} \mid \nu_t = x]$ under its stated conditions. If the observed shock density is symmetric and the target is the normalized restriction

$$\varpi(x) = \frac{f_\nu(x) \mathbf{1}\{x \in I\}}{p_I}, \quad (35)$$

then (28), with μ_h in place of m_h , coincides with the transformed-shock population projection restricted to I , using the same regressors and normalization. If I is the full support, $p_I = 1$ and the target is simply f_ν .

For a state-conditional comparison, retain $m_h(x, z)$ and instead use the normalized conditional density of ν_t given $Z_{t-1} = z$ on I , when this conditional density is symmetric and assigns positive probability to I . Marginal symmetry alone does not establish equality between conditional and unconditional projection coefficients.

The projection is directional rather than omnibus: a nonzero component can average to zero under a particular weight. A small pre-specified collection of transformations reduces cancellation risk without recovering the full location information. Section 3.3 discusses the inferential scope of these summaries.

2.3 Average finite shifts and pointwise derivatives

A finite-shift estimand asks how the outcome changes when every realized structural shock is displaced by the same amount. Averaging over baseline shocks and predetermined controls gives

$$A_h(\delta) = \mathbb{E}[m_h(X_t + \delta, Z_{t-1}) - m_h(X_t, Z_{t-1})], \quad (36)$$

whenever shifted values remain in the relevant support. A transparent sufficient condition for the next identity is that, for almost every (X_t, Z_{t-1}) , the map $u \mapsto m_h(X_t + u, Z_{t-1})$ is absolutely continuous

on an interval Δ containing zero and

$$\mathbb{E} \left[\int_{\Delta} |\theta_h(X_t + u, Z_{t-1})| \, du \right] < \infty. \quad (37)$$

Then Fubini's theorem and the fundamental theorem of calculus imply

$$A_h(\delta) = \int_0^\delta \mathbb{E}[\theta_h(X_t + u, Z_{t-1})] \, du. \quad (38)$$

Proposition 2.3 (Finite shifts as integrated local responses). *Let Δ be an interval containing zero and suppose (38) holds for every $\delta \in \Delta$. Define $\bar{\theta}_h(u) = \mathbb{E}[\theta_h(X_t + u, Z_{t-1})]$. Then*

$$A_h(\delta) = \int_0^\delta \bar{\theta}_h(u) \, du. \quad (39)$$

If $\bar{\theta}_h$ is continuous, then $A'_h(\delta) = \bar{\theta}_h(\delta)$.

The derivative of the finite-shift curve is a translated average marginal response. The state-conditional LL-LP formulation instead retains the starting shock and state. Under the triangular assignment conditions below, both are causal functionals of the same nonlinear response function, but the one-dimensional empirical LL-LP target additionally requires the composition bridge in Section 2.5.

2.4 State-conditional causal interpretation

Let $Y_{t+h}(x)$ be the potential outcome when the date- t aggregate shock is set to x , other contemporaneous and future structural shocks are held fixed or averaged over a specified continuation distribution, and endogenous variables evolve according to the model. The continuation rule is part of the intervention. Let

$$m_h^*(x, z) = \mathbb{E}[Y_{t+h}(x) \mid Z_{t-1} = z] \quad (40)$$

be a selected potential-response curve. Throughout the causal results, displayed conditional means are finite and use jointly measurable versions. State-conditional identification statements apply to P_Z -almost every relevant state; identification at an arbitrary exceptional state is not asserted without additional regularity in z . When averaging under a specified state distribution, the causal assumptions must hold almost everywhere under that distribution, as imposed below.

Assumption 2.1 (Consistency and compatible conditional-mean versions). $Y_{t+h} = Y_{t+h}(X_t)$ almost surely. In addition, the observed and potential conditional-mean versions are chosen compatibly: on the neighborhood N_x used below, they satisfy

$$m_h(u, z) = \mathbb{E}[Y_{t+h}(u) \mid X_t = u, Z_{t-1} = z]$$

for $P_{X|Z=z}$ -almost every $u \in N_x$ at the relevant states. The diagonal expectation is not selected independently for each potential outcome at a null conditioning event.

Assumption 2.2 (Local mean ignorability). For the point (x, z) of interest, there is a neighborhood

N_x of x such that, for almost every $u \in N_x$ under the conditional law of $X_t \mid Z_{t-1} = z$,

$$\mathbb{E}[Y_{t+h}(u) \mid X_t = u, Z_{t-1} = z] = \mathbb{E}[Y_{t+h}(u) \mid Z_{t-1} = z]. \quad (41)$$

The observed curve $m_h(u, z)$ and both sides of (41) denote the compatible versions continuous in u on N_x .

Assumption 2.3 (Local overlap and smoothness). The conditional density of X_t given $Z_{t-1} = z$ is positive on N_x , and $m_h^*(u, z)$ is differentiable in u at x .

Theorem 2.1 (Causal interpretation of the local derivative). *Under Assumptions 2.1–2.3, for the selected continuous versions,*

$$\theta_h(x, z) = \partial_x m_h(x, z) = \partial_x m_h^*(x, z) \equiv \theta_h^*(x, z). \quad (42)$$

The conclusion is local. A causal interpretation of a full derivative curve requires the assumptions over the full reported support, and a finite contrast requires them over the integration interval in (23). A ratio of output and spending derivatives requires additional restrictions before it can be read as a policy multiplier. Without these conditions, the estimand remains reduced form.

A primitive triangular sufficient condition. Consider

$$X_t = \phi(Z_{t-1}) + \varepsilon_{1t}, \quad (43)$$

$$Y_{t+h} = \Psi_h(X_t, Z_{t-1}, \eta_{t,h}), \quad (44)$$

where $\eta_{t,h}$ collects the other contemporaneous and future structural disturbances relevant for horizon h . The function Ψ_h is unrestricted in its first argument and may incorporate direct shock nonlinearity and shock-dependent propagation.

Assumption 2.4 (Additive triangular structural shock). For each fixed h , equations (43)–(44) hold, ε_{1t} is conditionally independent of $\eta_{t,h}$ given Z_{t-1} , and the conditional support of X_t contains a neighborhood of the evaluation point for almost every relevant state. The potential outcome under $X_t = x$ is $Y_{t+h}(x) = \Psi_h(x, Z_{t-1}, \eta_{t,h})$. Assume Ψ_h is jointly measurable and $\mathbb{E}|Y_{t+h}| < \infty$. Let $K_{h,z}$ be a regular conditional distribution of $\eta_{t,h}$ given $Z_{t-1} = z$, and assume $\int |\Psi_h(u, z, e)| K_{h,z}(de) < \infty$ for P_Z -almost every relevant z and every supported u under consideration. Use this same kernel to define the potential and diagonal conditional means.

Proposition 2.4 (Triangular sufficient condition). *Under Assumption 2.4,*

$$\mathbb{E}[Y_{t+h}(x) \mid X_t = x, Z_{t-1} = z] = \mathbb{E}[Y_{t+h}(x) \mid Z_{t-1} = z] \quad (45)$$

for P_Z -almost every z and $P_{X|Z=z}$ -almost every supported x . If, in addition, the compatible continuous versions in Assumption 2.2 are used on a neighborhood and Assumption 2.3 holds there, then equality of the observed and potential mean curves extends over that neighborhood and

$$\partial_x \mathbb{E}[Y_{t+h} \mid X_t = x, Z_{t-1} = z] = \partial_x \mathbb{E}[Y_{t+h}(x) \mid Z_{t-1} = z]. \quad (46)$$

2.5 From state-conditional responses to the empirical LL-LP target

Theorem 2.1 concerns $\theta_h^*(s, z)$, the derivative conditional on the full relevant pre-shock state. The aggregate estimator developed below smooths only in the shock and adjusts for a finite-dimensional control vector. Let F_s denote a selected regular conditional distribution of Z_{t-1} given $X_t = s$. The raw one-dimensional response curve is

$$\mu_h(s) = \mathbb{E}[Y_{t+h} \mid X_t = s] = \int m_h(s, z) dF_s(z). \quad (47)$$

With predetermined controls W_{t-1} , define $q(s) = \mathbb{E}[W_{t-1} \mid X_t = s]$ and $V_t^X = W_{t-1} - q(X_t)$. Assume $\mathbb{E}\|W_{t-1}\|^2 + \mathbb{E}|Y_{t+h}|^2 < \infty$ and that $\Sigma_W^X := \mathbb{E}[V_t^X (V_t^X)^\top]$ is positive definite. Define the population partial-linear coefficient

$$\gamma_h = (\Sigma_W^X)^{-1} \mathbb{E}[V_t^X \{Y_{t+h} - \mu_h(X_t)\}]. \quad (48)$$

The empirical LL-LP target is the derivative of

$$m_h^r(s) = \mathbb{E}[Y_{t+h} - W_{t-1}^\top \gamma_h \mid X_t = s] = \mu_h(s) - q(s)^\top \gamma_h, \quad (49)$$

namely

$$\beta_h(s) = \{m_h^r(s)\}' = \mu_h'(s) - \dot{q}(s)^\top \gamma_h. \quad (50)$$

Assumption 2.5 (Differentiable state composition). At an interior evaluation point s , the map $u \mapsto F_u$ admits a two-sided signed-measure derivative $\dot{F}_s$ in total variation. The functions $m_h(s, z)$ and $\partial_s m_h(s, z)$ are integrable under the relevant measures, the product rule and differentiation under the integral in (47) are valid, and q is differentiable at s . Assumptions 2.1–2.3, including their compatibility and continuous-version requirements, hold on a common neighborhood of s for F_s -almost every relevant state.

Proposition 2.5 (Causal bridge for the empirical LL-LP derivative). *Under Assumption 2.5,*

$$\beta_h(s) = \underbrace{\int \theta_h^*(s, z) dF_s(z)}_{\bar{\theta}_h^{\text{obs}}(s)} + \underbrace{\left\{ \int m_h^*(s, z) d\dot{F}_s(z) - \dot{q}(s)^\top \gamma_h \right\}}_{C_h(s)}. \quad (51)$$

The first term is an observed-composition average of state-conditional causal derivatives. Consequently,

$$\beta_h(s) = \bar{\theta}_h^{\text{obs}}(s) \iff C_h(s) = 0. \quad (52)$$

If $C_h(s) \neq 0$, the LL-LP target remains the derivative of a well-defined residualized conditional mean but combines causal response heterogeneity with residual changes in the composition of pre-shock states.

Even when $C_h(s) = 0$, the weighting distribution F_s may vary with s . Thus $\beta_h(s_1)$ and $\beta_h(s_2)$ can be causal averages for different populations. A comparison across shock values isolates shock-size heterogeneity for a common population only under an additional balance or standardization condition.

Corollary 2.1 (Local balance and a fixed population). *Suppose Proposition 2.5 applies. If $\dot{F}_s = 0$ and $\dot{q}(s) = 0$, then*

$$\beta_h(s) = \int \theta_h^*(s, z) dF_s(z). \quad (53)$$

If $W_{t-1} = w(Z_{t-1})$, local constancy of F_s implies $\dot{q}(s) = 0$ under the stated integrability conditions. If Proposition 2.5 applies at every point of a reported shock interval, $F_u = F^0$ throughout that interval, and $q(u)$ is constant there, then

$$\beta_h(s) = \int \theta_h^*(s, z) dF^0(z) \quad (54)$$

for every s in that interval.

Corollary 2.2 (Exact partial-linear state adjustment). *Suppose $W_{t-1} = w(Z_{t-1})$, $\mathbb{E}\|W_{t-1}\|^2 < \infty$, and assume there exist a vector γ_h^0 , a differentiable function g_h , and a measurable set $\mathcal{Z}_0$ such that*

$$m_h(u, z) = g_h(u) + w(z)^\top \gamma_h^0 \quad (55)$$

for every supported u and every $z \in \mathcal{Z}_0$, with $F_u(\mathcal{Z}_0) = 1$ for every supported u . Assume the displayed conditional-mean versions are continuous in u on a common neighborhood of s , the state-conditional causal assumptions hold there, and the Robinson residualized-control moment matrix is nonsingular. Then the population Robinson coefficient equals γ_h^0 and

$$\beta_h(s) = g_h'(s) = \theta_h^*(s, z) \quad (56)$$

for F_s -almost every z . In this special case the one-dimensional derivative equals the state-conditional causal derivative, not merely an average of it.

For a pre-specified distribution F^0 supported on states for which the causal curve is identified, define, whenever differentiation may pass through the integral,

$$M_h^{F^0}(s) = \int m_h^*(s, z) dF^0(z), \quad \Theta_h^{F^0}(s) = \frac{d}{ds} M_h^{F^0}(s) = \int \theta_h^*(s, z) dF^0(z). \quad (57)$$

This target asks how the same reference population would respond locally at different shock values. The baseline LL-LP target does not generally equal (57). Sufficient conditions for equality throughout an interval include the interval-wide balance conditions of Corollary 2.1 with that F^0 , or the exact-adjustment conditions of Corollary 2.2 imposed throughout the interval on an F^0 -full set of relevant states. Otherwise, explicit standardization or valid reweighting can define a separate fixed-population procedure; it is not automatic from finite-control residualization.

Why mean-zero shock exogeneity is not enough. Let $Z_{t-1} \in \{0, 1\}$, let $X_t = \sigma(Z_{t-1})\varepsilon_t$ with $\mathbb{E}[\varepsilon_t \mid Z_{t-1}] = 0$, and let $Y_{t+h}(x) = a(Z_{t-1})x + \eta_{t,h}$, with $\mathbb{E}[\eta_{t,h} \mid X_t, Z_{t-1}] = 0$. Assume $p(s) = \Pr(Z_{t-1} = 1 \mid X_t = s)$ is differentiable at the evaluation points. Then the state-conditional causal derivative is $a(z)$ and does not vary with shock size. If the shock variance differs across states, however, $p(s)$ varies with s , and

$$\mu_h'(s) = a(0) + \{a(1) - a(0)\}\{p(s) + sp'(s)\}. \quad (58)$$

The one-dimensional slope can therefore vary strongly with $|s|$ even though there is no shock-size heterogeneity within either state. A level control for Z_{t-1} does not generally remove this interaction-driven composition effect. Section 4.2 illustrates the mechanism.

Table 1: Causal hierarchy of response derivatives

Object	Interpretation	Conditions for causal use
$\theta_h^*(s, z)$	Local effect at shock value s for initial state z	Local consistency, mean ignorability, overlap, and smoothness conditional on z
$\beta_h(s)$	Derivative of the one-dimensional residualized response targeted by LL-LP	Equals the observed-composition causal average when $C_h(s) = 0$; equals a state-specific derivative under exact partial linearity
$\Theta_h^{F^0}(s)$	Fixed-population average derivative under a chosen state distribution F^0	State-conditional causal identification plus standardization to the same F^0 at every s

3 Local Linear Local Projections: Estimand and Estimation

The shock is now denoted by ν_t to emphasize that it is the identified shock series used in empirical work. The estimand is reduced form: a causal label additionally requires Theorem 2.1 and an appropriate bridge condition in Proposition 2.5 or its corollaries. The local smoothing remains one-dimensional: it targets the slope of the residualized response curve near a chosen shock value. The smoothing dimension is the shock value, not the horizon; this differs from smooth local projections, which regularize the response sequence across horizons (Barnichon & Brownlees, 2019).

3.1 Residualized response curves

Let W_{t-1} collect predetermined controls. For outcome ℓ and horizon h , define $Y_{\ell,t,h}$ as the horizon- h outcome. Let

$$\mu_{\ell,h}(s) = \mathbb{E}[Y_{\ell,t,h} \mid \nu_t = s], \quad q(s) = \mathbb{E}[W_{t-1} \mid \nu_t = s]. \quad (59)$$

Assume $\mathbb{E} \|W_{t-1}\|^2 + \mathbb{E} |Y_{\ell,t,h}|^2 < \infty$. Define $V_t = W_{t-1} - q(\nu_t)$ and assume $\Sigma_W = \mathbb{E}[V_t V_t^\top]$ is positive definite. The population control coefficient after residualizing with respect to the shock is

$$\gamma_{\ell,h} = \Sigma_W^{-1} \mathbb{E}[V_t \{Y_{\ell,t,h} - \mu_{\ell,h}(\nu_t)\}]. \quad (60)$$

The residualized response curve is

$$m_{\ell,h}(s) = \mathbb{E}[Y_{\ell,t,h} - W_{t-1}^\top \gamma_{\ell,h} \mid \nu_t = s] = \mu_{\ell,h}(s) - q(s)^\top \gamma_{\ell,h}, \quad (61)$$

and the derivative estimand is

$$\beta_{\ell,h}(s) = m'_{\ell,h}(s) = \mu'_{\ell,h}(s) - \dot{q}(s)^\top \gamma_{\ell,h}. \quad (62)$$

Controls change the estimand as well as its precision. Unless they exactly remove the composition term in (51), the controlled derivative need not equal a fixed-population causal average or a derivative at a particular state. The baseline interpretation is the derivative of a residualized conditional mean.

3.2 The LL-LP estimator

Estimate $\gamma_{\ell,h}$ and form residualized outcomes

$$\tilde{Y}_{\ell,t,h} = Y_{\ell,t,h} - W_{t-1}^\top \hat{\gamma}_{\ell,h}. \quad (63)$$

For the first-stage estimator, $\mu_{\ell,h}$ and q are estimated using fixed folds and a growing temporal buffer between each evaluation block and its training sample. Under the nuisance-rate and buffered sample-splitting conditions in Appendix D, Proposition D.2 gives

$$\hat{\gamma}_{\ell,h} - \gamma_{\ell,h} = O_p(T^{-1/2}). \quad (64)$$

At a fixed evaluation point s , set $U_{t,s} = (\nu_t - s)/b$ and define

$$(\hat{a}_{\ell,h}(s), \hat{c}_{\ell,h}(s)) = \arg \min_{a,c} \sum_{t=1}^T K(U_{t,s}) \{\tilde{Y}_{\ell,t,h} - a - cU_{t,s}\}^2. \quad (65)$$

Since the local regressor is scaled by b , the derivative with respect to the original shock is

$$\hat{\beta}_{\ell,h}(s) = \frac{\hat{c}_{\ell,h}(s)}{b}. \quad (66)$$

This is the standard local-linear derivative construction (Fan & Gijbels, 1996; Loader, 1999). A local-constant fit has no slope coefficient and is therefore not suited to this target.

3.3 First-stage results and inferential scope

Appendix D establishes the population residualization identities, a buffered cross-fitted score bound, and a root- T expansion for the finite-dimensional Robinson coefficient. These first-stage results do not by themselves establish the sampling properties of the local-slope estimator. In particular, Proposition D.2 does not show that replacing $\gamma_{\ell,h}$ by $\hat{\gamma}_{\ell,h}$ is negligible for (66). Inference for the derivative curve and its scalar sign and magnitude summaries remains to be developed; the reported curves and projections are descriptive.

For descriptive use, evaluation points should remain in a well-supported part of the shock distribution, and applications should report the shock density or local effective counts together with bandwidth sensitivity. Estimated shocks, mass points, heaping, boundary points, and data-selected evaluation locations require arguments beyond this draft.

4 Descriptive Monte Carlo Evidence

The simulations address three questions. The first compares global, transformed-shock, and LL-LP estimators in a sign-oriented design. The second shows how state composition can generate apparent shock-size heterogeneity. The appendix reports a magnitude-only design and a fully specified sensitivity calculation for alternative symmetric targets. No interval estimator is constructed or evaluated.

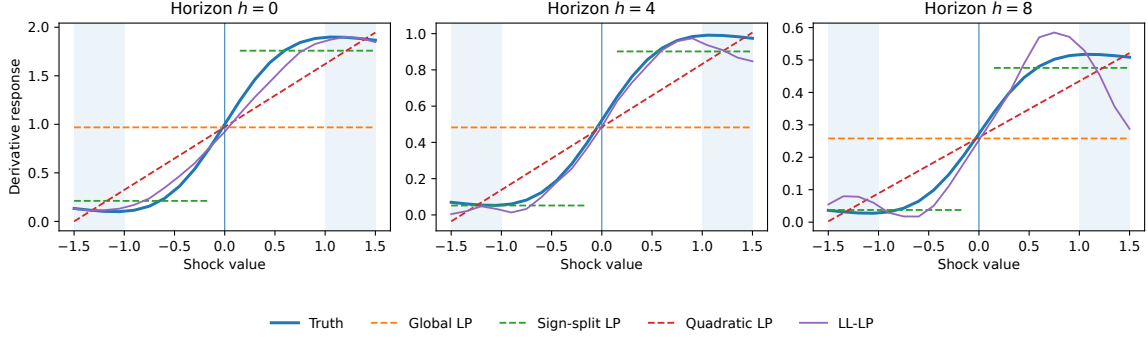


Figure 2: Repeated-sample estimator comparison in the sign-oriented design. Lines report the true derivative and Monte Carlo mean estimates at $T = 250$ and $R = 200$. Shaded regions mark the adjacent tails, $1 < |s| \leq 1.5$.

4.1 Sign-oriented nonlinear response

The repeated-sample experiment uses

$$Y_t = \sum_{j=0}^8 a_j f(X_{t-j}) + u_t, \quad a_j = 0.85^j, \quad (67)$$

$$f(x) = x\{1 + 0.75 \tanh(1.10x)\}, \quad u_t = 0.35u_{t-1} + \eta_t. \quad (68)$$

The sequences $\{X_t\}$ and $\{\eta_t\}$ are iid, mutually independent, with $X_t \sim N(0, 1)$ and $\eta_t \sim N(0, 0.60^2)$. The AR(1) disturbance is initialized from its stationary distribution, and a further 300 observations are discarded as burn-in. The target at horizon h is $a_h f'(s)$. Since

$$\frac{f'(s) + f'(-s)}{2} = 1, \quad (69)$$

all nonconstant variation belongs to the sign component.

Each of $R = 200$ replications at $T = 250$ estimates: (i) a global linear projection on $(1, X_t)$; (ii) a sign-split projection with sign-specific intercepts and slopes; (iii) a quadratic projection on $(1, X_t, X_t^2)$; and (iv) a local-linear derivative with Epanechnikov kernel, bandwidth 0.75, and 21 points on $[-1.5, 1.5]$. There are no nontrivial controls in this design, so the Robinson first step is degenerate. The sign-split derivative is left undefined at zero. Appendix E records the implementation and seed.

Figure 2 shows that the global LP cannot reproduce variation over shock values. The sign-split specification captures parts of the curve but imposes piecewise affinity, allowing different intercepts and slopes. The quadratic specification instead imposes constant curvature. LL-LP follows the smooth profile without imposing either shape, at the cost of greater repeated-sample variation.

The restricted nonlinear specifications have lower RMSE in this particular design because their low-dimensional shapes approximate the DGP well. LL-LP retains more flexibility and has higher tail RMSE because fewer observations receive substantial local weight. The comparison illustrates a flexibility–precision trade-off rather than a ranking that must hold across designs.

Table 2: Descriptive Monte Carlo performance at horizon $h = 0$ ($T = 250$, $R = 200$)

Estimator	Interior bias	Interior RMSE	Tail RMSE
Global LP	−0.031	0.662	0.895
Sign-split LP	−0.015	0.380	0.297
Quadratic LP	−0.027	0.331	0.283
LL-LP	−0.038	0.507	0.705

Notes: The interior is $|s| \leq 1$ and the adjacent tails are $1 < |s| \leq 1.5$. The zero grid point is omitted from sign-split summaries because its derivative is undefined. Entries are generated using `reproduce_figures_tables.py` and describe point-estimate bias and RMSE.

4.2 Composition-induced apparent shock-size heterogeneity

Let $\{Z_{t-1}\}$ be an iid Bernoulli(1/2) sequence and

$$X_t = \sigma_{Z_{t-1}} \varepsilon_t, \quad Y_t(x) = a_{Z_{t-1}} x + \eta_t, \quad (70)$$

where the three sequences $\{Z_{t-1}\}$, $\{\varepsilon_t\}$, and $\{\eta_t\}$ are mutually independent, $\varepsilon_t \sim N(0, 1)$ iid, $\eta_t \sim N(0, 0.60^2)$ iid, $(\sigma_0, \sigma_1) = (0.7, 1.3)$, and $(a_0, a_1) = (0.7, 1.3)$. Conditional on the state, the causal derivative is exactly a_z and does not vary with shock size. The fixed-population average derivative is therefore one.

Let $p(s) = \Pr(Z_{t-1} = 1 \mid X_t = s)$. Then

$$\mu(s) = s\{a_0 + (a_1 - a_0)p(s)\}, \quad \mu'(s) = a_0 + (a_1 - a_0)\{p(s) + sp'(s)\}. \quad (71)$$

Because tail shocks disproportionately come from the high-variance state, $p(s)$ changes with $|s|$ and $\mu'(s)$ is nonlinear. By symmetry, the population Robinson coefficient on the level control $W_{t-1} = Z_{t-1}$ is zero. Indeed, with $V_t = Z_{t-1} - p(X_t)$,

$$Y_t - \mu(X_t) = X_t(a_1 - a_0)V_t + \eta_t, \quad (72)$$

so the Robinson numerator is $(a_1 - a_0)\mathbb{E}[X_t p(X_t)\{1 - p(X_t)\}] = 0$: the marginal density of X_t and the function $p(s)\{1 - p(s)\}$ are even. Hence the population residualized target equals the raw derivative. The level adjustment does not remove the omitted state–shock interaction.

Figure 3 plots the analytic target and the Monte Carlo mean local-linear estimate. The nonlinearity is entirely compositional and can move the one-dimensional derivative outside $[a_0, a_1]$, which is impossible for a convex average of the two state-specific derivatives. This design shows why the bridge remainder must be treated as an identifying condition rather than a generic benefit of adding predetermined controls.

5 Conclusion

Nonlinear shock transmission can be organized as variation in the slope of a response curve. A conventional linear projection reports one projection coefficient. LL-LP instead targets the slope over a supported range of shock values, helping researchers explore where transmission changes with the size or sign of the shock.

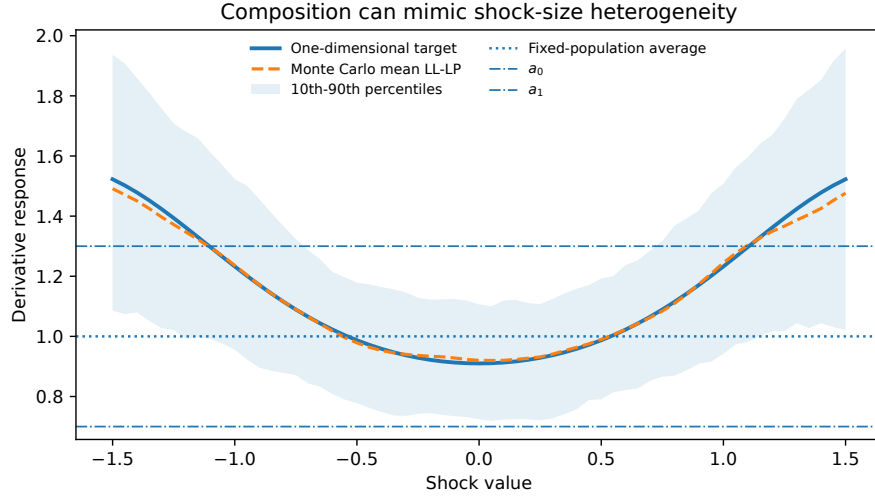


Figure 3: Composition can mimic shock-size heterogeneity. State-specific causal derivatives and their fixed-population average are constant, whereas the one-dimensional target varies because the state distribution changes with the shock value. The Monte Carlo uses $R = 400$, $T = 1000$, and a local-linear bandwidth of 0.45; the shaded region reports the 10th–90th percentiles of the descriptive point estimates.

Nonlinearity need not be present on impact. It can emerge as the shock changes economic conditions that subsequently govern transmission. Even when the impact response is linear for a given initial state, responses at later horizons can therefore be nonlinear in the initial shock. The resulting curve captures the combined effect of direct shock nonlinearity and propagation through states changed by the shock. Separating these mechanisms requires additional economic structure.

The interpretation of the one-dimensional curve also requires care. A state-conditional local slope is causal when nearby shocks are comparable after conditioning on the relevant pre-shock information. But a curve estimated after finite-dimensional residualization need not automatically average these state-conditional causal effects. If the composition of initial states changes with the shock value, it can generate apparent shock-size heterogeneity. This distinction is an identifying condition, not an automatic consequence of including controls.

The pointwise slope curve should remain the primary descriptive object, while population sign and magnitude projections provide lower-dimensional summaries whose interpretation depends on the chosen reference distribution. Reporting the supported range, reference distribution, bandwidth sensitivity, and conditions underlying any causal interpretation provides a transparent way to study nonlinear shock transmission without forcing it into a small number of predetermined categories.

References

- Ascari, G., & Haber, T. (2022). Non-linearities, state-dependent prices and the transmission mechanism of monetary policy. *Economic Journal*, 132(641), 37–57. <https://doi.org/10.1093/ej/ueab049>
- Bai, X., Laumer, S., & Rich, K. M. (2026, May 7). *Understanding the effects of government spending shocks: It's in the size as well* (Working paper No. SSRN 5925643). <https://doi.org/10.2139/ssrn.5925643>
- Barnichon, R., & Brownlees, C. (2019). Impulse response estimation by smooth local projections. *Review of Economics and Statistics*, 101(3), 522–530. https://doi.org/10.1162/rest_a.00778
- Barnichon, R., Debortoli, D., & Matthes, C. (2022). Understanding the size of the government spending multiplier: It's in the sign. *Review of Economic Studies*, 89(1), 87–117. <https://doi.org/10.1093/restud/rdab029>
- Barnichon, R., Matthes, C., & Ziegenbein, A. (2022). Are the effects of financial market disruptions big or small? *Review of Economics and Statistics*, 104(3), 557–570. https://doi.org/10.1162/rest_a.00972
- Berbee, H. C. P. (1979). *Random walks with stationary increments and renewal theory*. Mathematisch Centrum. <https://ir.cwi.nl/pub/12694>
- Caravello, T. E., & Martínez Bruera, P. (2024, June). *Disentangling sign and size non-linearities* (Working paper No. SSRN 4704050). <https://doi.org/10.2139/ssrn.4704050>
- David, J. M., Giacomini, R., Jiao, X., & Wang, W. (2026, September 3). *Causal state-dependent local projections* (Working paper No. arXiv:2605.05404v3). <https://arxiv.org/abs/2605.05404v3>
- Davydov, Y. A. (1968). Convergence of distributions generated by stationary stochastic processes. *Theory of Probability and Its Applications*, 13(4), 691–696. <https://doi.org/10.1137/1113086>
- Fan, J., & Gijbels, I. (1996). *Local polynomial modelling and its applications*. Chapman; Hall. <https://www.routledge.com/p/book/9780412983214>
- Gonçalves, S., Herrera, A. M., Kilian, L., & Pesavento, E. (2024). State-dependent local projections. *Journal of Econometrics*, 244(2), Article 105702. <https://doi.org/10.1016/j.jeconom.2024.105702>
- Gonçalves, S., Herrera, A. M., Kilian, L., Pesavento, E., & Holban, I. K. (2026). *Semiparametric local projections* (Working paper No. 2616). Federal Reserve Bank of Dallas. <https://doi.org/10.24149/wp2616>
- Gouriéroux, C., & Lee, Q. (2025, August 25). *Nonlinear impulse response functions and local projections* (Working paper No. arXiv:2305.18145v2). <https://arxiv.org/abs/2305.18145v2>
- Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. *American Economic Review*, 95(1), 161–182. <https://doi.org/10.1257/0002828053828518>
- Kilian, L., & Vigfusson, R. J. (2011). Are the responses of the U.S. economy asymmetric in energy price increases and decreases? *Quantitative Economics*, 2(3), 419–453. <https://doi.org/10.3982/QE99>
- Kolesár, M., & Plagborg-Møller, M. (2025). Dynamic causal effects in a nonlinear world: The good, the bad, and the ugly. *Journal of Business & Economic Statistics*, 43(4), 737–754. <https://doi.org/10.1080/07350015.2025.2539478>

- Koop, G., Pesaran, M. H., & Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics*, 74(1), 119–147. [https://doi.org/10.1016/0304-4076\(95\)01753-4](https://doi.org/10.1016/0304-4076(95)01753-4)
- Loader, C. (1999). *Local regression and likelihood*. Springer. <https://doi.org/10.1007/b98858>
- Potter, S. M. (2000). Nonlinear impulse response functions. *Journal of Economic Dynamics and Control*, 24(10), 1425–1446. [https://doi.org/10.1016/S0165-1889\(99\)00013-5](https://doi.org/10.1016/S0165-1889(99)00013-5)
- Rambachan, A., & Shephard, N. (2025, January 22). *When do common time series estimands have nonparametric causal meaning?* (Working paper No. arXiv:1903.01637v4). <https://arxiv.org/abs/1903.01637v4>
- Ramey, V. A., & Zubairy, S. (2018). Government spending multipliers in good times and in bad: Evidence from US historical data. *Journal of Political Economy*, 126(2), 850–901. <https://doi.org/10.1086/696277>
- Robinson, P. M. (1988). Root- N -consistent semiparametric regression. *Econometrica*, 56(4), 931–954. <https://doi.org/10.2307/1912705>
- Semenova, V., Goldman, M., Chernozhukov, V., & Taddy, M. (2023). Inference on heterogeneous treatment effects in high-dimensional dynamic panels under weak dependence. *Quantitative Economics*, 14(2), 471–510. <https://doi.org/10.3982/QE1670>

A The Symmetrized Observed Target

Let $I = [-\bar{r}, \bar{r}]$ be a symmetric interval and let

$$p_I = \Pr(|\nu_t| \leq \bar{r}) > 0. \quad (73)$$

Define

$$\varpi^{\text{sym}}(x) = \frac{f_\nu(x) + f_\nu(-x)}{2p_I} \mathbf{1}\{|x| \leq \bar{r}\}. \quad (74)$$

The normalization is unnecessary when I is the full support and $p_I = 1$.

Proposition A.1 (Magnitude-preserving symmetrization). *The density in (74) is symmetric and integrates to one. For every function g integrable under this density,*

$$\int_I g(x) \varpi^{\text{sym}}(x) dx = \frac{1}{2p_I} \mathbb{E}[\mathbf{1}\{|\nu_t| \leq \bar{r}\} \{g(|\nu_t|) + g(-|\nu_t|)\}]. \quad (75)$$

Hence the target distribution of $|x|$ equals the observed distribution of $|\nu_t|$ conditional on $|\nu_t| \leq \bar{r}$, while the target sign is independent of magnitude and assigns probability one half to each sign.

Proof. Symmetry follows by replacing x with $-x$. For the second component of the integral, substitute $u = -x$ and use symmetry of I . This gives (75); setting $g \equiv 1$ verifies normalization. Applying the identity to functions of $|x|$ preserves the conditional magnitude distribution, while applying it to sign indicators gives probability one half to each sign. $\square$

Where $f_\nu(x) > 0$, the density ratio is

$$\rho^{\text{sym}}(x) = \frac{\varpi^{\text{sym}}(x)}{f_\nu(x)} = \frac{1}{2p_I} \left\{ 1 + \frac{f_\nu(-x)}{f_\nu(x)} \right\} \mathbf{1}\{|x| \leq \bar{r}\}. \quad (76)$$

The target density in (74) is mathematically defined even when one sign has zero observed density at a particular magnitude, because it symmetrizes by borrowing mass from the opposite sign. Density-ratio reweighting requires the target measure to be absolutely continuous with respect to the observed shock distribution: $f_\nu(x) > 0$ for $\varpi^{\text{sym}}(x) dx$ -almost every x . Choose I within bilateral common support. Regularization may stabilize weights on this support, but it does not restore missing support and may alter the effective target.

B Proofs for Causal and Component Results

B.1 Proof and scaling for the triangular sufficient condition

Proof of Proposition 2.4. Use the regular conditional kernel $K_{h,z}$ in Assumption 2.4. Conditional on $Z_{t-1} = z$, equation (43) maps ε_{1t} one-to-one into X_t . Hence conditional independence gives the conditional joint law $P_{X|Z=z} \otimes K_{h,z}$ for $(X_t, \eta_{t,h})$, for P_Z -almost every z . Define the finite compatible potential mean by

$$m_h^*(u, z) = \int \Psi_h(u, z, e) K_{h,z}(de).$$

The same kernel defines the diagonal potential conditional mean. The conditional joint-law factorization and the structural outcome equation give

$$m_h(x, z) = \int \Psi_h(x, z, e) K_{h,z}(de) = m_h^*(x, z) \quad (77)$$

for $P_{X|Z=z}$ -almost every x , on a common P_Z -full set of states. This proves (45) for the compatible diagonal versions and also supplies the compatibility in Assumption 2.1. When the selected observed and potential means are continuous on a common neighborhood, positivity of the conditional density extends their almost-everywhere equality to every point there. Theorem 2.1 then yields (46).

If X_t is itself the primitive structural shock, then $\phi \equiv 0$. If instead $X_t = \phi(Z_{t-1}) + \lambda \varepsilon_{1t}$ with known $\lambda \neq 0$, the derivative with respect to the primitive innovation is $\lambda \theta_h(x, z)$. Equation (43) uses the normalization $\lambda = 1$. $\square$

B.2 Proof of the causal bridge and its corollaries

Proof of Proposition 2.5. Assumption 2.5 and the signed-measure product rule give

$$\mu'_h(s) = \int \partial_s m_h(s, z) dF_s(z) + \int m_h(s, z) d\dot{F}_s(z). \quad (78)$$

Total-variation differentiability also implies $\dot{F}_s \ll F_s$. To see this, let A be measurable with $F_s(A) = 0$. The two-sided expansion

$$F_{s+h}(A) = h\dot{F}_s(A) + o(|h|) \quad (79)$$

holds uniformly over measurable sets. Nonnegativity for $h > 0$ gives $\dot{F}_s(A) \geq 0$, while nonnegativity for $h < 0$ gives $\dot{F}_s(A) \leq 0$. Hence $\dot{F}_s(A) = 0$.

For F_s -almost every relevant state, the same local equality argument used in the proof of Theorem 2.1 gives $m_h(s, z) = m_h^*(s, z)$, while the theorem gives $\partial_s m_h(s, z) = \theta_h^*(s, z)$ for the selected local versions. Absolute continuity of $\dot{F}_s$ with respect to F_s makes the level equality valid $|\dot{F}_s|$ -almost everywhere as well. Substitution into (78) and then into (50) yields (51). Since F_s is a probability measure, the first term is the observed-composition average $\bar{\theta}_h^{\text{obs}}(s)$. The remaining bracket is $C_h(s)$, so equality to that particular average holds if and only if $C_h(s) = 0$. $\square$

Proof of Corollary 2.1. If $\dot{F}_s = 0$ and $\dot{q}(s) = 0$, the composition term in (51) is zero. If $W_{t-1} = w(Z_{t-1})$ and F_u is locally constant, then

$$q(u) = \int w(z) dF_u(z) \quad (80)$$

is locally constant under the stated integrability conditions. When Proposition 2.5 applies at every point of the reported interval, $F_u = F^0$ and constancy of $q(u)$ there give the stated fixed-population equality point by point. $\square$

Proof of Corollary 2.2. Let

$$\epsilon_{h,t} = Y_{t+h} - m_h(X_t, Z_{t-1}). \quad (81)$$

Then $\mathbb{E}[\epsilon_{h,t} | X_t, Z_{t-1}] = 0$. Under (55),

$$\mu_h(X_t) = g_h(X_t) + q(X_t)^\top \gamma_h^0 \quad (82)$$

and therefore

$$Y_{t+h} - \mu_h(X_t) = \{W_{t-1} - q(X_t)\}^\top \gamma_h^0 + \epsilon_{h,t}. \quad (83)$$

Writing $V_t = W_{t-1} - q(X_t)$, iterated expectations give $\mathbb{E}[V_t \epsilon_{h,t}] = 0$. Nonsingularity of $\mathbb{E}[V_t V_t^\top]$ therefore implies that the Robinson population coefficient is γ_h^0 . The residualized curve is consequently

$$m_h^r(u) = \mu_h(u) - q(u)^\top \gamma_h^0 = g_h(u) \quad (84)$$

for every supported u , so $\beta_h(s) = g_h'(s)$. On the common full-measure state set $\mathcal{Z}_0$, differentiation of (55) at s gives $\partial_s m_h(s, z) = g_h'(s)$. The state-conditional causal assumptions and Theorem 2.1 then yield $\theta_h^*(s, z) = g_h'(s)$ for F_s -almost every z . $\square$

B.3 Symmetry-balanced component projections

Proof of Proposition 2.2. The weighted Frisch–Waugh–Lovell theorem gives

$$\tau_{h,f}^\varpi = \frac{\int_I e_f^\varpi(x) m_h(x) \varpi(x) dx}{d_f^\varpi}. \quad (85)$$

Because e_f^ϖ is orthogonal to the intercept,

$$H_f^\varpi(-\bar{r}) = H_f^\varpi(\bar{r}) = 0. \quad (86)$$

Absolute continuity of m_h and integration by parts therefore imply

$$\int_I e_f^\varpi(x) m_h(x) \varpi(x) dx = - \int_I H_f^\varpi(x) \theta_h(x) dx, \quad (87)$$

which proves (31).

If f is even, symmetry of ϖ implies that its projection coefficient on x is zero, so e_f^ϖ is even. Its zero-total-integral primitive H_f^ϖ is odd, as is ω_f^ϖ . Pairing r and $-r$ leaves only $\theta_h^{\text{sign}}(r)$.

If f is odd, its projection intercept is zero and the residual after projection on x remains odd. Hence H_f^ϖ and ω_f^ϖ are even. Moreover, integration by parts gives

$$\int_{-\bar{r}}^{\bar{r}} H_f^\varpi(x) dx = - \int_{-\bar{r}}^{\bar{r}} x e_f^\varpi(x) \varpi(x) dx = 0, \quad (88)$$

where the final equality is projection orthogonality to x . Pairing positive and negative arguments leaves only $\theta_h^{\text{mag}}(r)$, and the zero integral removes any constant derivative component. $\square$

C Proofs for the Population Results

C.1 Proof of Proposition 1.1

Proof. For $x \neq 0$, the maintained representation gives

$$Q_h(x, z) = B_h(G_h(z, x), z). \quad (89)$$

The multivariate chain rule gives (12). The pointwise sufficient conditions and the interval qualification follow immediately. $\square$

C.2 Proof of Proposition 1.2

Proof. Necessity is immediate: if two nonzero shock values generate the same state path, evaluating B_h at that path gives the same response per unit of shock. For sufficiency, for every $g \in \mathcal{S}_z$ choose any nonzero x such that $G_h(z, x) = g$ and define

$$B_h(g, z) = \frac{R_h(x, z)}{x}. \quad (90)$$

Condition (14) makes this definition independent of the selected preimage and gives (13). If $G_h(z, \cdot)$ is one-to-one, the compatibility condition is automatic. To extend the representation at zero, enlarge the domain by $G_h(z, 0)$ if necessary and set $B_h(G_h(z, 0), z) = \partial_x R_h(0, z)$. The stated compatibility requirement prevents a conflict with a nonzero shock attaining the same state-path value. $\square$

C.3 Proof of Proposition 2.1

Proof. The conditional population projection slope is

$$\alpha_h^{\text{LP}}(z) = \frac{\text{Cov}(X_t, Y_{t+h} \mid Z_{t-1} = z)}{\sigma_z^2}. \quad (91)$$

The conditional-mean residual is orthogonal to functions of X_t , so

$$\alpha_h^{\text{LP}}(z) = \frac{1}{\sigma_z^2} \mathbb{E}[(X_t - \mu_z) m_h(X_t, z) \mid Z_{t-1} = z]. \quad (92)$$

Absolute continuity gives

$$m_h(x, z) = m_h(\underline{x}_z, z) + \int_{\underline{x}_z}^x \theta_h(u, z) \, du. \quad (93)$$

The imposed absolute integrability permits Fubini's theorem. The constant term drops out, and (92) becomes

$$\alpha_h^{\text{LP}}(z) = \int_{\underline{x}_z}^{\bar{x}_z} \theta_h(u, z) \left\{ \frac{1}{\sigma_z^2} \int_u^{\bar{x}_z} (x - \mu_z) f_z(x) \, dx \right\} \, du. \quad (94)$$

For $u \geq \mu_z$ the inner integral is nonnegative. For $u < \mu_z$, it equals

$$- \int_{\underline{x}_z}^u (x - \mu_z) f_z(x) \, dx \geq 0. \quad (95)$$

Finally, another application of Fubini gives

$$\int_{\underline{x}_z}^{\bar{x}_z} \omega_z(u) \, du = \frac{1}{\sigma_z^2} \int_{\underline{x}_z}^{\bar{x}_z} (x - \mu_z)(x - \underline{x}_z) f_z(x) \, dx \quad (96)$$

$$= \frac{\mathbb{E}[(X_t - \mu_z)(X_t - \underline{x}_z) \mid Z_{t-1} = z]}{\sigma_z^2} = 1. \quad (97)$$

This proves the result. $\square$

C.4 Proof of Theorem 2.1

Proof. By consistency and the compatible-version requirement in Assumption 2.1, for $P_{X|Z=z}$ -almost every $u \in N_x$,

$$\mathbb{E}[Y_{t+h} \mid X_t = u, Z_{t-1} = z] = \mathbb{E}[Y_{t+h}(u) \mid X_t = u, Z_{t-1} = z]. \quad (98)$$

Local mean ignorability makes the right-hand side equal to $m_h^*(u, z)$. Thus the selected versions of $m_h(u, z)$ and $m_h^*(u, z)$ agree $P_{X|Z=z}$ -almost everywhere on N_x . Their continuity and positivity of the conditional density extend equality to the whole neighborhood: a nonzero continuous difference would remain nonzero on an open interval of positive conditional probability. Differentiability of m_h^* at the interior point x then gives differentiability of the equal observed curve and (42), for the relevant states covered by the assumptions. $\square$

D Population Identities and Robinson Estimation

D.1 Scope and population objects

This appendix establishes the population residualization identities, a buffered cross-fitted score bound, and a root- T expansion for the finite-dimensional Robinson coefficient.

Let $k = (\ell, h)$ index a fixed outcome and horizon, and write $Y_{k,t} := Y_{\ell,t,h}$. Assume throughout this population subsection that $\mathbb{E} \|W_{t-1}\|^2 + \mathbb{E} |Y_{k,t}|^2 < \infty$. Define

$$\mu_k(r) = \mathbb{E}[Y_{k,t} \mid \nu_t = r], \quad q(r) = \mathbb{E}[W_{t-1} \mid \nu_t = r], \quad (99)$$

$$V_t = W_{t-1} - q(\nu_t), \quad \Sigma_W = \mathbb{E}[V_t V_t^\top]. \quad (100)$$

Assume Σ_W is positive definite. The finite second moments above ensure that the population coefficient and orthogonality moments below are well defined. The Robinson population coefficient is

$$\gamma_k = \Sigma_W^{-1} \mathbb{E}[V_t \{Y_{k,t} - \mu_k(\nu_t)\}]. \quad (101)$$

Define

$$m_k(r) = \mathbb{E}[Y_{k,t} - W_{t-1}^\top \gamma_k \mid \nu_t = r] = \mu_k(r) - q(r)^\top \gamma_k, \quad (102)$$

$$\beta_k(r) = m'_k(r), \quad (103)$$

$$u_{k,t} = Y_{k,t} - W_{t-1}^\top \gamma_k - m_k(\nu_t). \quad (104)$$

Proposition D.1 (Population orthogonality and derivative identity). *For every fixed k ,*

$$Y_{k,t} = W_{t-1}^\top \gamma_k + m_k(\nu_t) + u_{k,t}, \quad \mathbb{E}[u_{k,t} \mid \nu_t] = 0, \quad \mathbb{E}[V_t u_{k,t}] = 0. \quad (105)$$

If μ_k and q are differentiable at s , then

$$\beta_k(s) = \mu'_k(s) - \dot{q}(s)^\top \gamma_k. \quad (106)$$

Proof. The conditional-mean restriction follows directly from (102)–(104). Moreover,

$$Y_{k,t} - \mu_k(\nu_t) = V_t^\top \gamma_k + u_{k,t}. \quad (107)$$

Multiplication by V_t and (101) give $\mathbb{E}[V_t u_{k,t}] = 0$. Differentiating (102) gives (106). $\square$

Three distinct derivatives. The raw one-dimensional conditional-mean derivative is $\mu'_k(s)$. The LL-LP estimand is the residualized derivative $\beta_k(s)$. A state-conditional causal derivative such as $\partial_s \mathbb{E}[Y_{k,t}(s) \mid Z_{t-1} = z]$ is a third object. These population identities concern $\beta_k(s)$; its causal interpretation additionally requires the bridge conditions in Section 2.5.

D.2 Sufficient conditions

Assumption D.1 (Stationarity, finite dimension, and absolute regularity). For fixed H and fixed d_W , the stacked process

$$\mathcal{Z}_t = (\nu_t, W_{t-1}^\top, (Y_{y,t,h}, Y_{g,t,h})_{h=0}^H)^\top \quad (108)$$

is strictly stationary and absolutely regular with coefficients satisfying

$$\beta_{\mathcal{Z}}(j) \leq C_\beta \rho^j, \quad j \geq 1, \quad 0 < \rho < 1. \quad (109)$$

Consequently, $\sum_{j \geq 1} j^2 \beta_{\mathcal{Z}}(j)^c < \infty$ for every $c > 0$. The fixed- H stacked process already incorporates the finite overlap among horizon outcomes; a growing horizon sequence would require a separate argument.

Assumption D.2 (Moments and nonsingularity). There exists $\eta > 0$ such that, for every fixed k ,

$$\mathbb{E} \|W_{t-1}\|^{4+\eta} < \infty, \quad \mathbb{E} |Y_{k,t}|^{4+\eta} < \infty, \quad \mathbb{E} |u_{k,t}|^{4+\eta} < \infty, \quad \lambda_{\min}(\Sigma_W) > 0. \quad (110)$$

The $Y_{k,t}$ moment ensures that the conditional means and population score are well defined; the displayed $u_{k,t}$ moment is the condition used directly for the leading-score bound.

Assumption D.3 (Buffered cross-fitting and Robinson nuisance rates). Let a fixed number F_0 of contiguous evaluation blocks $I_{\ell,T}$ partition $\{1, \dots, T\}$, with $n_{\ell,T} = |I_{\ell,T}|$ and $n_{\ell,T}/T \rightarrow \pi_\ell \in (0, 1)$. The training sample for fold ℓ excludes every index within distance g_T of $I_{\ell,T}$, where $g_T \rightarrow \infty$ and $g_T = o(T)$. Let $\mathcal{A}_{\ell,T}$ be the sigma-field generated by the training observations and any learner randomization independent of the data. The estimators $\hat{q}_\ell$ and $\hat{\mu}_{k,\ell}$ are $\mathcal{A}_{\ell,T}$ -measurable.

For each fold, work on an extension of the probability space supporting a coupled copy $\{\mathcal{Z}_t^{*(\ell)} : t \in I_{\ell,T}\}$ that has the same joint law as the observed evaluation block, is independent of $\mathcal{A}_{\ell,T}$, and satisfies

$$\Pr \left(\exists t \in I_{\ell,T} : \mathcal{Z}_t^{*(\ell)} \neq \mathcal{Z}_t \right) \leq C\beta_{\mathcal{Z}}(g_T). \quad (111)$$

This construction follows the neighbors-left-out sample-splitting logic of Semenova et al. (2023). A direct total-variation argument controls the middle block jointly with both training sides. Let L , M , and R collect respectively the left training observations, the evaluation block, and the right training observations. Write $d_{\text{TV}}(P, Q) = \sup_B |P(B) - Q(B)|$ and use the corresponding absolute-regularity

convention. All product laws below are placed in coordinate order (L, M, R) . The triangle inequality gives

$$\begin{aligned} d_{\text{TV}}(P_{LMR}, P_M \otimes P_{LR}) &\leq d_{\text{TV}}(P_{LMR}, P_L \otimes P_{MR}) \\ &\quad + d_{\text{TV}}(P_L \otimes P_{MR}, P_L \otimes P_M \otimes P_R) \\ &\quad + d_{\text{TV}}(P_L \otimes P_M \otimes P_R, P_M \otimes P_{LR}) \\ &\leq 3\beta_{\mathcal{Z}}(g_T). \end{aligned}$$

Indeed, the three terms are respectively the absolute-regularity coefficients between L and (M, R) , between M and R , and between L and R . Each is bounded by $\beta_{\mathcal{Z}}(g_T)$ using an ordered past–future cut and monotonicity in the sigma-fields. Adjoining learner randomization independent of the entire data does not increase the coefficient. Therefore

$$\beta(\sigma(\mathcal{Z}_t : t \in I_{\ell,T}), \mathcal{A}_{\ell,T}) \leq 3\beta_{\mathcal{Z}}(g_T). \quad (112)$$

An edge block needs only one past–future cut. Applying Berbee’s coupling lemma (Berbee, 1979) to the evaluation block and the joint training sigma-field yields (111). The copy preserves dependence within the evaluation block; it is independent of training, not internally iid. Since F_0 is fixed and mixing is geometric, the sum of fold-specific coupling-failure probabilities is $o(1)$. The constant 3 leaves all stated rates unchanged.

For a fold-specific random function $g_{\ell,T}$ and a measurable function $S_t^{*(\ell)}$ of the coupled evaluation observation, define

$$\|g_{\ell,T}\|_{P,r} = \left\{ \mathbb{E} \left[\left\| g_{\ell,T}(S_t^{*(\ell)}) \right\|^r \mid \mathcal{A}_{\ell,T} \right] \right\}^{1/r}. \quad (113)$$

Because the coupled evaluation block is independent of the training sigma-field and has the stationary evaluation law, this conditional norm is well defined; the conditioning is suppressed for nonrandom functions.

For $t \in I_{\ell,T}$, define

$$e_{q,t} = \hat{q}_{\ell}(\nu_t) - q(\nu_t), \quad e_{m,k,t} = \{\hat{\mu}_{k,\ell}(\nu_t) - \mu_k(\nu_t)\} - e_{q,t}^{\top} \gamma_k. \quad (114)$$

For some $\delta > 0$, uniformly over fixed folds and fixed k ,

$$\|e_{q,\ell}\|_{P,4+\delta} + \|e_{m,k,\ell}\|_{P,4+\delta} = o_p(1), \quad \sqrt{T} \|e_{q,\ell}\|_{P,2} \|e_{m,k,\ell}\|_{P,2} = o_p(1). \quad (115)$$

On the score support, for some possibly smaller $\delta > 0$,

$$\sup_s \mathbb{E}[\|V_t\|^{2+\delta} \mid \nu_t = s] < \infty, \quad \sup_s \mathbb{E}[|u_{k,t}|^{2+\delta} \mid \nu_t = s] < \infty. \quad (116)$$

The results below use the untrimmed population and sample scores as displayed. No trimming extension is claimed.

D.3 Buffered empirical-process bounds and the Robinson step

Lemma D.1 (Buffered cross-fitted score bound). *Maintain Assumptions D.1 and D.3. Let S_t and ζ_t be fixed-dimensional measurable functions of $\mathcal{Z}_t$. For fold ℓ , let $g_{\ell,T}$ be a fixed-dimensional random func-*

tion of S_t , measurable with respect to $\mathcal{A}_{\ell,T}$. Assume the displayed product $g_{\ell,T}(S_t)\zeta_t$ is conformable; in the applications below one factor is scalar, and vector or matrix conclusions are interpreted coordinatewise.

(i) If $\mathbb{E}[\zeta_t | S_t] = 0$, $\sup_s \mathbb{E}[\|\zeta_t\|^{2+\delta_0} | S_t = s] < \infty$ for some $\delta_0 > 0$, and

$$\max_{\ell \leq F_0} \|g_{\ell,T}\|_{P,2+\delta_0} = o_p(1), \quad (117)$$

then

$$\frac{1}{\sqrt{T}} \sum_{\ell=1}^{F_0} \sum_{t \in I_{\ell,T}} g_{\ell,T}(S_t)\zeta_t = o_p(1). \quad (118)$$

(ii) If $\max_{\ell \leq F_0} \|g_{\ell,T}\|_{P,2+\delta_0} = o_p(1)$, then

$$\frac{1}{\sqrt{T}} \sum_{\ell=1}^{F_0} \sum_{t \in I_{\ell,T}} \{g_{\ell,T}(S_t) - P_{\ell}g_{\ell,T}\} = o_p(1), \quad (119)$$

where

$$P_{\ell}g_{\ell,T} = \mathbb{E}[g_{\ell,T}(S_t^{*(\ell)}) | \mathcal{A}_{\ell,T}]. \quad (120)$$

Both conclusions hold coordinatewise for fixed-dimensional vectors and matrices.

Proof. Prove the claim one fold at a time; no joint coupling of all folds is required. Fix ℓ and work first with the coupled evaluation block. Conditional on $\mathcal{A}_{\ell,T}$, that block has its stationary joint law, is independent of the random function $g_{\ell,T}$, and has mixing coefficients bounded by those in Assumption D.1.

For part (i), let $X_t^* = g_{\ell,T}(S_t^{*(\ell)})\zeta_t^{*(\ell)}$ and set $r = 2 + \delta_0$. Conditional centering follows from

$$\mathbb{E}[X_t^* | \mathcal{A}_{\ell,T}] = \mathbb{E}[g_{\ell,T}(S_t^{*(\ell)})\mathbb{E}[\zeta_t^{*(\ell)} | S_t^{*(\ell)}] | \mathcal{A}_{\ell,T}] = 0. \quad (121)$$

The conditional moment bound gives

$$\|X_t^*\|_{r|\mathcal{A}_{\ell,T}} \leq C \|g_{\ell,T}\|_{P,r}. \quad (122)$$

Davydov's covariance inequality (Davydov, 1968), together with $\alpha(j) \leq \beta_{\mathcal{Z}}(j)$, yields

$$|\text{Cov}(X_t^*, X_{t-j}^* | \mathcal{A}_{\ell,T})| \leq C \beta_{\mathcal{Z}}(j)^{1-2/r} \|g_{\ell,T}\|_{P,r}^2. \quad (123)$$

The geometric coefficients are summable at exponent $1 - 2/r = \delta_0/(2 + \delta_0) > 0$, so

$$\text{Var}\left(\frac{1}{\sqrt{n_{\ell,T}}} \sum_{t \in I_{\ell,T}} X_t^* \middle| \mathcal{A}_{\ell,T}\right) \leq C \|g_{\ell,T}\|_{P,r}^2 = o_p(1). \quad (124)$$

To pass from the random conditional bound to an unconditional probability statement, for any $\epsilon, \xi > 0$ use conditional Chebyshev on the event that the right-hand side of (124) is at most ξ :

$$\Pr\left(\left|n_{\ell,T}^{-1/2} \sum X_t^*\right| > \epsilon\right) \leq \Pr(C \|g_{\ell,T}\|_{P,r}^2 > \xi) + \xi/\epsilon^2. \quad (125)$$

First let $T \rightarrow \infty$ and then $\xi \downarrow 0$. This proves the coupled fold bound. Since $n_{\ell,T}/T \rightarrow \pi_\ell$, the same conclusion holds under $T^{-1/2}$ normalization.

For part (ii), apply the same argument to

$$X_t^* = g_{\ell,T}(S_t^{*(\ell)}) - P_\ell g_{\ell,T}. \quad (126)$$

Its conditional L^r norm is bounded by $2 \|g_{\ell,T}\|_{P,r}$.

Finally, transfer each fold-specific result from the coupled block to the observed block using (111). Combine the fixed number of foldwise bounds by the triangle inequality and sum their coupling-failure probabilities. This sum is $o(1)$, proving (118)–(119). $\square$

Let $\ell(t)$ denote the evaluation fold containing t and define

$$\widehat{V}_t = W_{t-1} - \widehat{q}_{\ell(t)}(\nu_t), \quad \widehat{\Sigma}_W = \frac{1}{T} \sum_{t=1}^T \widehat{V}_t \widehat{V}_t^\top. \quad (127)$$

On the event that $\widehat{\Sigma}_W$ is nonsingular, define

$$\widehat{\gamma}_k = \widehat{\Sigma}_W^{-1} \frac{1}{T} \sum_{t=1}^T \widehat{V}_t \{Y_{k,t} - \widehat{\mu}_{k,\ell(t)}(\nu_t)\}. \quad (128)$$

Proposition D.2 (Root- T Robinson expansion). *Under Assumptions D.1–D.3, for every fixed k ,*

$$\sqrt{T}(\widehat{\gamma}_k - \gamma_k) = \Sigma_W^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^T V_t u_{k,t} + o_p(1). \quad (129)$$

In particular, $\widehat{\gamma}_k - \gamma_k = O_p(T^{-1/2})$.

Proof. For $t \in I_{\ell,T}$, $\widehat{V}_t = V_t - e_{q,t}$. At the true coefficient,

$$Y_{k,t} - \widehat{\mu}_{k,\ell(t)}(\nu_t) - \widehat{V}_t^\top \gamma_k = u_{k,t} - e_{m,k,t}. \quad (130)$$

The score therefore decomposes exactly as

$$\frac{1}{T} \sum_t \widehat{V}_t \{Y_{k,t} - \widehat{\mu}_{k,\ell(t)}(\nu_t) - \widehat{V}_t^\top \gamma_k\} \quad (131)$$

$$= \frac{1}{T} \sum_t V_t u_{k,t} - \frac{1}{T} \sum_t V_t e_{m,k,t} - \frac{1}{T} \sum_t e_{q,t} u_{k,t} + \frac{1}{T} \sum_t e_{q,t} e_{m,k,t}. \quad (132)$$

Lemma D.1(i), first with $S_t = \nu_t$, $g = e_{m,k,\ell}$, and $\zeta_t = V_t$, and then with $g = e_{q,\ell}$ and $\zeta_t = u_{k,t}$, gives

$$T^{-1/2} \sum_t V_t e_{m,k,t} = o_p(1), \quad T^{-1/2} \sum_t e_{q,t} u_{k,t} = o_p(1). \quad (133)$$

For the product term, let $g_{\ell,T}(\nu) = e_{q,\ell}(\nu) e_{m,k,\ell}(\nu)$. Hölder's inequality and (115) imply

$$\|g_{\ell,T}\|_{P,2+\delta/2} \leq \|e_{q,\ell}\|_{P,4+\delta} \|e_{m,k,\ell}\|_{P,4+\delta} = o_p(1). \quad (134)$$

Lemma D.1(ii) yields

$$\frac{1}{T} \sum_t e_{q,t} e_{m,k,t} = \sum_{\ell=1}^{F_0} \frac{n_{\ell,T}}{T} P_{\ell}(e_{q,\ell} e_{m,k,\ell}) + o_p(T^{-1/2}). \quad (135)$$

The population term is $o_p(T^{-1/2})$ by Cauchy–Schwarz and the product-rate condition in (115).

For the Gram matrix,

$$\widehat{V}_t \widehat{V}_t^{\top} = V_t V_t^{\top} - V_t e_{q,t}^{\top} - e_{q,t} V_t^{\top} + e_{q,t} e_{q,t}^{\top}. \quad (136)$$

The stationary absolute-regularity law of large numbers gives $T^{-1} \sum_t V_t V_t^{\top} \rightarrow_p \Sigma_W$. Lemma D.1(i), using $\mathbb{E}[V_t | \nu_t] = 0$, makes the two cross terms $o_p(1)$. For the quadratic term, Hölder’s inequality gives $\|e_{q,\ell} e_{q,\ell}^{\top}\|_{P,2+\delta/2} = o_p(1)$; Lemma D.1(ii) controls its centered sample average, while

$$\|P_{\ell}(e_{q,\ell} e_{q,\ell}^{\top})\| \leq \|e_{q,\ell}\|_{P,2}^2 = o_p(1). \quad (137)$$

Hence $\widehat{\Sigma}_W \rightarrow_p \Sigma_W$ and is nonsingular with probability approaching one.

The normal equations and (132) now give (129) by Slutsky’s theorem. It remains only to verify the order of the leading score. Jensen’s inequality implies $\mathbb{E} \|V_t\|^{4+\eta} < \infty$. Hölder’s inequality then gives

$$\mathbb{E} \|V_t u_{k,t}\|^{2+\eta/2} < \infty. \quad (138)$$

Since $\mathbb{E}[V_t u_{k,t}] = 0$, Davydov’s inequality with exponent $\eta/(4+\eta) > 0$ and geometric absolute regularity gives summable covariances. Therefore

$$\text{Var} \left(T^{-1/2} \sum_{t=1}^T V_t u_{k,t} \right) = O(1), \quad (139)$$

so the leading score is $O_p(1)$. This proves root- T consistency of the finite-dimensional Robinson coefficient only; no conclusion about the local-linear derivative estimator follows from this proposition alone. $\square$

E Additional Descriptive Monte Carlo Evidence

E.1 Implementation details

All figures and tables in this section are generated using `reproduce_figures_tables.py`. The sign-oriented design uses seed 412731, $R = 200$, $T = 250$, a 300-observation burn-in, and the estimator definitions stated in Section 4. The local-linear fit uses the Epanechnikov kernel

$$K(u) = \frac{3}{4} (1 - u^2) \mathbf{1}\{|u| \leq 1\} \quad (140)$$

with bandwidth 0.75. Performance is evaluated at 21 equally spaced points on $[-1.5, 1.5]$. The comparison design uses seed 725912, $R = 400$, $T = 1000$, 61 evaluation points, and bandwidth 0.45. The numerical choices are finite-sample tuning values for descriptive comparisons, not an asymptotic bandwidth sequence.

E.2 Sensitivity to the symmetric target

This subsection examines how the symmetric reference distribution affects the population sign and magnitude projections. Let the observed shock density be

$$f_\nu(x) = 0.70 \phi_{-0.35, 0.55}(x) + 0.30 \phi_{0.80, 0.85}(x), \quad (141)$$

where $\phi_{\mu, \sigma}$ denotes a Gaussian density with mean μ and standard deviation σ . All targets are restricted and normalized on $I = [-2.5, 2.5]$. The sign-oriented response is

$$m_S(x) = x\{1 + 0.75 \tanh(1.10x)\}, \quad (142)$$

and its scalar summary is the coefficient on x^2 in a target-weighted projection on $(1, x, x^2)$. The magnitude-oriented response is

$$m_M(x) = x + 0.90 \left\{ x - \frac{\tanh(1.10x)}{1.10} \right\}, \quad (143)$$

and its scalar summary is the coefficient on x^3 in a target-weighted projection on $(1, x, x^3)$. The targets are the symmetrized observed density from (74) and centered Gaussian densities with standard deviations 0.6, 0.8, and 1.0, each truncated and renormalized on I .

Table 3: Sensitivity of population symmetry-balanced projections to the target distribution

Component	Symmetric target	Population coefficient
Sign	Gaussian, $\sigma = 0.6$	0.490
	Gaussian, $\sigma = 0.8$	0.408
	Gaussian, $\sigma = 1.0$	0.368
	Symmetrized observed	0.388
Magnitude	Gaussian, $\sigma = 0.6$	0.132
	Gaussian, $\sigma = 0.8$	0.098
	Gaussian, $\sigma = 1.0$	0.084
	Symmetrized observed	0.090

Notes: Coefficients are deterministic numerical quadratures of the defined population projections. Their variation reflects how the target weights central and tail magnitudes; these are population coefficients rather than estimates from repeated samples.

E.3 Magnitude-only design

The magnitude-only design replaces the response function in (67) by

$$f_M(x) = x + \gamma_M \left\{ x - \frac{\tanh(\lambda x)}{\lambda} \right\}, \quad f'_M(x) = 1 + \gamma_M \tanh^2(\lambda x), \quad (144)$$

with $\gamma_M = 0.90$ and $\lambda = 1.10$. Its derivative is even and rises with $|x|$. The experiment uses seed 199731, $R = 300$, $T = 250$, and the same innovation law and stationary initialization as the sign-oriented design. It compares the global and sign-split LPs, a cubic projection on $(1, x, x^3)$, an odd

piecewise-linear projection on

$$(1, x, \text{sgn}(x)(|x| - 0.75)_+), \quad (145)$$

and LL-LP. The derivative of the odd piecewise component changes at ± 0.75 ; the two-sided derivative at either knot need not exist.

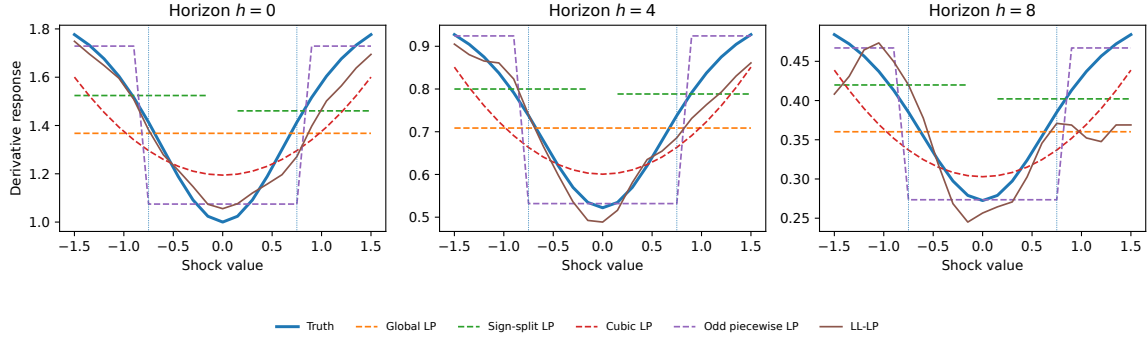


Figure 4: Magnitude-only response at $T = 250$ and $R = 300$. Lines report true and Monte Carlo mean derivative curves. Dotted vertical lines mark the odd piecewise-linear knots at ± 0.75 . Values plotted at the knots use the inner-segment slope solely as a plotting convention.

The odd transformations summarize magnitude dependence under the symmetric Gaussian shock design, while LL-LP estimates a smooth profile over shock values. As in the sign-oriented experiment, the comparison concerns point-estimate performance.