Uncovering Nonlinearities with Local Projections

Paul Bousquet*

November 10, 2025
[Click here for latest version]

Abstract

A common approach to estimate the effects of macroeconomic policy is to use a time series of plausibly exogenous policy changes ("shocks") in a linear regression framework. This pairing can be powerful: under standard identification assumptions, linear methods estimate a weighted average of a shock's true marginal effects, even when the effects are arbitrarily non-linear. The price of a vanilla regression's lack of sensitivity is a black box, as one point estimate cannot reveal where and to what extent nonlinearities exist. I show how to exploit the mechanics of local projection and develop specifications to jointly test if marginal effects have sign and size dependence. Using monetary policy shocks as an application, I find persistent nonlinearities in US data that cannot be replicated by a New Keynesian model with asymmetric rigidities in price and wage setting.

1 Introduction

Many of the most important advances in economics amount to progress in the struggle to identify and estimate the effects of policy. A large literature surveyed in Ramey (2016) and Montiel Olea et al. (2025) has sought to circumvent the eternal endogeneity problem by constructing time series of plausibly exogenous shifts in policy ("shocks") through various identification strategies. Recent work by Rambachan and Shephard (2025) and Kolesár and Plagborg-Møller (2025) looks "under the hood" of these efforts and underscores their potential power: simple Jordà (2005) local projections (LP) or vector autoregressions (VAR) estimate a weighted average of a shock's true marginal effects, even if there are arbitrary nonlinearities. The precedent for the "weighted average" interpretation comes from well-known results in microeconometrics (e.g., Imbens and Angrist, 1994), but its utility has been revived for macroeconomists thanks to assumptions researchers are willing to make about these policy shock series. Kolesár and Plagborg-Møller (2025) demonstrate if the shock series is indeed independent of the other components of an outcome variable's data generating process, the weights themselves can be easily estimated and will also be non-negative, meaning we can recover estimates of causal effects.

Though a regression of y_{t+h} on x_t can remarkably estimate a weighted average of a possibly non-linear object of interest, this does not help answer whether nonlinearities exist to begin with. Even though we have an approximation for the weights on x_t, the marginal effects at x_t are still unknown. The properties of marginal effects are important. Linear models are popular in macroeconomics for their transparent dynamics and computational efficiency, but this simplicity comes with a cost. Optimal policy in a linear world often amounts to just pushing hard enough on the right instruments, but in practice policymakers must constantly confront diminishing returns to intervention (e.g., "pushing on a string" in Fisher, 1935). Other than comparing linear estimation to some non-linear benchmark, there's no formally established, straightforward procedure to test if nonlinearities are present in policy transmission. Some attempts have been made to gauge non-linear behavior with intuitively specified local projections, but Caravello and Martínez Bruera (2024) show the lack of explicit econometric backing can lead to incorrect inference.[1]

This paper presents a method to detect *size effects* (disproportionate impact of big and small shocks) and *sign effects* (asymmetry of positive and negative shocks) in data. The approach extends the work of Rambachan and Shephard (2025) and Kolesár and Plagborg-Møller (2025) by showing that if the right functions of x_t are included in a local projection, the corresponding weights will have properties that allow for testing for sign and size effects. In related work, Caravello and Martínez Bruera (2024) show several properties of weights in LPs and show it's easy to conflate sign and size. They recommend two independent series of regressions that will separate sign and size effects if the shock's distribution is symmetric. This paper's approach allows for joint testing of sign and size with the same separation result, while also being less sensitive to the shock's distribution and offering more appealing

---

[1]Some examples of using LP to assess nonlinearities can be found in Tenreyro and Thwaites (2016); Ascari and Haber (2022); Alessandri et al. (2025) for monetary policy shocks and Ramey and Zubairy (2018); Ben Zeev et al. (2023) for government spending shocks.

finite sample properties (e.g., speed of weight convergence) and interpretability. The proposed framework is also related to semiparametric estimation with prespecified weight functions (see Kolesár and Plagborg-Møller (2025) for an overview), but given data limitations common in macroeconomic settings, it's convenient to have a stripped down procedure with a similar objective.

Formally, the purpose of the paper is to show what functional regressors have the best weighting properties to detect nonlinearities in LPs. The basis for this task is implicit regression weights depend only on the shock (not the outcome variable), so the goal is to find specifications placing weight in the desired parts of a shock's support. Broadly, if we consider 4 types of shocks along the dimensions of big vs. small and positive vs. negative, there should be 4 corresponding coefficients in the regression. "Corresponding" in this context means including just the right combinations of regressors so that, for example, the regression weights $\omega(\varepsilon)$ on the big, positive shock coefficient are only non-zero for ε sufficiently large. Appropriate weighting justifies labeling $\beta_{i,j}$ with combinations of $i = \{\text{big}, \text{small}\}$ and $j = \{\text{positive}, \text{negative}\}$. Testing for nonlinearities is then a simple task: for size effects, the null hypothesis is $\beta_{\text{big},j} = \beta_{\text{small},j}$ and for sign effects it's $\beta_{i,\text{pos}} = \beta_{i,\text{neg}}$. An advantage is coefficient *differences* may be significant even if the underlying point estimates aren't.

For implementation, a naive starting point of disjoint indicator functions turns out to be a safe way to carry out the procedure: under an arbitrary shock distribution, there will be no false positives in population estimates and the weights converge quickly in finite samples. Ideally, we would like to also be robust to false negatives and avoid having to set fixed thresholds for defining a "big" shock (i.e., the paradox of the heap).[2] I also provide more general ways to construct functional regressors with desirable properties, but given the sample sizes in most settings, simple indicator functions may often be the best choice. Similar limitations arise for trying to expand the list of categories from big vs. small and positive vs. negative, but making these generalizations more tractable could be fruitful explorations in future research.[3]

For readers solely interested in applying these methods, Section 4 gives an implementation guide and an illustration with monetary policy shocks. I find nonlinearities for all variables generally peaking in the medium to long-run, with firmest indications for size effects for both positive and negative shocks and sign effects in big shocks. Barnichon and Matthes (2018) find opposing sign effects for unemployment and inflation and broadly conjecture that a New Keynesian model with asymmetric adjustment costs in price and wage setting (Kim and Ruge-Murcia, 2009) should be able to deliver whatever size and sign effects are found. I use a Metropolis-Hastings routine to estimate the Aruoba et al. (2017) extension of the model and confirm certain parametrization can deliver various nonlinearities on impact but they quickly vanish. This disconnect lends support to Friedman (1960)'s "long and variable lags", but in an era where central banks don't exert control over monetary aggregates (Cochrane, 2024), it's not clear what mechanism would yield such a transmission path.

---

[2]"One grain of sand is not a heap of sand, two grains of sand is not a heap of sand,..., one million grains of sand is a heap of sand"

[3]A promising new strand of literature has tried to make progress on estimating state, size, and sign dependent effects (Barnichon and Matthes, 2018; Gonçalves et al., 2024). General non-linear estimation is difficult and often sensitive to misspecification (White, 1980).

2 Current Paradigm

2.1 Environment

Consider an arbitrary data generating process (DGP) $\psi_h : \mathbb{R} \times \mathbb{R}^L \to \mathbb{R}$ for an outcome variable Y at time $t + h$

$$Y_{t+h} = \psi_h(\varepsilon_t, \boldsymbol{S_{t+h}}) \tag{1}$$

Here, ε_t is the structural shock of interest at time t and $\boldsymbol{S_{t+h}}$ is "everything else" in the system, which could for instance include the information set at time t as well as leads and lags of ε_t (and other shocks). Following Rambachan and Shephard (2025) and Kolesár and Plagborg-Møller (2025), the working definition of a shock, with respect to a DGP of the form in (1), is that it satisfies $\varepsilon_t \perp \boldsymbol{S_{t+h}} \ \forall h \geq 0$. In that case, note that the conditional mean $m_h(a) \equiv \mathbb{E}[\psi_h(a, \boldsymbol{S_{t+h}}) | \varepsilon_t = a]$ is equal to the average structural function $\Psi_h(a) \equiv \mathbb{E}[\psi_h(a, \boldsymbol{S_{t+h}})]$.

Now we turn to the estimands of interest. For a group of N functions $\{f_i(\cdot)\}_{i=1}^{N}$ and control set $\boldsymbol{W_t}$, suppose we regress Y_{t+h} on $\left\{1, \{f_i(\varepsilon_t)\}_{i=1}^{N}, \boldsymbol{W_t}\right\}$. The specification is

$$Y_{t+h} = \alpha + \beta_1 f_1(\varepsilon_t) + \cdots + \beta_N f_N(\varepsilon_t) + \gamma' \boldsymbol{W_t} + u_{t+h} \tag{2}$$
$$= \alpha + \boldsymbol{\beta}' \boldsymbol{X_t} + \gamma' \boldsymbol{W_t} + u_{t+h}$$

where $\boldsymbol{X_t}$ is a concatenation of $\{f_i(\varepsilon_t)\}_{i=1}^{N}$. If ε_t is a shock and continuously distributed on an interval $I \subset \mathbb{R}$, Kolesár and Plagborg-Møller (2025)'s Proposition 1 can be extended to show that

$$\beta_i = \int_I \omega_i(a) \cdot m_h'(a) \mathrm{d}a \tag{3}$$
$$\text{with} \ \ \omega_i(a) = \frac{\mathrm{Cov}(\mathbf{1}_{\{a \leq \varepsilon_t\}}, X_i^{\perp})}{\mathrm{Var}(X_i^{\perp})} \tag{4}$$

where $X_i^{\perp}$ is the residual from projecting the ith element of $\boldsymbol{X_t}$ on the remaining $N - 1$ elements.[4] Because ε_t is a shock, estimands are a weighted average of $\Psi_h'(\cdot)$, the data generating process's true marginal effects. In Appendix A.4, the Frisch-Waugh-Lovell truncation is expanded to provide more explicit closed form solutions.

Estimands are a weighted average of marginal effects that can be arbitrarily non-linear, but estimation is a black box with output that sheds no light on the existence of nonlinearities, namely *size effects* (disproportionate impact of big and small shocks) and *sign effects* (asymmetric impact of positive and negative shocks). This representation also contrasts with a "unit change" interpretation of regression, as each β_i is a weighted average of the *same* object (marginal effects of ε_t, not $f_i(\varepsilon_t)$). So picking functions for inclusion in (2) boils down to weight engineering. Several more implications of (3) and (4) are explored in the paper. (i) the weights only depend on ε_t but may not if ε_t is not a shock (ii) if we instead use a proxy z_t in place of ε_t (if ε_t is not observable), the weights still depend on ε_t (iii) the estimand's form says nothing about the finite sampling properties of an estimator $\widehat{\beta}_i$.

---

[4]And a constant. Also need $\{f_i(\varepsilon_t)\}_{i=1}^{N}$ s.t rank condition holds, as will be discussed in Online Appendix B.2. Notice the weights can be easily estimated using a loop of OLS regressions, see Kolesár and Plagborg-Møller (2025) for more details.

2.2 Past Efforts To Estimate and Identify Non-Linear Marginal Effects

A large literature in applied macroeconomics has tried to estimate the effects of policy (e.g., interest rates or government spending) by using Jordà (2005) local projection or vector autoregression in conjunction with a constructed shock series meant to represent plausibly exogenous changes (e.g., Romer and Romer, 2004). The default is to use a completely linear structure. Relative to the framework of (2), this means the only regressors are the shock itself (identity function) and the control set. Some work has included other functions of the shock, like $f(\varepsilon) = \varepsilon^2$, in addition to the identity function in an attempt to capture non-linear effects of shocks. Caravello and Martínez Bruera (2024) provide a survey of many past efforts and find such specifications are sometimes incorrectly characterized. They consider a special case of (2)

$$Y_{t+h} = \alpha + \beta_1 \varepsilon_t + \beta_2 f(\varepsilon_t) + \gamma' W_t + u_{t+h} \tag{5}$$

With respect to (5), they show if ε_t is a shock that follows a symmetric distribution then

(i): $f(\cdot)$ is even & DGP features no sign effects $\implies \beta_2 = 0$

(ii): $f(\cdot)$ is odd & DGP features no size effects $\implies \beta_2 = 0$

These results provide important clarity on past work (e.g., ε^2 as a regressor isn't informative about size effects) and a strategy to test for nonlinearities. Because these statements hold regardless of the DGP's other properties, the presence of sign-dependence won't distort the detection of size-dependence and vice versa. While this separation property is valuable, it still leaves some questions unanswered. For example, if we include $f(\varepsilon) = \varepsilon^3$ and reject the null hypothesis that $\beta_2 = 0$, we might feel comfortable concluding there are size effects but cannot say more. There are many possibilities for the nature of the nonlinearity – in the extreme case, only negative shocks have size effects (and positive shocks don't) or vice versa. These possibilities, which we can't distinguish between at present, carry vastly different implications. This is also merely an identification result; it says nothing about finite sample properties of hypothesis testing coefficients in (5). Later parts of the paper will show simulations illustrating instances where performance may be lacking, even in ideal circumstances where the identification results hold exactly because the shock is symmetrically distributed. As the distribution becomes more asymmetric, as is the case for the monetary policy shock application in Section 4, their approach is less useful.

Besides a conflation of size and sign effects, some past work with specifications like (2) incorrectly ascribed causal meaning to the estimands. Kolesár and Plagborg-Møller (2025) show that unless the data generating process (1) matches the regression structure exactly, causal inference is not possible. For example, suppose we use (5) with $f(\varepsilon) = \varepsilon^2$. Unless the conditional mean of Y is a quadratic function in ε, $\beta_1 + 2\beta_2\varepsilon$ does not have a useful interpretation. This is because a corollary to their Proposition 1 is in specification (5), there must be negative weight placed on β_2 (see Appendix A.2 for a proof). In general, specifications that include functions of ε as regressors cannot be used to estimate causal effects (White, 1980) but are rather a means to detect nonlinearities.

In sum, linear regression is a surprisingly powerful tool for estimating non-linear marginal effects of a shock. The important qualifier is estimates represent an approximation to a weighted average across a shock's entire support. While the weights' form is known, underlying marginal effects are not; in other words $\sum_i^M \omega_i \cdot m_i' = \beta$ is still one equation with M unknowns. Recovering the exact marginal effect of a given value of ε is not possible, but it is possible to test whether the marginal effect function is non-linear by augmenting linear regressions with the proper functions. There does seem to be room to expand past approaches along the extensive margin (i.e., what kinds of nonlinearities) which may even open the door to statements about the intensive margin (i.e., how non-linear). The rest of the paper will focus on how to use linear regression to be more descriptive about the types of nonlinearities that exist in a DGP.

2.3   An Illustration of The Problem

The objective of Section 2 is to describe the status quo as concisely as possible, which thus far mostly involved extending the analysis of more technical papers like Rambachan and Shephard (2025) and Kolesár and Plagborg-Møller (2025). But what it means to have a weighting scheme $\frac{\text{Cov}(\mathbf{1}_{\{a \leq \varepsilon_t\}}, X_i^\perp)}{\text{Var}(X_i^\perp)}$ is not obvious, so an example is useful. Under the following DGP[5]

$$x_t = \varepsilon_t^d, \ \pi_t = c(x_t) + \beta \mathbb{E}_t[\pi_{t+1}] + \varepsilon_t^s \quad \text{where } \varepsilon_t^d \sim \mathcal{U}[-a, a], \ \varepsilon_t^s \sim \mathcal{N}(0, \sigma^2), \ c(x) = \begin{cases} \kappa x^b & \text{if x} > 0 \\ 0 & \text{o.w} \end{cases} \quad (6)$$

note that $\mathbb{E}_t[\pi_{t+1}]$ will be a constant. So a regression of π_t on x_t, or functions of x_t, should be revealing. Because of the simple structure, we might expect a specification of

Example 1: $$\pi_t = \alpha + \beta_1 x_t \cdot \mathbf{1}_{x_t \leq 0} + \beta_2 x_t \cdot \mathbf{1}_{x_t > 0} + u_t$$

to perform well in estimating marginal effects. In the context of the previous discussion, the logic is the following: (3) showed regression estimands are weighted averages, so shouldn't weight only be placed where the indicator functions are equal to 1 (active)? But this is not the case. Using the form in (4) we can plot the weights. Figure 1a shows that while the *aggregate* weight where the indicators are not active is indeed 0, this is only because there is positive and negative weight that cancels out. For the estimand on $y \cdot \mathbb{1}_{x>0}$, there is no issue because marginal effects are 0 where the indicator is not active. However, $\widehat{\beta}_1$ will not converge to 0 unless marginal effects are constant for $x > 0$ (i.e., only if $b = 0, 1$). This result holds more generally under standard choices for the distribution of y_t. Related, another possibly surprising revelation from Figure 1a is weights are not relatively equal across the relevant parts of the shock's support, even though it follows a uniform distribution. In fact, the weight plots look similar when x_t follows a normal distribution. A different instinct some researchers might have

---

[5]Motivated by a basic New Keynesian model. Caravello and Martínez Bruera (2024) use a special case to illustrate their separation result, and I found tinkering with it was very helpful to understand the broader mechanics of the weights.

is to estimate positive and negative effects separately, for instance by truncating Example 1 to only estimate β_1. Figure 1b shows this also poses an inference problem – over 30% of the weight is put on positive values when using a standard normal shock.
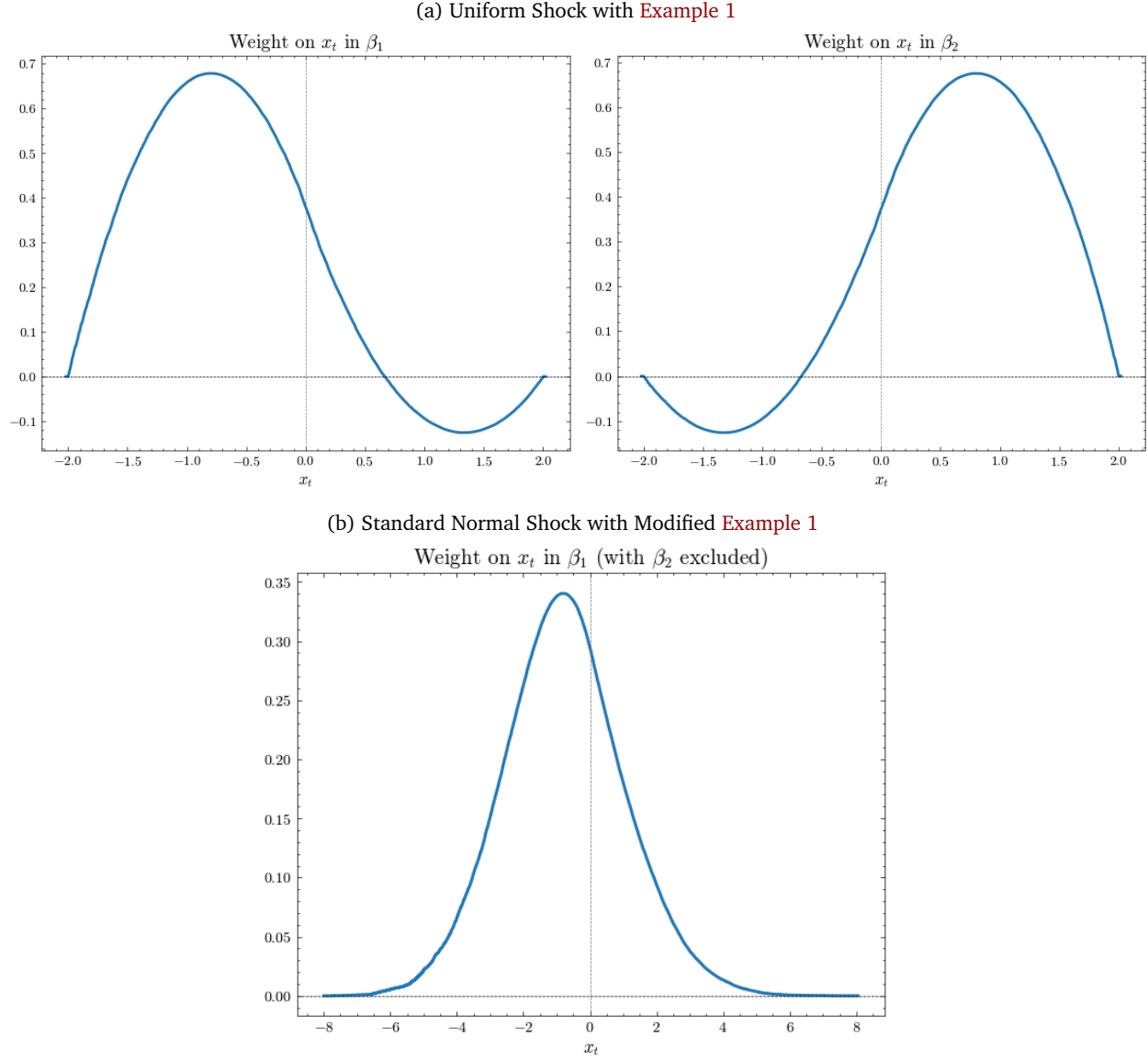
(a) Uniform Shock with Example 1

(b) Standard Normal Shock with Modified Example 1

Figure 1

One common specification to detect asymmetry is to instead use x_t itself in addition to $\max\{x_t, 0\}$

Example 2:
$$\pi_t = \alpha + \beta_1 x_t + \beta_2 \max\{x_t, 0\} + u_t$$

Figure 2 shows this specification can work for a symmetrically distributed shock, but only for a narrow purpose. The right panel shows negative weights, but they serve a purpose: the estimand is the effects of positive shocks relative to negative shocks. So if $\beta_2 \equiv 0$, there are indeed symmetric effects from a positive and negative shock. But the left panel shows β_1 has no clear interpretation; the intuitive description of this specification that "$\beta_1 + \beta_2$

7

is the (weighted) average effect of a positive shock" does not fit. This is one reason why in the Caravello and Martínez Bruera (2024) benchmark for nonlinearity detection for this case is based on

Example 3:
$$\pi_t = \alpha + \beta_1 x_t + \beta_2 f(x_t) + u_t$$

they recommend using an even function to gauge sign effects (e.g., x^2 or $|x|$) and an odd function (e.g., x^3) to test for size. In that case, there will be no negative weight in β_1, while the weights in β_2 will look the same as in Example 2 when testing sign effects. However, when the shock's distribution is not symmetric, we lose the clean separation of size and sign. Figure 3a shows that shocks of a given magnitude need not have the same weight, which means that it's possible positive and negative shocks of the same size have the same (absolute) effect but $\beta_2 \neq 0$, meaning this is not a proper specification for sign effects.
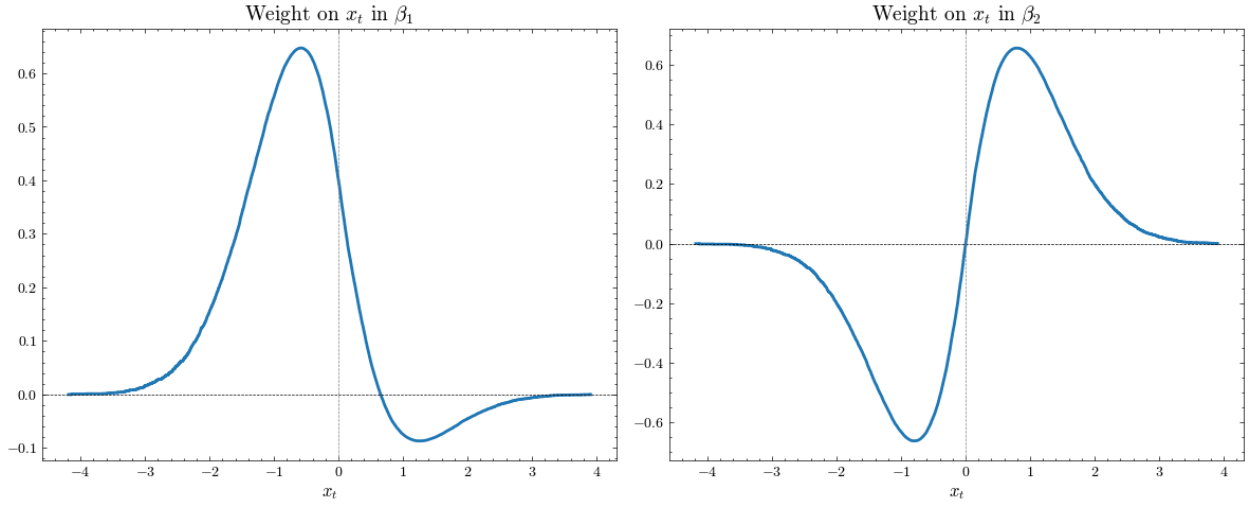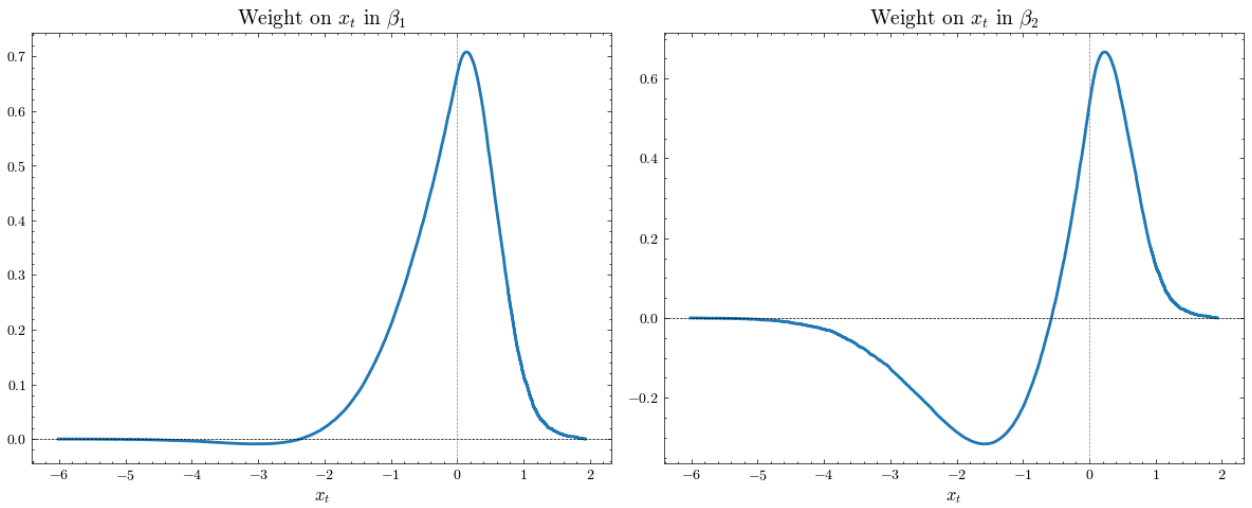

Figure 2: Standard Normal Shock with Example 2

Figure 3a: Left Skewed Shock with Example 3

Even in the symmetric shock case, the Caravello and Martínez Bruera (2024) can have mixed results. For our example DGP, in the $b = 1$ case, there are only sign effects. With uniform shocks, even though the structure of the DGP is simple and the shock distribution is symmetric, the detection performance is poor with a realistic sample

size: in only 16% of 10,000 simulations with $n = 300$, a null hypothesis of no size effects is rejected in a level-.05 test. The performance is better with standard normal shocks, rejecting in 73% of simulations. Similarly, for $b = 2$ and standard normal shocks, there are now size effects for positive shocks, but the null of no size effects is not rejected in 35% of simulations. This highlights the limited power an identification result has in finite samples. To make the failure more transparent, Figure 3b plots the weights in the size effect specification for one of the simulations next to its limit.[6] Even in the most aspirational scenario when shocks follow a well-behaved, symmetric distribution, the weights may be far from converging.[7]
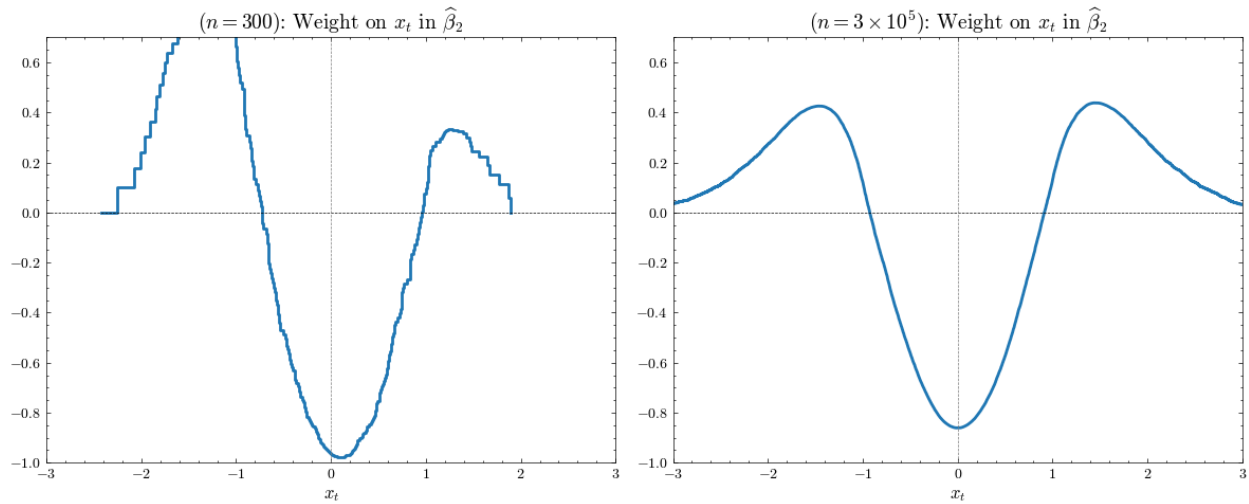

Figure 3b: Standard Normal Shock with Selected Simulation of Example 3

Another issue is interpretability. After conducting hypothesis tests on the Example 3 specification, we are still pretty much in the dark about the underlying DGP. Even if the nulls are properly rejected, we can at best operate under the belief that positive shocks have generally larger effects than negative shocks and (in the $b = 2$ case) big shocks generally have disproportionately larger effects than small shocks, but this is imprecise. At a minimum, we should seek to get more specific than "generally".

To explicitly highlight the themes from this section:

1. **Regression weighting is not obvious ex ante**; researchers should always check the weights for their shock series and specification (Kolesár and Plagborg-Møller, 2025).

2. **Including polynomials not sufficient to gauge nonlinearity** (White, 1980)

3. **It's easy to conflate size and sign effects** (Caravello and Martínez Bruera, 2024)

4. **Traditional approaches are sensitive to sample size and the shock's distribution**

Section 3 builds on these lessons and formalizes a path forward using indicator functions.

---

[6]Here, $f(x) = \mathbb{1}_{x \geq \overline{x}} \cdot (x - \overline{x}) + \mathbb{1}_{x \leq -\overline{x}} \cdot (x + \overline{x})$, where $\overline{x}$ is σ away from the mean (0). Results are similar for $f(x) = x^3$.

[7]The $n = 300$ graph in Figure 3b varies across samples. The median simulated error relative to the sum of the area in each quadrant is 20%. In the language of Caravello and Martínez Bruera (2024), we would say this is problematic because the odd weights are non-trivial.

3   Uncovering Nonlinearities

Formally, we are interested in the effects of a shock ε_t on an outcome Y_{t+h} for different values of ε_t. Section 2 showed if several functions of $\{f_i(\varepsilon_t)\}_{i=1}^N$ are included in the regression, β_i will represent a weighted average of the marginal effects of ε_t on Y_{t+h} on the region where there is non-zero correlation between the shock and the residuals from projecting f_i on everything else in the regression. This illuminates a path towards a more precise characterization of the marginal effect function. If we define disjoint regions $\{I_i\}_{i=1}^N$, we want to have an estimate of weighted marginal effects on each region. At a minimum, the weights must be non-negative and ideally they should not overlap. Caravello and Martínez Bruera (2024) show another desirable property is to make the regions and the weights $\{\omega_i(\cdot)\}_{i=1}^N$ symmetric – if shock realizations $a \in I_i$ and $-a \in I_j$ satisfy $\omega_i(a) = \omega_j(-a)$, then we can compare if the marginal effect function on regions i and j is asymmetric, without results being distorted by the effects of size. Section 3.1 shows certain indicator functions can perform relatively well for these standards and discusses the difficulties in improvements when sample size is limited. Section 3.2 discusses the basic implementation and Section 3.3 outlines broader best practices for working with shock series.

3.1   The Power of Indicators

Disjoint indicator functions feel like they should meet the objectives, but Section 2 revealed two necessary conditions: not interacting the indicators with the shock and including all regions of interest. To see this, we will add a final example with standard normal shocks. One appealing property of using regressions to detect nonlinearities is the implicit weighting is invariant to the outcome variable. So really, we don't need a model to evaluate the weights, just a time series for the shock process. So when we consider

Example 4: $\pi_t = \alpha - \beta_{\text{small, neg}} \cdot \mathbb{1}_{-x_t \in [\frac{1}{100}, \frac{5}{4}]} - \beta_{\text{big, neg}} \cdot \mathbb{1}_{x_t < \frac{5}{4}} + \beta_{\text{small, pos}} \cdot \mathbb{1}_{x_t \in [\frac{1}{100}, \frac{5}{4}]} + \beta_{\text{big, pos}} \cdot \mathbb{1}_{x_t > \frac{5}{4}} + u_t$

note that the weight plots in Figure 4 are the same no matter the left-hand-side outcome variable. The motivation for this form is to set reasonable cutoffs for big and small shock magnitudes (e.g., for standard normal, $x_t = 1$ is a standard deviation and so on). The broader structure seeks to distinguish the effects of both size ($i = \{\text{big, small}\}$) and sign ($j = \{\text{positive, negative}\}$). To test for specific size effects, the null hypothesis is $\beta_{\text{big},j} = \beta_{\text{small},j}$, for sign effects it's $\beta_{i,\text{pos}} = \beta_{i,\text{neg}}$, and for general effects a joint test can be used. In over 99.9% of $n = 300$ simulations of the DGP (6) using the same parameterizations as Example 3, when size or sign effects are present, the appropriate nulls of no effect are rejected. Figure 4 shows weight plots for this specification.[8] The reason for the drastic improvement in performance is evident: the weights here are much further along in converging and also are more directly placing weight where desired.[9] Still, this is not perfect – "big shock" estimates put significant weight on smaller values. But overall it's clearly beneficial to have everything work through a single regression, where each

---

[8]Like Figure 3a, this varies across simulations, but the variance here is concentrated exclusively at the endpoints for the big shock weights.
[9]Appendix A.1 provides more simulation evidence on performance relative to the Caravello and Martínez Bruera (2024) benchmark.

region of interest has its own corresponding estimand. The point estimates themselves are also more revealing, as taking the difference in coefficients provides an indication of how quickly a linear approximation would diverge.
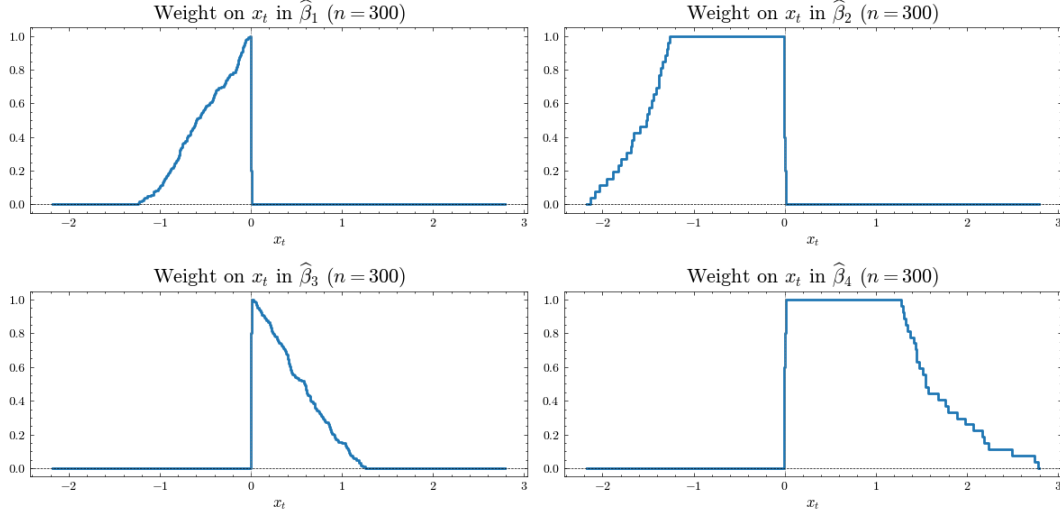

Figure 4: Standard Normal Shock with Example 4

In terms of why these indicator functions work, a formal proof of their properties can be found in Online Appendix B.1 and B.2. Broadly, under an arbitrary shock distribution, all estimands will be the same if the DGP is linear in the shock. In terms of disentangling the effects of size and sign, there is still a complete separation when the shock is symmetrically distributed as in Caravello and Martínez Bruera (2024). For the non-symmetric case, if there are no size effects, estimands on regions of the same sign will be the same, meaning size effects can also be cleanly detected.[10] Online Appendix B.3 shows that without a symmetric distribution, it is generally not possible for *any* method in finite samples to distinguish between a true sign effect and positive and negative shocks having the same kind of size effect. This procedure does provide a simple robustness check to guard against this edge case: defining the small regions so that the weights on $\beta_{\text{small, neg}}, \beta_{\text{small, pos}}$ are as similar as possible.[11] Besides consulting Online Appendix B for the explicit econometric backing, the next subsection details the implementation and tries to weave in some of the intuition for the identification proof.

3.2 Implementation

Here are the broad steps to implement this procedure

1. Define regions $\{I_i\}_{i=1}^N$

2. Exclude 0

3. Define f_i to be $\text{sign}(x) \cdot \mathbb{1}_{x \in I_i}$

4. Rescale f_i by coefficient in projection of the shock on $\{f_i\}_{i=1}^N$

We will go through each step and conclude with demonstrating what can go wrong when deviating from them.

---

[10]Note that this is not the case for the Caravello and Martínez Bruera (2024) approach.

[11]"Robustness check" refers to how the point estimates change. Since the definitions are conditional, inference is more complicated.

1. Define regions

This procedure can be extended to an arbitrary number of regions, but because the number of large shocks will typically be limited, the previously discussed baseline of 4 is recommended. Needing to set a threshold for what a "big" shock is not ideal. One thing that helps is first transforming the shock so that the set of non-zero observations has standard deviation 1 and the entire series is mean 0. Thresholds in terms of standard deviations are simpler to reason about. A good rule of thumb is to ex ante commit to one threshold and then to see how results change ex post after it's mildly perturbed in both directions.

2. Exclude 0

Including indicator functions on all possible values of a shock will result in perfectly collinear regressors. So for the regions on either side of the real line, zero should be excluded. Shock series with lots of zero values can run into problems for small samples (Barnichon and Mesters, 2025), but here there's a benefit of being able easily circumvent linear dependence. It's good practice to exclude a small interval centered around zero as well, especially if there aren't many zero values. A natural question is why can't we instead exclude some other region. Considering the DGP from the last section, one might propose using

Example 5:
$$\pi_t = \alpha - \beta_1 \cdot \mathbb{1}_{-x_t \in (0,1]} - \beta_2 \cdot \mathbb{1}_{-x_t < 1} - \beta_3 \cdot \mathbb{1}_{x_t \in [0,1]} + u_t$$

With respect to the use of dummy variables in microeconometrics, one might expect that estimands represent effects relative to the excluded region. Figure 5 shows while the estimands do have a clear interpretation, in some cases it initially seems at odds with the Example 5 specification. For each estimand β_i, say the corresponding indicator is active on interval I_i. Figure 5 shows the estimands are weighted averages of marginal effects on the interval from the left endpoint of I_i to ∞.[12] So these objects may have some use; under a linear DGP, these estimands will all be the same. But the weights are, with respect to the indicators they relate to, poorly targeted.
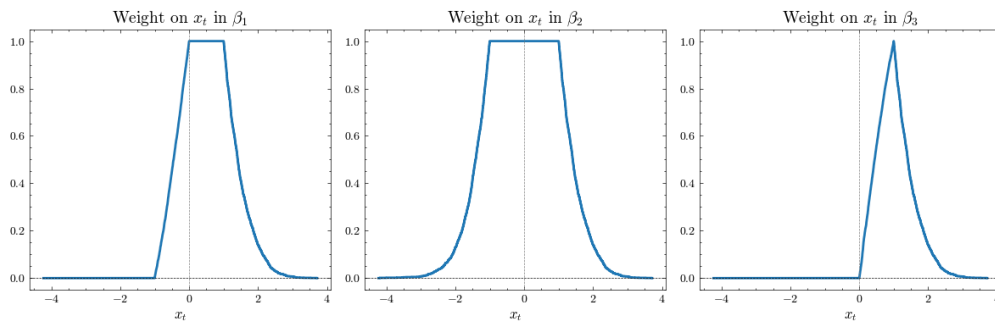

Figure 5: Standard Normal Shock with Example 5

3. Define $f_i(x) = \text{sign}(x) \cdot \mathbb{1}_{x \in I_i}$

f_i should be an indicator of whether the shock is in a given region multiplied by the sign of the shock.[13]

---

[12]Because weights are negligible where ε_t has low density, weights are essentially 0 starting from ≈ 3 standard deviations.

[13]Multiplying by the sign function is not critical but is rather for ease of interpretation. Without this transformation, the sign of the estimands on negative regions would be the opposite of average marginal effects.

4. Rescale f_i

Notice in Figure 4 the weights do not all integrate to the same value, meaning comparing estimates could be distorted by a scaling issue. If we rescale the coefficients by the projection of the shock on all the functions, all the weights will integrate to 1 by construction, meaning they are both comparable and interpretable as a traditional weighted average. Because the functions are now generated regressors, standard errors require a delta method adjustment, but the adjustment turns out to be small in practice (see Appendix A.3 as well as the code on this paper's GitHub repository).[14]

Where deviating from these steps can go wrong

Including the shock itself or interacting the shock with the indicator functions can lead to unexpected results. To fix ideas, we again use DGP (6) with standard normal shocks and consider

Example 6: $$\pi_t = \alpha - \beta_1 \cdot \mathbb{1}_{-x_t \in (0,1]} - \beta_2 \cdot \mathbb{1}_{-x_t < 1} + \beta_3 \cdot \mathbb{1}_{x_t \in (0,1]} + \beta_4 \cdot \mathbb{1}_{x_t \in [0,1]} + \beta_5 x_t + u_t$$

Figure 6 shows the resulting weights behave erratically and none of the estimands have any meaningful interpretation. The weights can also behave counterintuitively if the shock is interacted with an indicator function. This is why for any specification used, it's always best to first simulate the weights under the simplest, standard normal case and ensure the description of the estimation is justified. Once more, Figure 6 also highlights the disconnect between our instinct when looking at regression coefficients as representing the effect of a unit change and what the estimands actually imply. All estimands on functions of $f(\varepsilon_t)$ are weighted averages of the same object. The only difference is how the weighting occurs, and it's often not what we would expect. Online Appendix B sheds some more light on the properties of the linear regression weights that result in this behavior.
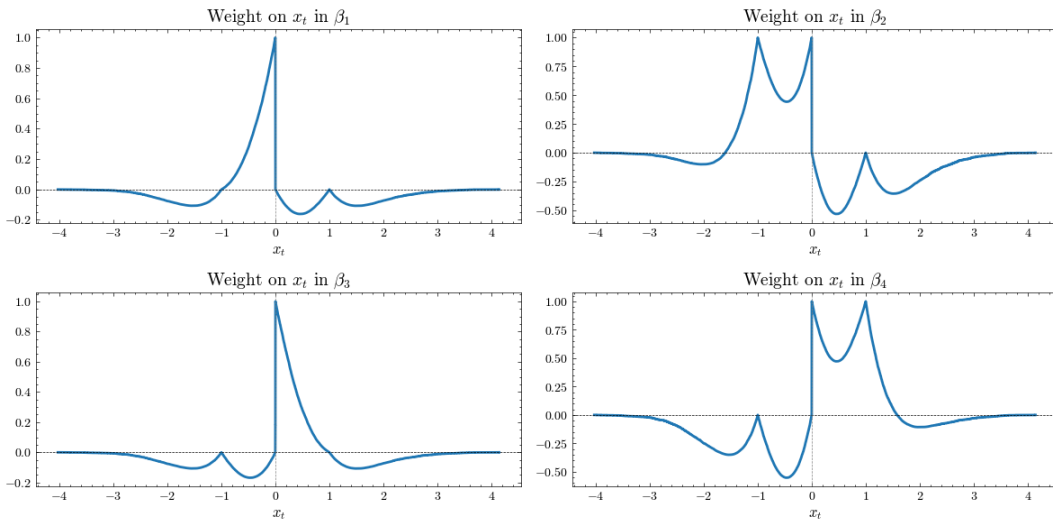

Figure 6: Standard Normal Shock with Example 6

---

[14]The repository is located at https://github.com/paulbousquet/UncoveringNonlin

3.3 General Best Practices

Just as slight perturbations in the local projection specification can yield vastly different results, it's also important to be cautious about identifying assumptions about a given shock series. The results in Section 2 rely on using the actual structural shock ε_t in the regression. In practice, we will have some shock proxy z_t. Kolesár and Plagborg-Møller (2025) note that if ε_t is unobserved, the form of the weights is unknown, so plotting the weights under the assumption that $\varepsilon_t = z_t$ provides the "best case scenario". Ideally, the proxy should indeed be a (classical) proxy – departures from the structural shock amount to noisy measurement. But a measurement error process can be complicated, even if induced from noise alone. Chen et al. (2011) show that the usual "attenuation bias" relationship to estimands does not hold if the measurement error is non-linear. Appendix A.1 explores consequences of classical measurement error and thankfully shows under a rich set of measurement error types (e.g., nonlinearities, heteroskedasticity, state-dependent noise) the "best case" weights are good approximations for the true weights. Even though the bias can go in either direction, that does not matter for the hypothesis testing procedure.

The more important issue is that our shock proxy may not be "clean". For example, we may only be willing to argue a shock series satisfies independence conditional on some control set. Kolesár and Plagborg-Møller (2025) show in this case we lose the powerful identification results because there will generally be negative weights. If the proxy is tied to more than one structural shock (and not accounted for), Koo et al. (2024) show inference will be incorrect.[15] A complication of the guidelines for assessing proxy quality a lack of an adequate sensitivity analysis procedure. There's a natural concern that certain control variables can drive results if a proxy is not a pure shock, and to this effect Kolesár and Plagborg-Møller (2025) recommend dropping various controls and seeing if results change, a ubiquitous robustness check strategy. Unfortunately, these results may be misleading. In the simulations for Appendix A.1, I find even when the outcome DGP is linear in the shock, (small sample) point estimates vary considerably when different control sets are used. So we cannot distinguish between variation indicating sensitivity to controls or sensitivity to sample size. This raises a broader point: another common robustness check is to redo the main analysis with different shock proxies. But because these proxies are constructed so differently (see, e.g., Brennan et al. (2025); Adams and Barrett (2025) for comparisons of monetary policy shock series), in general there's no reason the implied weights and therefore results should be similar (McKay and Wolf, 2023). Until better sensitivity analysis tests are available, the best strategy is to have a convincing argument $\varepsilon_t \approx z_t$.

Once a shock series has been selected, there are other considerations for estimation. A great deal of work has been done to clarify the differences between LP and vector autoregression (VAR). (Plagborg-Møller and Wolf, 2021) prove LP and VAR are asymptotically equivalent in the limit (if lag order is high enough). So even researchers who prefer VAR estimation should run the LP analogue and plot the weights to have a better sense of what is being estimated. While there will be finite sample differences (Li et al., 2024; Montiel Olea et al., 2024),

---

[15]For proxies with many zeros, finite sample correlation is inherent, see Barnichon and Mesters (2025) for discussion and a proposed fix.

plotting the weights in the LP will still be informative. Borrowing from Kolesár and Plagborg-Møller (2025)'s example, if all the weights on a government spending shock are being placed on positive values, what estimation is actually uncovering are the effects of spending buildups. They show this phenomenon largely holds for several popular shock series, so it's important, no matter the preference for LP vs. VAR, to plot the weights. Overall, Montiel Olea et al. (2025) provides a survey on considerations for LP and VAR estimation and gives an overall recommendation of LPs for finite sample flexibility. For the procedure outlined in this paper, there are also separate interpretation issues, which are discussed in context of the application in Section 4.2. Online Appendix B also discusses generalizations, constructing functions using the empirical CDF or machine learning, which I find have more limited use with typical sample sizes in macroeconomics.

4 Application

Section 4.1 applies the guidance in Section 3.3 for choosing a monetary shock proxy. Because many great papers and people have been committed to the topic, the discussion is specialized and may not be of interest to general readers. Likewise, interested readers may be disappointed that some of the arguments are not fully fleshed out, which I leave to separate work. Section 4.2 shows the nonlinearity detection results and briefly outlines an attempt to match them with a non-linear equilibrium model.

4.1 Selecting a Monetary Shock Series

To select a series for assessing possible nonlinearities in the transmission of U.S. monetary policy, we have to address a question for which there is surprisingly not a straightforward answer: what is a (structural) monetary policy shock? Unless one is willing to argue that central banks have a systematic way to set rates they decide arbitrarily to deviate from, which seems like a poor description of an institution like the Federal Reserve and its army of economists, monetary shocks are changes in policy unanticipated by private agents. This makes the high-frequency measures of forecast errors backed out from price changes in futures markets a natural choice.

Within the class of high-frequency measures, there are several options. Bu et al. (2021) is currently popular because of its ability to easily handle the zero lower bound period by creating a single measure to represent shocks across the entire yield curve. At the same time, their measure cannot be easily mapped into a candidate data generating process, so it's less clear what would be estimated (Brennan et al., 2025). Another issue is that because private agents do not know perfectly the central bank's reaction function and there may not be a single information set for all agents, changes in futures markets may be representing combinations of multiple structural shocks, which is a challenge. There are several measures that look at changes to expected future interest rates, rather than the current period, and try to decompose them into "forward guidance" vs. "information shocks" (e.g., Jarociński and Karadi, 2020), but because these measures are estimation-specific, there is a risk that the deviation from

structural shocks is systematic or sample-dependent, rather than pure noise. Instead, sacrificing performance at the zero lower bound and looking at changes to the Fed's expected change to its target in the current period seems to be the most practical option. All concerns about the possible tangling of effects from forward guidance, information, credibility, preferences, etc. are moot when looking at the current period because once an action is announced, the adjustment is not a function of ambiguity about any of those things because the Fed chair has essentially written the futures price correction in stone.[16] This leads to a selection of the Jarociński (2024) MP1 series, originally developed by Kuttner (2001), as the proxy of choice.

Before moving on to discussing other approaches more in-depth, it should be noted there are many concerns specific to the high-frequency series. When outcome variables are not also high-frequency, Jacobson et al. (2024) warn of temporal aggregation bias because the Federal Reserve's meeting calendar fluctuates and sometimes multiple shocks occur within the same month. Absent getting better data, the best response is likely to not put much stock in the results at the shortest horizons. Casini and McCloskey (2024) also point out that using a narrow observation is not actually a magic identification wand, though they show the Nakamura and Steinsson (2018) measure is relatively robust to the potential concerns. A final consideration is that these futures markets are not fully saturated with participants, particularly during the zero lower bound period, and past work has shown many popular high-frequency series are correlated with macroeconomic fundamentals (Miranda-Agrippino and Ricco, 2021; Bauer and Swanson, 2023). This predictability concern has rightly been a focal point of the recent literature (Acosta, 2023), but the results may not be as damning as they seem. Leaving aside that these markets may be innately "inefficient", it seems more likely that these finite sample results are showing the effects of heteroskedasticity. When there is more movement in macro fundamentals, it is more likely for central banks to act, thus creating more variance for structural shocks. Basic heteroskedasticity itself does not significantly disturb the utility of proxies (Section 3) but more work is needed to clarify the relationship between (endogenous) second-moment dependence and "selection bias" (Rambachan and Shephard, 2025).

Another popular method in this literature is projection orthogonalization, or using the residuals from a linear regression. This is the basis for Romer and Romer (2004), who represent the change in interest rates unrelated to the Fed's information with the residuals in a regression of changes in the federal funds rate on Fed forecasts.[17] But the residuals themselves are extremely sensitive to the estimated coefficients, and we should not have faith that $\varepsilon_t \approx z_t$ – Cochrane (2011) demonstrates this won't occur even in the simplest case where the data generating process is linear (a basic New Keynesian model with a Taylor Rule). Miranda-Agrippino and Ricco (2021) and Bauer and Swanson (2023) use orthogonalization by residualizing existing measures of monetary policy shocks to guard against claims of predictability (see Acosta (2023) for a survey). These adjustments will likewise be sensitive to the realized OLS point estimates, which really have bite given the sample size. For instance, Bauer and Swanson

---

[16]There is risk of contamination in the few instances where there were shocks in the days before the formal announcement of the target.
[17]Aruoba and Drechsel (2025) argue these forecasts don't span the information set. They extend the methodology with text analysis.

(2023)'s shocks are based on a 1988-2023 monthly sample. If they had originally done this procedure in 2015, the median percent difference in shock magnitude between the original and "updated" series would be over 100%.[18]

4.2 Nonlinearities in the Effects of Monetary Policy Shocks

I look for evidence of nonlinearities in monetary policy transmission by applying the described procedure to outcome variables of industrial production, consumer price index (CPI), consumption, and unemployment from November 1988 to January 2020 using the MP1 series.[19] I use log differences and cumulate them over future horizons so that the left-hand side variables represent "percent change since the shock occurred". Before detailing the inference procedure, it's useful to first establish how to interpret point estimates. Recall within this paradigm, we define size and sign effects in terms of coefficient differences. For size effects, the big and small estimates will be the same (in population) if and only if marginal effects on the region where there is non-overlapping weight are not constant. For example, a negative sign effect for big shocks and industrial production can be interpreted as the expansionary effect for big negative shocks (on IP) is smaller than the contractionary effect from big positive shocks. As discussed in Section 3.1, sign effects are trickier because of the difficulties in falsifying the edge case where there are symmetric size effects (Caravello and Martínez Bruera, 2024). Ideally, sign effect tests (H_0: $\beta_{\mathrm{pos}} - \beta_{\mathrm{neg}} = 0$) can provide an indication for or against a narrative similar to "pushing on a string" (Fisher, 1935; Tenreyro and Thwaites, 2016); $\beta_{\mathrm{pos}} < \beta_{\mathrm{neg}}$ for output proxies indicates positive shocks have a disproportionately larger contractionary effect relative to the expansionary effect of negative shocks.

Figure 7 shows the point estimates using penalized local projection (Barnichon and Brownlees, 2019) indicate size effects for positive shocks and sign effects for big shocks. For all variables with the exception of CPI, larger positive shocks have a disproportionately more contractionary effect. The sign effect plots broadly support a "pushing on a string" story (for unemployment, the signs flip: $\beta_{\mathrm{pos}} > \beta_{\mathrm{neg}}$ means larger contractionary effects).

The more relevant question is whether the point estimates are meaningful indications of persistent nonlinearities. For inference, Barnichon and Brownlees (2019) note that a downside to penalization methods is generally complicated asymptotics, especially when using cross-validation. To over-compensate for any potential bias, I fix the penalty parameter at a mild level ex ante and use 99% confidence intervals with a Huber-White variance-covariance matrix from an even more under-smoothed estimate (Montiel Olea and Plagborg-Møller, 2021). Also, it's tempting to think about Figure 7 through the lens of the difference between an (average) impulse to two different types of shocks. This interpretation more so holds for sign effects, but the weighting need not be the same for big and small shocks for the same size category (Caravello and Martínez Bruera, 2024). For a robustness check, I find that results do not differ when the big shock regions are manipulated to have more similar weighting. For size effects, there is weight overlap for shock regions of the same size. So we can instead interpret those plots as

---

[18]Sims (1998) cautions against scrutinizing shock magnitudes in VARs, which are relative to a given information set. The concern here is distinct. Again, $\varepsilon_t \approx z_t$ means the bias should be from systematic measurement noise (so shouldn't be mechanically sample-dependent).

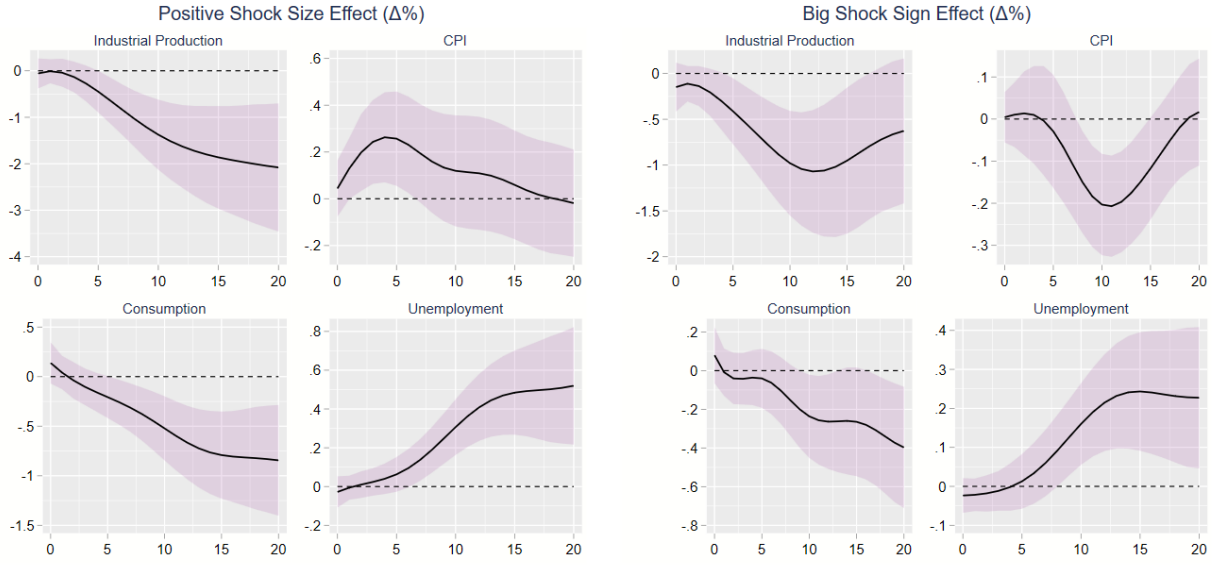[19]More details can be found in Online Appendix D.4.

Figure 7: Indicator Approach with MP1 Shocks

how fast a linear approximation diverges. Appendix A.5 features the remaining visualizations. In particular, there is evidence big negative shocks have a more expansionary effect in the long run than small shocks (Figure 11). For sign effects, not much can be said about asymmetries for small shocks (Figure 12).

The nonlinearities are therefore statistically meaningful, but there's a point to be made about whether they are *economically* meaningful. For example: suppose marginal effects for positive shocks are β and for negative shocks $\beta + \varepsilon$. A population hypothesis test will reject a null hypothesis of linearity, even though a linear model is appropriate. Because the indicator functions are normalized so that their individual weights integrate to 1, coefficient differences can give some insight into whether the degree of nonlinearity matters because they have a reasonable interpretation as a difference in means. For the application in this section, this translates to measuring the nonlinearity in terms of difference in percentage change in outcome since the shock occurred. Referencing Figure 7 confirms the economically significant effects.[20]

Overall, a picture is painted that is hard to square with standard models: nonlinearities that peak in the medium to long run. The next step after finding results like this is to try and explain them. To compare to the results from US data, a basic point of reference would be using a model that features meaningful nonlinearities to generate data and then run the same regressions. Barnichon and Matthes (2018) conjecture that sign effects that work in opposite directions for unemployment and inflation, which is what we observed in the last section, can be rationalized in a model with downward-rigid prices and wages (Kim and Ruge-Murcia, 2009). In this setting,

---

[20]To avoid haggling over what constitutes meaningful nonlinearity, one option is to normalize by the linear estimate's standard deviations, so coefficient differences are still in units of effect sizes but in some sense have an interpretation similar to t-statistics (i.e., gesturing towards the likelihood parameters were drawn from the same distribution). Results are also more easily comparable to DSGE model output by minimizing unimportant scaling distortion from finite sample properties of time series and model-simulated data. Details are in Online Appendix D.6.

firms seeking to change their price at a rate different from steady-state inflation face an adjustment cost

$$\Phi_t^p(\pi_t) = \frac{\phi_p}{\psi_p^2} \left(e^{-\psi_p(\pi_t - \pi^*)} + \psi_p(\pi_t - \pi^*) - 1 \right)$$

For $\psi_p > 0$, it's more costly to decrease prices than raise them (downward-rigid), for $\psi_p < 0$ prices are upward-rigid, and the function limits to symmetric adjustment costs as $\psi_p \to 0$. Nominal wage adjustment costs take on the same structure. Past estimation of this model has found evidence of downward rigidity in prices and wages, consistent with empirical evidence dating back to Keynes (1936) and Tobin (1972).

Since the relevance for this paper is largely motivation, I relegate most details about the model and the estimation to Online Appendix D.5. Using the same sample period of US data, the Aruoba et al. (2017) extension of the downward-rigidity model is estimated to second order via a standard random walk Metropolis-Hastings algorithm and particle filter (Fernandez-Villaverde and Rubio-Ramirez, 2007). I use the distribution of parameters generated by this exercise to simulate data and run the same local projections procedure to create Bayesian analogue (i.e., using credible sets instead of confidence intervals) for the empirical results. These exercises show (full results in Online Appendix D.5) that while the model can generate nonlinearities, in general the observed asymmetric effects for both size and sign occur on impact and then quickly dissipate. I also take the posterior mode of all parameters and then vary both asymmetry parameters (one at a time, in both directions, and then both at once in the same direction) while keeping everything else fixed, then simulate data and estimate for each combination. This exercise provides some clarity: on impact, certain combinations of the asymmetry parameters can generate any desired nonlinearities, but it cannot be sustained.

Looking at the impulse response functions directly from the model (rather than running a LP) corroborates the above interpretations. Figure 8 shows impulse responses for both negative and positive shocks of different sizes. By a horizon of 5 periods after the shock, the magnitude of responses is near or below zero. Online Appendix D.5 discusses various extensions to the model, like adding autocorrelated shocks, that ultimately don't help much.

One reason why the effects of monetary shocks may not have a lasting effect is the lack of inertia in interest rate setting. Even though the Metropolis-Hastings produced draws with moderately high persistence in the Taylor Rule (posterior mode of $\rho_r \approx .67$), an inspection of model-simulated data reveals that whenever a large monetary shock takes the central bank away from its (nominal) target i^*, it generally doesn't take long to get back. Table 1 in Online Appendix D.5 shows the results of 10,000 simulations at the posterior mode. For each simulation, I take the median distance between the target interest rate and the current interest rate h periods after a big change in interest rates (magnitude greater than 10%) and then average across simulations. In periods in which the central bank heavily adjusts the interest rate, the target is relatively far away, but this is almost completely undone 2 periods later. There is also a large asymmetry on impact that quickly becomes less dramatic. Regardless of model, consecutive, large realizations of white noise innovations are unlikely, but the staying power of shocks can vary. These results suggest that the nonlinearities observed in data may not have an explanation in our standard class
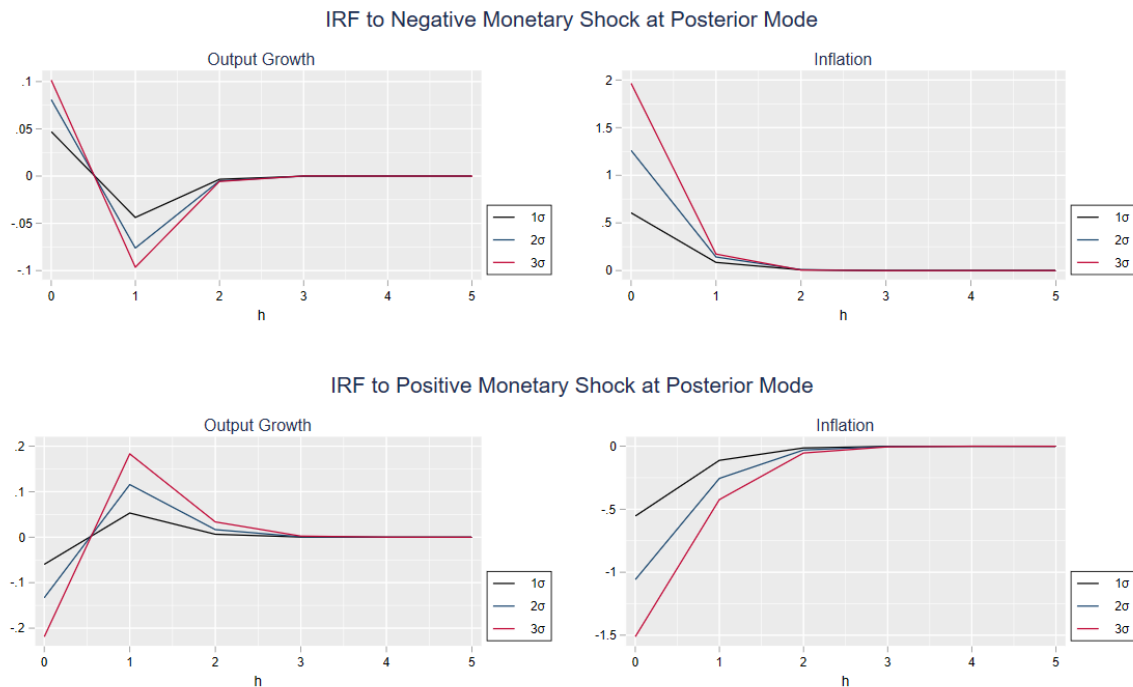
Figure 8

of models and warrant further exploration of channels in monetary policy transmission. There should also be some broader considerations added in model selection. Linearized general equilibrium models, appealing because of a reduction of analytical and computational complexity, can output suboptimal normative prescriptions if the economy actually follows a data generating process with strong non-linear components.

5 Conclusion

This paper demonstrates a new method to test for nonlinearities in data exploiting properties of least squares regression that are consequences of assumptions about proxies for structural shocks that are commonly made in the applied macroeconomics literature. Three new approaches within this framework were characterized, but the simplest (disjoint indicator functions) seem to be the most useful in practice. While this seems to be yet another example of the power of OLS in spite of its simplicity, there are some limitations setting the table for future work. There is a tension that emerges between making the weights appear in the desired places and the efficiency of estimates. The disjoint indicator functions are the most efficient option at the expense of having relatively dispersed weight. So while we can view coefficient differences as a good gauge of deviations from linearity, we cannot interpret the estimands themselves as weighted averages of marginal effects on the areas the indicator functions are active (they are weighted averages over a larger region). It seems possible to expand along this dimension, but it's not immediately clear how. The procedure informed an application to monetary policy shocks, which showed results that are difficult to match even with a general equilibrium model that featured rich size and sign nonlinearities. Results like this can inform paths forward for better understanding the transmission of policy.

References

Acosta, Miguel. 2023. "The Perceived Causes of Monetary Policy Surprises." February, Working paper.

Adams, Jonathan J., and Philip Barrett. 2025. "What Are Empirical Monetary Policy Shocks? Estimating the Term Structure of Policy News." June, Working Paper.

Alessandri, Piergiorgio, Òscar Jordà, and Fabrizio Venditti. 2025. "Decomposing the monetary policy multiplier." *Journal of Monetary Economics*, 152, p. 103783.

Aruoba, S. Boragan, and Thomas Drechsel. 2025. "Identifying Monetary Policy Shocks: A Natural Language Approach." *American Economic Journal: Macroeconomics*, forthcoming.

Aruoba, S. Borağan, Luigi Bocola, and Frank Schorfheide. 2017. "Assessing DSGE model nonlinearities." *Journal of Economic Dynamics and Control*, 83 34–54.

Ascari, Guido, and Timo Haber. 2022. "Non-Linearities, State-Dependent Prices and the Transmission Mechanism of Monetary Policy." *The Economic Journal*, 132(641): 441–478.

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

Barnichon, Regis, and Christian Matthes. 2018. "Functional Approximation of Impulse Responses." *Journal of Monetary Economics*, 99 41–55.

Barnichon, Régis, and Geert Mesters. 2025. "Innovations meet Narratives -improving the power-credibility trade-off in macro." January, Working paper.

Bauer, Michael D., and Eric T. Swanson. 2023. "An Alternative Explanation for the "Fed Information Effect"." *American Economic Review*, 113(3): 664–700.

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

Brennan, Connor M., Margaret M. Jacobson, Christian Matthes, and Todd B. Walker. 2025. "Monetary Policy Shocks: Data or Methods?" finance and economics discussion series, Federal Reserve Board, Washington, D.C.

Bu, Chunya, John Rogers, and Wenbin Wu. 2021. "A unified measure of Fed monetary policy shocks." *Journal of Monetary Economics*, 118 331–349.

Caravello, Tomás E., and Pedro Martínez Bruera. 2024. "Disentangling Sign and Size Non-linearities." June.

Casini, Alessandro, and Adam McCloskey. 2024. "Identification and Estimation of Causal Effects in High-Frequency Event Studies." June, Working paper.

Chen, Xiaohong, Han Hong, and Denis Nekipelov. 2011. "Nonlinear Models of Measurement Errors." *Journal of Economic Literature*, 49(4): 901–937.

Cochrane, John. 2011. "Determinacy and Identification with Taylor Rules." *Journal of Political Economy*, 119(3): 565–615.

Cochrane, John. 2024. "Expectations and the neutrality of interest rates." *Review of Economic Dynamics*, 53 194–223.

Fernandez-Villaverde, Jesus, and Juan F. Rubio-Ramirez. 2007. "Estimating Macroeconomic Models: A Likelihood Approach." *Review of Economic Studies*, 74(4): 1059–1087.

Fisher, Irving. 1935. *100% Money*.: Adelphi Company, , 1st edition 94.

Friedman, Milton. 1960. *A Program for Monetary Stability*.: Fordham University Press.

Gonçalves, Sílvia, Ana María Herrera, Lutz Kilian, and Elena Pesavento. 2024. "State-dependent local projections."

Imbens, Guido W., and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica*, 62(2): 467–475.

Jacobson, Margaret M., Christian Matthes, and Todd B. Walker. 2024. "Temporal Aggregation Bias and Monetary Policy Transmission." March, Working Paper.

Jarociński, Marek. 2024. "Estimating the Fed's unconventional policy shocks." *Journal of Monetary Economics*, 144, p. 103548.

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.

Keynes, John Maynard. 1936. *The General Theory of Employment, Interest and Money*. Cambridge: Macmillan Cambridge University Press.

Kim, Jinill, and Francisco J. Ruge-Murcia. 2009. "How much inflation is necessary to grease the wheels?" *Journal of Monetary Economics*, 56 365–377.

Kolesár, Michal, and Mikkel Plagborg-Møller. 2025. "Dynamic Causal Effects in a Nonlinear World: the Good, the Bad, and the Ugly." Working Paper.

Koo, Bonsoo, Seojeong Lee, and Myung Hwan Seo. 2024. "What Impulse Response Do Instrumental Variables Identify?" April, Working paper.

Kuttner, Kenneth N. 2001. "Monetary Policy Surprises and Interest Rates: Evidence from the Fed Funds Futures Market." *Journal of Monetary Economics*, 47(3): 523–544.

Li, Dake, Mikkel Plagborg-Møller, and Christian K. Wolf. 2024. "Local Projections vs. VARs: Lessons From Thousands of DGPs."

Loria, Francesca, Christian Matthes, and Donghai Zhang. 2025. "Assessing Macroeconomic Tail Risk." *The Economic Journal*, 135(665): 264–284.

McKay, Alisdair, and Christian K. Wolf. 2023. "What Can Time-Series Regressions Tell Us About Policy Counterfactuals?" *Econometrica*, 91(5): 1695–1725.

Miranda-Agrippino, Silvia, and Giovanni Ricco. 2021. "The Transmission of Monetary Policy Shocks." *American Economic Journal: Macroeconomics*, 13(3): .

Montiel Olea, José Luis, and Mikkel Plagborg-Møller. 2021. "Local Projection Inference is Simpler and More Robust Than You Think." *Econometrica*, 89(4): 1789–1823.

Montiel Olea, José Luis, Mikkel Plagborg-Møller, Eric Qian, and Christian K. Wolf. 2024. "Double Robustness of Local Projections and Some Unpleasant VARithmetic."

Montiel Olea, José Luis, Mikkel Plagborg-Møller, Eric Qian, and Christian K. Wolf. 2025. "Local Projections or VARs? A Primer for Macroeconomists." In *NBER Macroeconomics Annual 2025*.

Nakamura, Emi, and Jón Steinsson. 2018. "High-Frequency Identification of Monetary Non-Neutrality: The Information Effect." *The Quarterly Journal of Economics*, 133(3): 1283–1330.

Plagborg-Møller, Mikkel, and Christian K. Wolf. 2021. "Local Projections and VARs Estimate the Same Impulse Responses." *Econometrica*, 89(2): 955–980.

Rambachan, Ashesh, and Neil Shephard. 2025. "When do common time series estimands have nonparametric causal meaning?", Working paper.

Ramey, Valerie. 2016. "Macroeconomic Shocks and Their Propagation." In *Handbook of Macroeconomics*. Chap. 2.

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.

Romer, Christina D., and David H. Romer. 2004. "A New Measure of Monetary Shocks: Derivation and Implications." *American Economic Review*, 94(4): 1055–1084.

Sims, Christopher A. 1998. "Comment on Glenn Rudebusch's "Do Measures of Monetary Policy in a VAR Make Sense?"." *International Economic Review*, 39(4): 933–941.

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

Tenreyro, Silvana, and Gregory Thwaites. 2016. "Pushing on a String: US Monetary Policy Is Less Powerful in Recessions." *American Economic Journal: Macroeconomics*, 8(4): 43–74.

Tobin, James. 1972. "Inflation and unemployment." *American Economic Review*, 62 1–18.

White, Halbert. 1980. "Using Least Squares to Approximate Unknown Regression Functions." *International Economic Review*, 21(1): 149–170.

Xu, Ke-Li. 2023. "Local Projection Based Inference under General Conditions." Working Paper.

Appendix[21]

A.1 Simulation Performance

This section will assess performance across a variety of data generating processes. The DGP in (6) from Example 3 and Example 4 gives a benchmark for the prevailing takeaway from this section. For the case of $b = 2$ with standard normal shocks (and sample size $n = 300$), a deep learning approach only offers a modest improvement over the Caravello and Martínez Bruera (2024) approach, failing to reject a null hypothesis of no size effects in 26% of simulations. The generated regressor approach fares even worse, correctly rejecting in only 60% of simulations – the methods developed to produce fewer false negatives failed to do so. This is because, while the weights look more appealing (see Online Appendix B.6), the variance of the coefficients limits the usefulness of this property. Online Appendix B discusses two alternatives. For the generated regressor approach, the functions corresponding to big shocks in particular have large variance, in part because less of the sample is concentrated there. For deep learning, the idiosyncrasies the neural network creates to respect the constraint set (see Online Appendix B.4) also create more variance. And to re-emphasize – these occur even with DGP (6), which outside of a single kink, is about as vanilla as it gets (outcome driven entirely by two i.i.d shocks with no autocorrelation). This is an indication that disjoint indicator functions should be the default method of choice, though it's still useful to report results from all.

One last, previously omitted case to discuss before moving on from DGP (6) is $b = .5$ (square-root). While the indicator functions perform the best in this environment, nulls are not rejected in nearly 40% of cases, with the other methods performing far worse. This presents a stark limitation – an important nonlinearity of economic interest would be diminishing returns to scale of policy intervention. However, these are intrinsically harder to detect, as the second derivative of the square-root function is essentially constant after moving away from 0. This is an unfortunate downside the framework is not as well-equipped to handle.

To get a more holistic picture of performance, we now consider a richer class of data generating processes. In particular, Li et al. (2024) assess the finite sample tradeoffs between local projections and vector autoregressions by making thousands of random selections of 5 variables from the Stock and Watson (2018) dataset and fitting a dynamic factor model to create a DGP (and back out the structural shocks). I use this as a starting point and compare all the above methods against a multitude of flavors for DGP. Because each DGP has its own structural shock, it's not feasible to train a neural network on each one. I find the point estimates are very similar to the disjoint indicator approach, which is not surprising given the functional forms it consistently converges to are broadly well-approximated by combinations of indicator functions and linear functions.

I modify a threshold-VAR model from Loria et al. (2025) who argue it captures some fundamental macroeconomic dynamics. The structure is centered around 3 components: growth of real activity y_t, a financial factor

---

[21]The Online Appendix can be found in the paper's GitHub repository `https://github.com/paulbousquet/UncoveringNonlin`

25

f_t, and a macroeconomic factor m_t. I keep this same system of 3 equations but add inflation π_t and additional fundamentals W_t. I use the Stock and Watson (2018) dataset and randomly select a series from the relevant group of variables for $y_t, f_t, m_t,$ and π_t and randomly select other variables from the remaining categories for W_t. The procedure detailed in Li et al. (2024) is used to generate a structural monetary policy shock X_t for this system using a dynamic factor model representation. The skeleton of the threshold-VAR DGP is

$$y_t = \beta_0 + \beta_1 f_t + \beta_2 m_t + \beta_3 \pi_t + \beta_w' W_{t-1} + u_t^y$$
$$f_t = \alpha_0 + \alpha_1 f_{t-1} + \alpha_2 m_t + \alpha_3 g(x_t) \cdot \mathbb{1}_x(f_{t-1}, m_{t-1}) + u_t^f \quad (7)$$
$$m_t = \gamma_0 + \gamma_1 m_{t-1} + \gamma_2 f_{t-1} + \gamma_3 g(x_t) \cdot \mathbb{1}_x(f_{t-1}, m_{t-1}) + u_t^m$$

where $g(x_t)$ is some non-linear function of the shock and $\mathbb{1}_x(f_{t-1}, m_{t-1})$ is a state-dependent multiplier. In the baseline calibration, I set $\mathbb{1}_x(f_{t-1}, m_{t-1}) = 3$ if the financial and macroeconomic factors are both negative and equal to 1 otherwise. I estimate the other parameters in this model using $g(x_t) = x_t$ and omitting the state-dependence. Then I simulate a time series for y_t, f_t, and m_t using (7) with a few different choices for $g(\cdot)$ and the data for π_t and W_t. For each choice of $g(\cdot)$, I run several local projections at horizon $h = 0$ for the different approaches for detecting nonlinearities. The LPs have 196 observations and include 4 lags of all variables except W_t, which is not included at all to mimic the presence of omitted variables. The results are averages across 100 variations of (7) with 10,000 simulations each. Inference is performed with Huber-White standard errors (Montiel Olea et al., 2024) and results are materially the same using more involved estimations of the variance-covariance matrix (Xu, 2023).

Figure 9 plots the power of hypothesis tests using the indicator function approach and the even/odd weight decomposition of Caravello and Martínez Bruera (2024) across 3 specifications for the non-linear shock function $g(x_t)$ that feature both size and sign effects. The point of the plots is to show how power changes as we scale a component of $g(\cdot)$ by θ. With a foundation of $c \cdot \mathbb{1}_{x \geq c} + x \cdot \mathbb{1}_{x < c}$, the first specification has the first term scaled by θ, likewise for the second specification and the second term. This can be thought of as two ways of adjusting a jump then plateau of effects. The third specification is $\theta x^2 \cdot \mathbb{1}_{x \geq c} + x \cdot \mathbb{1}_{x < c}$.[22] The shock is standardized and the cutoff c is set to 1 in the DGP. Across simulations, the structural shocks follow a roughly but not perfectly symmetric distribution, making the even/odd approach a valid choice ex-ante.

The plots show that the indicator function approach strictly dominates the even/odd decomposition, though the gap is decreasing in the size of these specific nonlinearities. The Caravello and Martínez Bruera (2024) procedure mostly dominates the generated regressor approach for the different parameterizations, so it's excluded for clarity. The indicator function approach has the advantage of insignificant coefficients not being the end of the story; the estimates may be different enough that a null of linearity can be rejected. In principle, this advantage should extend to the generated regressors, but their unconventional construction clearly leads to even more inefficiency.

---

[22]The first two specifications were selected to give all methods as difficult task, with the third being a more traditional non-linear structure. The results for the first two specifications are similar if we multiply x through their corresponding $g(\cdot)$.

Difference in Power: Indicators vs. Even/Odd Approach
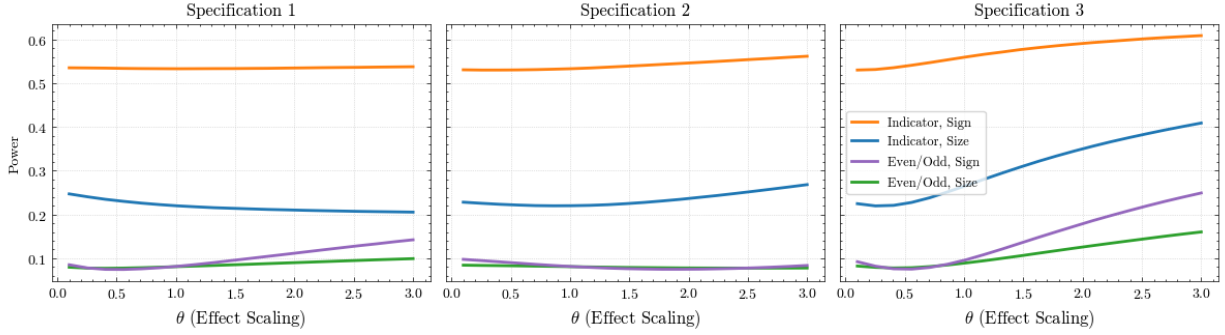
Size for the indicator function approach refers to a rejection of the null hypothesis that the coefficient on the big, positive shock and the small, positive shock are the same. Sign refers to a rejection that the big, positive and big, negative coefficient is the same. Even/Odd are simple significance tests of the coefficient on $f(x)$.

Figure 9

In every previous simulation, we have assumed the structural shock ε_t is perfectly observed. This will obviously not be the case in practice. As mentioned in Section 2, if the structural shock is not observed, the weights are unknown. Assuming that a proxy z_t is in fact the structural shock represents the ceiling on estimation quality. But this says nothing about how useful what we're actually estimating is. Kolesár and Plagborg-Møller (2025) show that if controls are needed for identification, we cannot have any faith in what we're estimating unless the propensity score is linear. Instead, the proxy should indeed be a proxy in a classical sense – departures from the structural shock amount to noisy measurement. This is merely a conjecture; measurement error itself can be complicated, even if induced from noise alone. Chen et al. (2011) show that the usual "attenuation bias" relationship to estimands does not hold if the measurement error is non-linear. Thankfully, I find in simulations that under a rich set of measurement error types (e.g., nonlinearities, heteroskedasticity, state-dependent noise) the "best case" weights are good approximations for the true weights. Even though the bias can go in either direction, that does not matter for the hypothesis testing procedure. Using a simulation of the first specification of DGP (7) with structural shock x_t, Figure 10 plots the case of $z_t = \text{sgn}(x)(x+u)^2$ where u is normally distributed, mean-0 noise with conditional variance $.01^2(1+x^2)$). If we run local projections using z_t and plot the weights as if $z_t = x_t$, the true weights are similar. The indicator function structure constrains deviations; out of several combinations tried, this was about as ugly as it got. One specific area of concern is the true weights are putting much more weight on shocks on the "wrong sign". But the consequences are limited to being less robust to false negatives.[23]

---

[23]This is another reason to use disjoint indicators. While it forces all coefficients to put some weight on the wrong sign, the values of ε_t where "wrong-sign weight" is placed will be the same (within the coefficient group). If we allow for overlap, we could have one coefficient with much more wrong-sign weight than the rest, and we can't know if it's an issue without the exact form of measurement error.
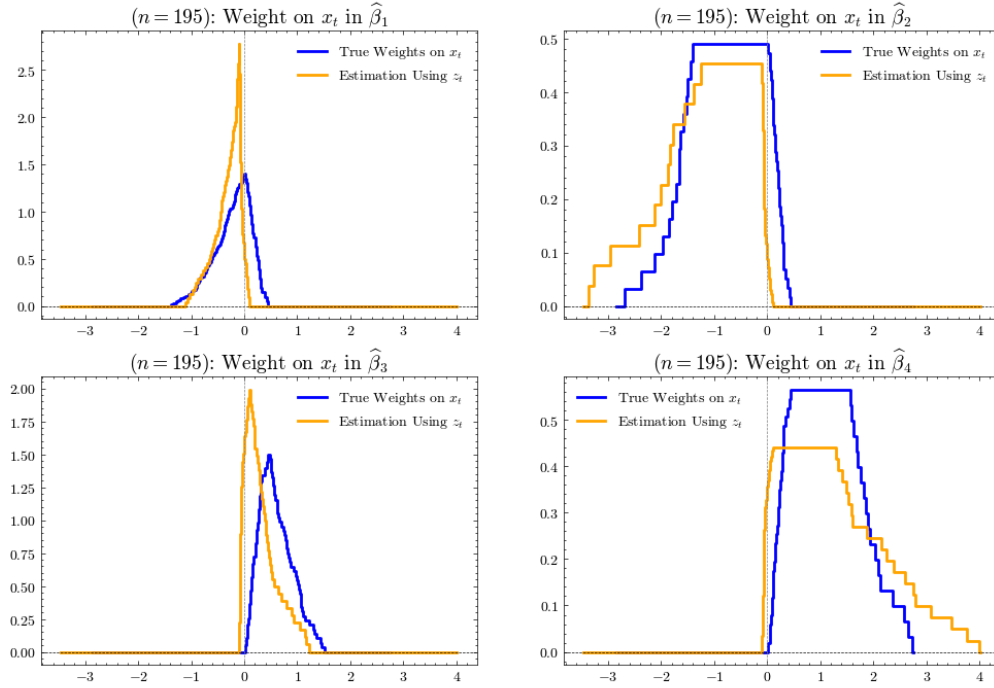
Figure 10: Selected Simulation of DGP (7) with Measurement Error

Finally, one issue that was not highlighted in the previous discussions is the issue of choosing thresholds. It's not ideal to have to mandate when a small shock becomes big etc. The first step to address this is to first standardize the data and then make the partitioning based on data realities. The threshold for magnitude in positive shocks need not be the same as for negative shocks. That should give an inkling as to a reasonable baseline to set, and then robustness exercises could involve moving this threshold around to see if results are sensitive to cutoffs. Interestingly, Figure 4 shows that while it *feels* like indicator functions involve setting paradoxical thresholds, there is a significant amount of weight overlap, so this is actually not as much of an issue. In fact, to decrease the amount of false negatives, one ironic way to address this is to allow indicators themselves to overlap. For example, the weights when using $f_1 = \mathbb{1}(\varepsilon_t < .01)$ and $f_2 = \mathbb{1}(\varepsilon_t < -1.5)$ have significantly less overlap than the disjoint case. The tension, as mentioned before, is that this increases standard errors. On the other hand, for generated regressors, there is no weight overlap and this paradox is unavoidable, though it can be diminished by fixing the peak at the median of the interval instead of having to choose it beforehand as well.

A.2 Inherency of Negative Weight

If ε_t is a continuously distributed shock on $I \subset \mathbb{R}$, note that[24]

$$\int_I \text{Cov}\big(\mathbf{1}_{\{a \geq \varepsilon_t\}}, f(\varepsilon_t)\big)\text{d}a = \int_I \Big\{\mathbb{E}[\mathbf{1}_{\{a \geq \varepsilon_t\}} f(\varepsilon_t)] - \mathbb{E}[\mathbf{1}_{\{a \geq \varepsilon_t\}}]\mathbb{E}[f(\varepsilon_t)]\Big\}\,\text{d}a$$

$$= \mathbb{E}\left[\big(f(\varepsilon_t) - \mathbb{E}[f(\varepsilon_t)]\big)\left(\int_{\underline{I}_t} \text{d}a\right)\right] = \mathbb{E}\big[(f(\varepsilon_t) - \mathbb{E}[f(\varepsilon_t)])\varepsilon_t\big] = \text{Cov}(f(\varepsilon_t), \varepsilon_t)$$

where $\underline{I}_t = \{x \in I : x \leq \varepsilon_t\}$. For a generic $f(\cdot)$ and $g(\cdot)$ (assume they are mean 0 WLOG), also notice

$$\int_I \text{Cov}\big(\mathbf{1}_{\{a \geq \varepsilon_t\}}, f(\varepsilon_t)\big) - \frac{\text{Cov}(f(\varepsilon), g(\varepsilon))}{\text{Var}(g(\varepsilon))} \int_I \text{Cov}\big(\mathbf{1}_{\{a \geq \varepsilon_t\}}, g(\varepsilon_t)\big) = \text{Cov}(f(\varepsilon_t), \varepsilon_t) - \frac{\text{Cov}(f(\varepsilon), g(\varepsilon))}{\text{Var}(g(\varepsilon))} \text{Cov}(g(\varepsilon_t), \varepsilon_t)$$

Recall from (4), the weight function on β_2 from (5) will follow $\omega_2(a) = \frac{\text{Cov}(\mathbf{1}_{\{a \leq \varepsilon_t\}}, f(\varepsilon)^\perp)}{\text{Var}(f(\varepsilon)^\perp)}$. Since $f^\perp(\varepsilon)$ in this case is $f(\varepsilon) - \frac{\text{Cov}(f(\varepsilon), \varepsilon))}{\text{Var}(\varepsilon)}$, $\int_I \omega_2(a)\text{d}a$ will be proportional to the above result when $g(\varepsilon) = \varepsilon$, yielding $\int_I \omega_2(a)\text{d}a \propto 0$. Thus, if $\omega_2(a) \neq 0$ for any a, ω_2 must necessarily take on negative values.

A.3 Standard Error Corrections

The necessary corrections are stated here and the proof can be found in Online Appendix B.2. Let $f_1(x)$ be the small, negative indicator, $f_2(x)$ be the big, negative indicator, and similarly for $f_3(x)$ and $f_4(x)$. The procedure uses generated regressors $g_i = \alpha_i f_i(x)$, where α_i is the coefficient on f_i in a projection of the shock x_t on all the functions (in a constant, and let the projection constant term be α_0). Let t_i be the t-statistic from the projection.

The point estimates of interests are coefficient sums, for which the variance is $\text{Var}\,\beta_i + \text{Var}\,\beta_j + 2\,\text{Cov}\,\beta_i, \beta_j$. Because the g_i functions are generated regressors, correction terms should be added, and then take a square root to deliver proper standard errors. Each point estimate will have 3 correction terms for each of the 3 parts of the coefficient sum. For the variances, $\frac{\beta_i^2}{t_i^2}$ should be added. For the off-diagonal term, it's $2\frac{\tilde{\beta}_i \tilde{\beta}_j}{\alpha_i \alpha_j} \text{Cov}(\alpha_i, \alpha_j)$, where $|\text{Cov}(\alpha_i, \alpha_j)|$ is actually just $\text{Var}(\alpha_0)$. These corrections will typically be negligible.

A.4 Unpacking the Weight Form

Recall the general form from (4) in Section 2

$$\omega_i = \frac{\text{Cov}(1_{a \leq \varepsilon_t}, X_i^\perp)}{\text{Var}(X_i^\perp)}$$

where $X_i^\perp$ is the residual from regressing X_i on the other elements in $\boldsymbol{X}_t$. We can unpack this definition to get things soley in terms of covariances and variance of terms of $\boldsymbol{X}_t$, which amounts to an expansion of the FWL theorem. To my knowledge, this expansion has not been done previously and for good reason – the full form amounts to several messy recursions that offer absolutely no insight to write out. However to motivate the use of

---

[24]This holds in the interior of I, see Kolesár and Plagborg-Møller (2025) Lemma 3 and Caravello and Martínez Bruera (2024) Lemma 1.

deep learning to address one of the central issues in this paper, it may be useful to see why it's difficult to conjure up functional forms that will produce appropriate weighting.

For what follows, consider X to be a generic matrix of N covariates in a regression (which can include a vector of 1s) and X_i to be its i-th element. Keeping with notation from earlier, $X_i^\perp$ is the residual from X_i on the remaining elements of X. WLOG, we will initially look at an example where $i = 1$. Further consider $X_n^{\perp_1}$ to be regressing the n-th element of X on its the remaining parts excluding X_1. Then

$$X_1^\perp = X_1 - \sum_{n=2}^{N} \frac{\text{Cov}(X_1, X_n^{\perp_1})}{\text{Var}(X_n^{\perp_1})} X_n$$

We can keep unpacking these terms but it should be clear that indexing is quickly going to become a nightmare because the "exclusions" will not be in a consistent ordering across the components (and sub-components, and sub-sub-components,...) of this summation. Things would have already got a bit messy notation wise had we done a formula for a generic $X_i^\perp$. So we will have to break this up into several parts. The details are tedious, so they are relegated to Online Appendix C.1. Those details allow us to explicitly write out the $N = 4$ special case of interest. Recall that the setting of interest is including functions $\{f_i(\varepsilon_t)\}_{i=1}^4$ in a regression, where ε_t is a shock. The weights $\omega_i(a)$ in β_i (corresponding to the i-th function) are

$$\omega_i(a) = \frac{C_{\mathbb{1},i} - \sum_{j \neq i} C_{\mathbb{1},j} \frac{C_{i,j} - \sum_{k \neq j} \frac{C_{i,k} C_{j,k}}{V_k}}{V_j - \sum_{k \neq j} \frac{C_{j,k}^2}{V_k}}}{V_i - \sum_{j \neq i} \frac{C_{\mathbb{1},j}^2 - 2C_{i,j} \sum_{k \neq j} \frac{C_{i,k} C_{j,k}}{V_k} + \sum_{k \neq j} \frac{C_{i,k}^2 C_{j,k}^2}{V_k^2}}{V_j - \sum_{k \neq j} \frac{C_{j,k}^2}{V_k}}}$$

where $C_{i,j}$ denotes the covariance between f_i and f_j, $C_{\mathbb{1},i}$ is the covariance between $\mathbb{1}_{(\varepsilon_t \geq a)}$ and f_i, and V_i is the variance of f_i. This $N = 4$ case is actually simple compared to the sprawling recursions of the general case. The representation above also implicitly assumes the functions are mean 0, which need not be the case.

As made explicit at the beginning of Section 3, the goal is to pick functions so that $\omega_i(a)$ are non-negative, relevant (don't put weight where we don't want), and hump-shaped. The inscrutable form above makes deep learning a natural solution to the complex function search in the case where we allow the functions to potentially be correlated.
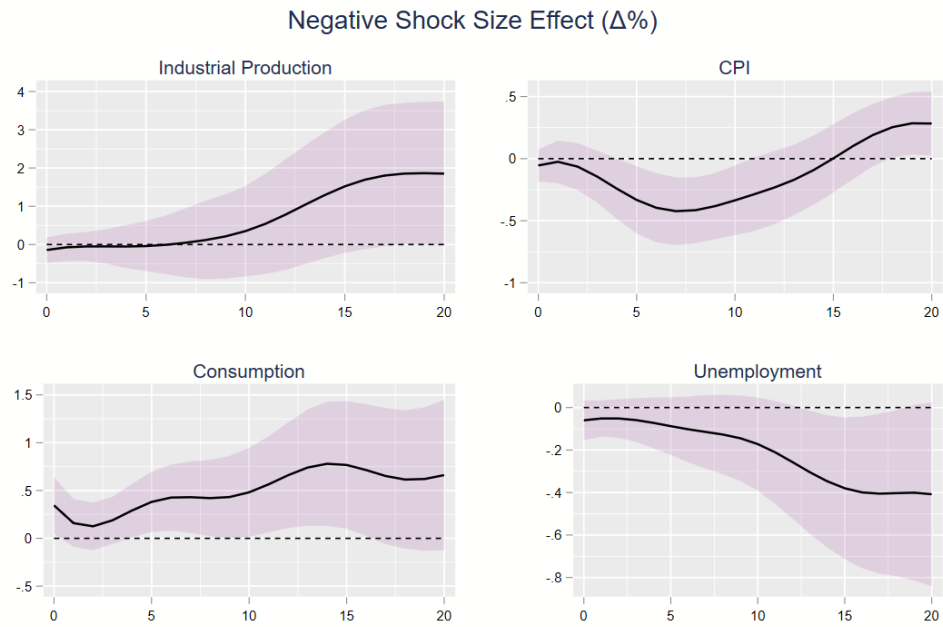
A.5 More Application Results
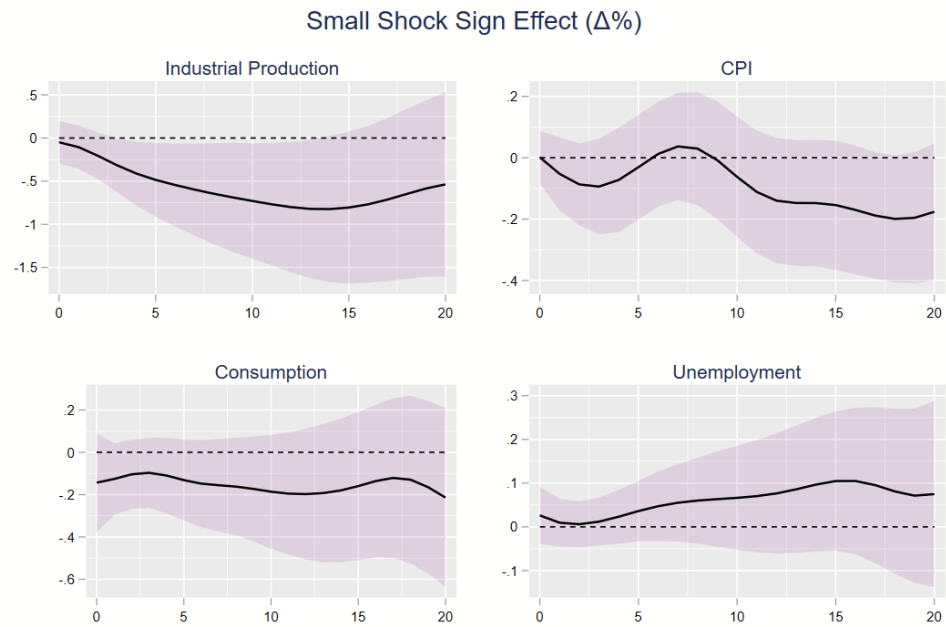

Figure 11: Indicator Function Approach with MP1 Shocks

Figure 12: Indicator Function Approach with MP1 Shocks