Discussion of: "Dynamic Causal Effects in a Nonlinear World: the Good, the Bad, and the Ugly"*

Sílvia Gonçalves[†], Ana María Herrera[‡]and Elena Pesavento[§]

April 14, 2025

Abstract

The ease of estimating linear local projections has made them a popular tool for impulse response function analysis. Kolesár and Plagborg-Møller's main goal is to inquire whether local projections (LP) estimands of impulse response functions have a causal interpretation when the data generating process (DGP) is nonlinear. This discussion focuses on two questions. First, how should we interpret the magnitude of the linear LP estimands in nonlinear environments? Second, is the linear LP useful when the researcher is interested in the effects of large shocks rather than small shocks?

JEL codes : C22, C32, H20, C51, E32, E52, E60, E62

*Acknowledgments: This work was supported by an NSERC grant from the government of Canada and NSF grants 2417534 and 2417535.

[†]McGill University, Department of Economics, 855 Sherbrooke St. W., Montréal, Québec, H3A 2T7, Canada. E-mail: silvia.goncalves@mcgill.ca.

[‡]University of Kentucky, Department of Economics, 550 South Limestone, Lexington, KY 40506-0034, USA. E-mail: amherrera@uky.edu.

[§]Emory University, Economics Department, 1602 Fishburne Dr. Atlanta, GA 30322, USA. E-mail: epesave@emory.edu.

"Functional misspecification is a fact of life — one almost never has information which justifies a particular linear or non-linear specification. Indeed, most econometric estimating relationships are intended as approximations, rather than as the 'truth'. It is therefore useful to realize the limitations of our approximations."

Halbert White (1980)
"Using Least Squares to Approximate Unknown Regression Functions"
International Economic Review, Vol. 21, No. 1, pp. 149-170

1 Introduction

Consider a practitioner who uses a linear local projection (LP) to estimate the impulse response of Y_{t+h} to a one-time shock in X_t given by

$$Y_{t+h} = \hat{\beta}_h X_t + \hat{\gamma}_h' W_t + \hat{r}_{t+h}, \tag{1}$$

where the coefficient $\hat{\beta}_h$ is obtained from an ordinary least squares (OLS) regression of Y_{t+h} on X_t and a set of controls W_t. Kolesár and Plagbrog-Møller ask: What does $\hat{\beta}_h$ estimate when the "true" model is nonlinear?

The simplicity of estimating local projections has made them a widely used tool to estimate impulse response functions in both linear and nonlinear models. While it is clear that LP estimates the causal dynamic effect of interest when the underlying DGP is linear, this is not obvious in the presence of nonlinearities. Yet, it has become common practice among practitioners to employ linear LPs to approximate the impulse response functions of nonlinear processes.

Kolesár and Plagborg-Møller's main finding is that the causal content of the linear LP estimand depends on how we identify the shock of interest. In particular, linear LP has a useful causal interpretation when the shock is observed (the "good"), but not necessarily when identification is achieved by heteroskedasticity (the "bad") or non-Gaussianity (the "ugly") restrictions. Given the predominance of linear LP in applied macroeconomics, this is a very timely and important paper. It offers a clear picture of the situations where linear LP can and cannot be useful for causal inference in macroeconomics, allowing for more informed decisions by researchers and practitioners using these methods.

In this discussion, we focus on the "good" and assume that the shock of interest is observed. This not only corresponds to the case where the linear LP estimand has a useful causal interpretation (as shown by this paper under very general conditions), but it is also empirically relevant as it covers the widely popular narrative identification approach. Our ultimate goal is to offer some suggestions intended to further clarify in what circumstances linear LP recovers useful notions of impulse response functions (IRFs) considered in the existing literature.

The rest of this discussion is organized as follows. In Section 2, we review two different definitions of IRFs that have been used in nonlinear impulse response analysis: the "average response function" (ARF) and the "average marginal response function" (AMRF). The first considers the effect of a fixed-sized δ shock whereas the second considers infinitesimally-sized shocks (i.e., $\delta \to 0$). Section 3 briefly reviews the form of the linear LP estimand as derived by the authors. We then relate it to the ARF and the AMRF. First, in Section 4.1, we show that the ability of the linear LP estimand to recover the average marginal response function depends on the type of nonlinearity and on the shape of the distribution of the shock of interest. In Section 4.2, we provide simulation evidence based on a model that includes different types of nonlinear regressors and different errors distributions. Linear LP may be strongly biased as an estimator of the average marginal response function when the nonlinearity captures a "size" effect as opposed to a "sign" effect. In contrast, a simple nonparametric estimator of the average marginal response function is able to eliminate this bias. In Section 5, we investigate the bias of the linear LP estimator when used to estimate the average response function. Our results suggest that this bias may be sizable unless the size of the shock is not too large relative to its standard deviation. Section 6 concludes.

2 Causal Effects in Nonlinear Models

Consider a special case of the causal framework in Kolesár and Plagborg-Møller given by

$$Y_{t+h} = \psi_h(X_t, U_{t+h}), \quad X_t \perp U_{t+h}$$

where ψ_h is a very general (unknown) function and U_{t+h} is a vector containing all variables (dated before, on, and after time t) that causally affect Y_{t+h}, other than X_t. The independence condition $X_t \perp U_{t+h}$ follows from the fact that X_t is a structural i.i.d. shock that is independent of the other shocks driving Y_t. The goal is to identify the causal dynamic effect of a one-time perturbation of size δ in X_t on the outcome variable Y_{t+h} for $h = 0, 1, \ldots$. This task is particularly challenging when the model is nonlinear as various definitions of the IRFs are available to the researcher.

A natural definition of this causal effect is the difference between the counterfactual when X_t is perturbed by δ and a baseline with no perturbation:

$$\underbrace{\psi_h(X_t + \delta, U_{t+h})}_{\equiv Y_{t+h}(\delta):\text{ counterfactual}} - \underbrace{\psi_h(X_t, U_{t+h})}_{\equiv Y_{t+h}:\text{ baseline}}. \tag{2}$$

Yet, the difficulty with obtaining the *ceteris paribus* effect of this perturbation stems from the fact that under nonlinear ψ_h, the difference defined in Equation (2) is random as it depends on X_t and (potentially) on U_{t+h}. The question then becomes: How should the researcher summarize the causal effect? An obvious answer would be to consider reporting the average impulse response function. But, which average? To clarify this point, we next review two possible average IRFs the researcher may consider.

Following Kolesár and Plagborg-Møller, we let

$$\Psi_h(x) \equiv E(\psi_h(x, U_{t+h}))$$

denote the *expected* potential outcome function. A definition considered in Gonçalves et al. (2021, 2024) is the average response function which considers the response to a shock of fixed size δ and is given as follows.

Definition 1 *The average response function of Y_{t+h} to a shock of fixed size δ in X_t is defined as*

$$ARF_h(\delta) = E(\Psi_h(X_t + \delta)) - E(\Psi_h(X_t)) = E(\psi_h(X_t + \delta, U_{t+h})) - E(\psi_h(X_t, U_{t+h})).$$

Such a definition would be of interest to a practitioner seeking to estimate the average effect of a "large shock", for instance, a 25 basis point shock to the fed funds rate or a 10% increase in oil prices on GDP growth.

A second definition available to the practitioner captures the response to an infinitesimally-sized shock, i.e. $\delta \to 0$, as follows.

Definition 2 *The average marginal response function of Y_{t+h} to a shock $\delta \to 0$ in X_t is defined as*

$$AMRF_h = E[\Psi_h'(X_t)] = \lim_{\delta \to 0} ARF_h(\delta)/\delta,$$

provided we can interchange the limit with the expectation operator.

Definition 2 is useful when capturing the effects of "small shocks", where "small" is defined in relation to the standard deviation of the X_t (see Gonçalves et al. (2024) for more on this point).

It is important to recall that while in linear models, these definitions coincide (up to scale), this is not the case in nonlinear models. Next, we briefly review the form of the linear LP estimand derived by Kolesár and Plagborg-Møller when X_t is an observed shock, with an eye geared towards relating it with Definitions 1 and 2. Our ultimate goal is to further clarify under what conditions linear LP can provide a good approximation to these IRF definitions.

3 The linear LP estimand

Proposition 1 of Kolesár and Plagborg-Møller proves that

$$\hat{\beta}_h \to_p \beta_h \equiv \int \omega(x) g_h'(x) dx,$$

where $g_h(x) \equiv E(Y_{t+h}|X_t = x)$. A causal interpretation of this estimand is obtained when X_t is independent of U_{t+h} since then $g_h(x) = \Psi_h(x) \equiv E(\psi_h(x, U_{t+h}))$, and $\beta_h = \int \omega(x)\Psi_h'(x)dx$. Importantly, when X_t is an observed shock –as it is the case when the shock is identified via a narrative approach– the weights $\omega(x)$ are "well behaved" in the sense that $\omega(x) \geq 0$ for all x and $\int \omega(x)dx = 1$. In other words, the linear LP can be interpreted as a convexly-weighted average of the marginal effects. In addition, the weights are humped-shaped and peak at the mean, $E(X_t)$. These properties guarantee that, as long as $g_h'(x)$ is always of the same sign, there is no sign reversal, i.e., the sign of the linear LP is consistent with the sign of $g_h'(x)$ for any x even when the DGP is nonlinear. In contrast, convex weights cannot be guaranteed when the shock is identified via heteroskedasticity ("the bad") or non-Gaussianity ("the ugly").

A very nice property of the proof is that the conditions under which the LP estimand can be given a causal interpretation are very general and allow for models with kinks (e.g. $f(x) = max(x, 0)$) and regime switching (state-dependent SVARs). In addition, the authors generalize the existing literature on this topic by allowing X_t to have unbounded support, which is key when discussing potential outcomes in macroeconomic settings where shocks are continuous and potentially unbounded. This is a great improvement over existing results in the literature that use bounded support.

While we agree that the property of no sign-reversal ensured by the convexity of the weights $\omega(x)$ is a minimum requirement for an estimand to have a causal interpretation, we question whether this is sufficient for justifying interest in linear LP when interest focuses on the magnitude of the response to a shock. For example, a long-standing question in the literature on government spending shocks has been what the magnitude is of the government spending multiplier (e.g., Ramey, 2011). The key question economists seek to answer is whether the multiplier is less than or more than unity. In this case, getting the sign of the multiplier correctly is not enough.

In the next section, we compare the LP estimand with the AMRF defined in Definition 2, which is a natural benchmark to the linear LP estimand. The AMRF may be easier to interpret as it relies on a pre-specified set of weights given by the density function of the shock X_t.

4 Linear LP and the Average Marginal Response

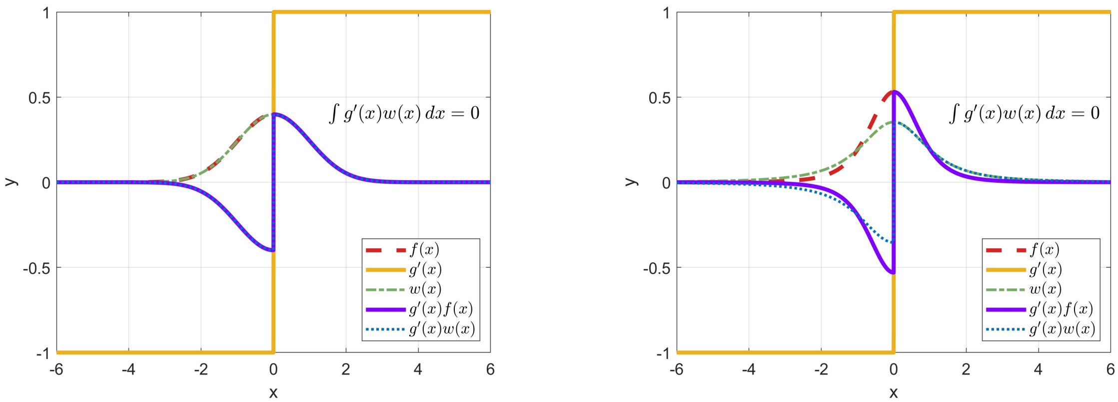
4.1 Large sample comparison

Letting $f(x)$ denote the density of X_t, we can write

$$AMRF_h = \int f(x) g_h'(x) dx.$$

The difference between β_h and $AMRF_h$ crucially depends on the functional form $g_h(x)$ and the density of X_t, $f(x)$. In particular, the two weighting schemes coincide if $\omega(x) = f(x)$, which is the case when $X_t \sim N(0, \sigma^2)$ as shown by Yitzhaki (1996). Moreover, as Caravello and Martínez Bruera (2024) demonstrate, if $f(x)$ is symmetric (even if not Gaussian), $\omega(x)$ is also symmetric. In this case, it can be shown that $\beta_h = AMRF_h$ if $g_h(x)$ is an even function (e.g., if $g_h(x) = |x|$), but not otherwise (e.g., if $g_h(x) = x^3$). Finally, when $f(x)$ is asymmetric, β_h is generally different from $AMRF_h$ and no ranking between β_h and $AMRF_h$ can be provided.

Figures 1 - 3 verify these statements and provide illustrations for functional forms of $g_h(x)$ often found in the literature and alternative densities for the shock of interest, X_t. Figure 1 displays the density $f(x)$, the weights, $\omega(x)$, the first derivative of the nonlinear function, $g_h'(x)$, and the products between $g_h'(x)$ and $f(x)$, and $g_h'(x)$ and $\omega(x)$, respectively, when $g_h(x) = \beta_0 x + |x|$ and $f(x)$ is symmetric. For simplicity, we set $\beta_0 = 0$. The linear LP estimand corresponds to the integral of $g_h'(x)\omega(x)$ whereas the AMRF is the integral of $g_h'(x)f(x)$. Note that here $g_h(x)$ is an even function, which could reflect sign nonlinearities (see Caravello and Martínez-Bruera (2024)). Panel (a) corresponds to the case where X_t



(a) $N(0,1)$ (b) $t(4)$

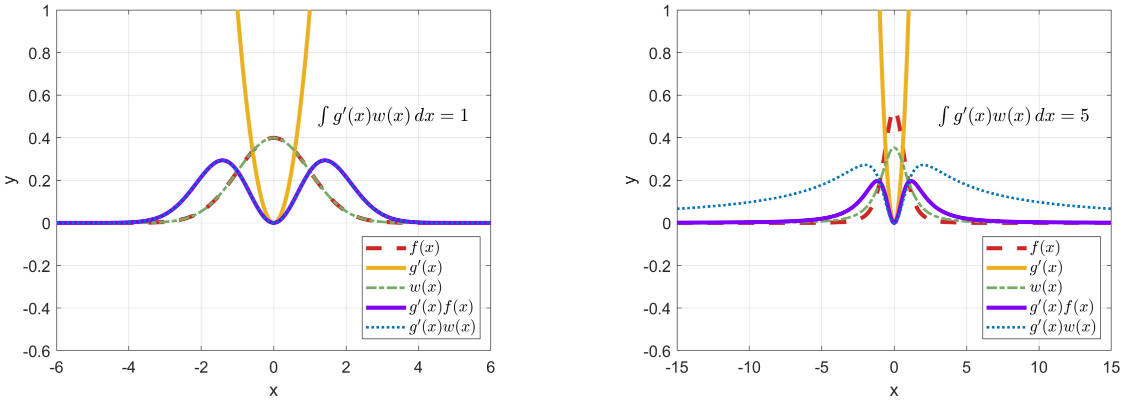
Figure 1: $g_h(x) = \beta_0 x + |x|$ (with $\beta_0 = 0$)

follows a standard normal distribution and panel (b) corresponds to the Student-t with 4 degrees of freedom standardized to have unit variance, which is intended to capture fat tails. As Figure 1 illustrates, the linear LP captures only the linear component (equal to zero in this example) of the IRF under symmetric (not necessarily Gaussian) shocks. Because the density is still symmetric, albeit with thicker tails, the integral of $g_h'(x)\omega(x)$ and $g_h'(x)f(x)$ coincide and are both zero, equal to the linear component of the IRF.[1]

Consider now the case where g_h is an odd function, capturing size dependence in the impulse response functions. Figure 2 depicts the weights, the function $g_h'(x)$, and their products when $g_h(x) = \beta_0 x + \frac{1}{3}x^3$ and the shocks are symmetrically distributed. Panel (a) illustrates the case when X_t follows a standard normal density and panel (b) when it follows a $t(4)$ distribution. As Figure 2 shows, in this case, the linear LP no longer equals zero, capturing the presence of "size" effects. Indeed, the magnitude of the linear LP remains unchanged if we add any even function to $g_h(x)$. A comparison between panels (a) and (b) reveals that $w(x) = f(x)$, i.e. the linear LP estimand coincides with the AMRF only when X_t is standard normal.

Finally, we illustrate the case where X_t follows an asymmetric distribution in Figure 3. To illustrate the difference between the linear LP estimand and AMRF, we let $f(x)$ be a Generalized Extreme Value (GEV) density with a shape parameter of 0, a scale parameter of 1, and a location parameter of 0, demeaned and standardized. As in the previous examples, in addition to the density function, we plot the implied causal weights, the nonlinear function $g_h'(x)$ and their products, for $g_h(x) = |x|$, an even function, and for

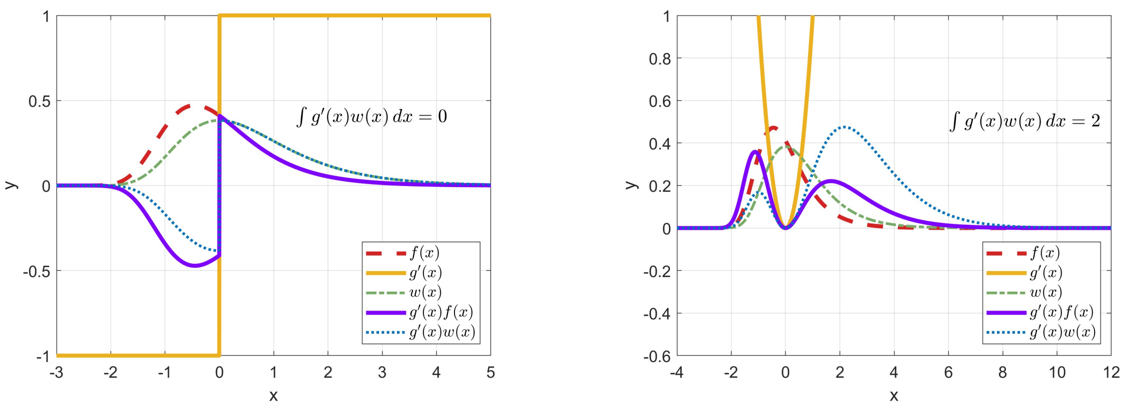
[1] We use a unit variance for ease of comparison across different distributions. However, as long as the variance is finite, it does not need to be equal to unity for the statements to hold.



(a) $N(0,1)$ (b) $t(4)$

Figure 2: $g_h(x) = \beta_0 x + \frac{1}{3}x^3$ (with $\beta_0 = 0$)

$g_h(x) = \frac{1}{3}x^3$, an odd function, in panels (a) and (b), respectively.



(a) $g_h(x) = |x|$ (b) $g_h(x) = \frac{1}{3}x^3$

Figure 3: $f(x)$ is $GEV(0,1,0)$

It is clear from Figure 3 that the linear LP no longer wipes out "sign" effects when the shock distribution is asymmetric. In general, the magnitude of the linear LP estimand depends on $\omega(x)$, $f(x)$, and the shape of $g_h(x)$, and no clear ranking between the linear LP estimand and the AMRF can be provided. For instance, in the case of the asymmetric $GEV(0,1,0)$ distribution illustrated in Figure 3, the magnitude of the linear LP estimates exceeds the corresponding AMRF.

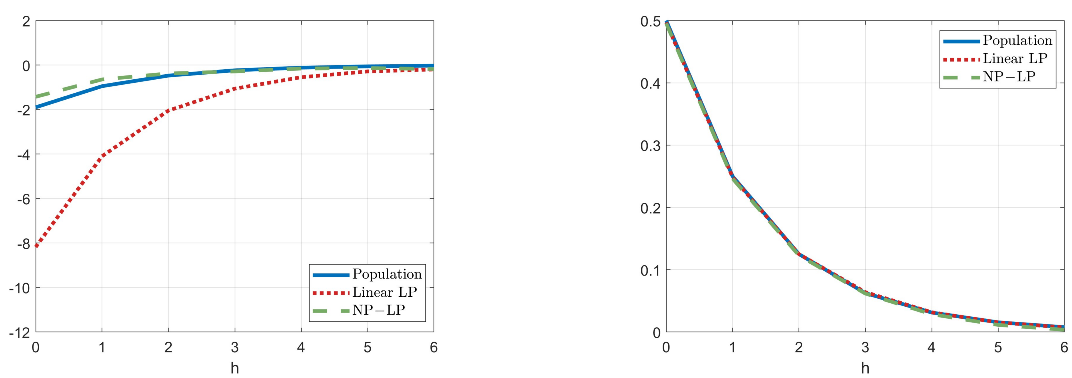
4.2 Small sample comparison

The previous section suggests that the discrepancy between the LP estimand and the AMRF depends on the shape of $g'_h(x)$ and the properties of $f(x)$. Nevertheless, employing such a straightforward method for estimating a parameter that may have a causal interpretation presents certain advantages. Specifically, as indicated by the authors, the linear LP obviates the necessity to nonparametrically estimate $g'_h(x)$, a task which can present considerable difficulties when confronted with small sample sizes commonly found in macroeconomics. In this section, we employ simulations to show that, despite these concerns, there are empirically relevant combinations of $g'_h(x)$ and $f(x)$ for which a nonparametric approach to estimate $AMRF_h$ seems promising in that it produces a small bias.

To do so we consider one of the DGPs in Gonçalves et al. (2021) where the shock is assumed to be identified via a narrative approach so that:

$$\begin{cases} x_t & = \varepsilon_{1t} \\ y_t & = 0.5x_t + 0.5y_{t-1} - 0.4h(x_t) + \varepsilon_{2t}. \end{cases}$$

We examine the scenarios in which the nonlinear regressor $h(x_t)$ either equals x_t^3 or $|x_t|$. The distribution of ε_{1t} is characterized as either symmetric, following a t-distribution with 4 degrees of freedom, or asymmetric, following a Generalized Extreme Value (GEV) distribution with a shape parameter of 0, a scale parameter of 1, and a location parameter of 0. Furthermore, ε_{2t} follows a $N(0,1)$ distribution. We compare the linear LP estimator and a nonparametric estimator of $AMRF_h = E[g'_h(X)]$, where $g'_h(x)$ is obtained via a local linear kernel regression. To illustrate the behavior of the estimators when the sample size is small we set $T = 250$ (slightly more than twenty years of monthly data or 80 years of quarterly data) and estimate the average response over 1,000 Monte Carlo draws.

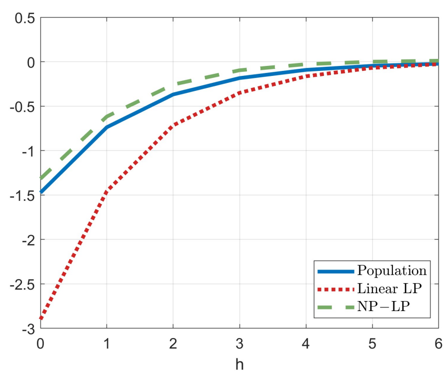


(a) $h(x) = x^3$ (b) $h(x) = |x|$

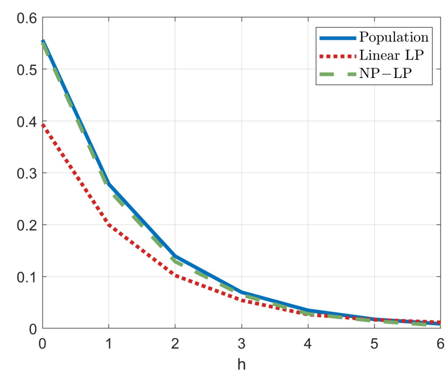
Figure 4: Comparison of linear LP and local linear, t_4 distribution.

Figure 4 plots the true impulse response function as defined by Definition 2, the average of the linear LP estimates and the nonparametric estimates over the 1,000 draws. The figure shows that when the distribution is symmetric but heavy-tailed, the nonparametric estimator differs from the local linear estimator when $h(x) = x^3$, as expected from the previous section. The local linear estimator is centered around the AMRF, the target parameter, whereas the linear LP is not (unsurprisingly). In contrast, the performance of the linear LP and the nonparametric estimator is similar when $h(x) = |x|$, as expected from our discussion in the previous section.

Figure 5 plots the simulation results when X_t follows an asymmetric distribution. As in the previous section we employ a GEV(0,1,0) distribution. We note that when the distribution is asymmetric the linear LP estimator underperforms the nonparametric estimator in terms of bias if the target IRF is the AMRF. These findings indicate that the nonparametric estimator constitutes a viable alternative, even in scenarios where researchers encounter the small sample sizes frequently observed in macroeconomic studies.



(a) $h(x) = x^3$



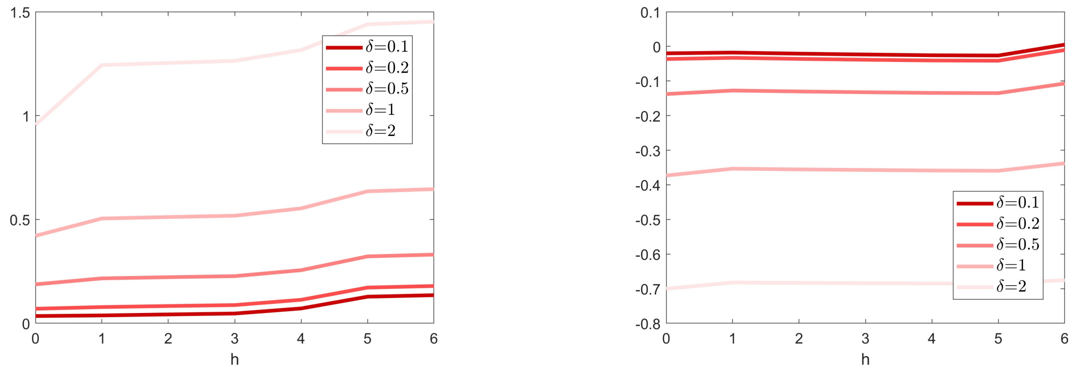
(b) $h(x) = |x|$

Figure 5: Comparison of LP with local linear, GEV distribution.

5 Linear LP and "large shocks"

Lastly, we explore what happens when the practitioner is interested in estimating the impulse response to a fixed-size shock instead of $\delta \to 0$. Our interest in answering this question stems from empirical applications in which the researcher is interested in estimating the response of GDP growth to oil price shocks that exceed one standard deviation (e.g., an unexpected 10% increase in the oil price), the impact of a military spending shock that amounts to the military build-ups experienced during a war (e.g., the World War II military spending shock is 12 times the standard deviation over the 1890Q1 to 2015Q4 sample), or government spending multipliers when the size of the shock exceeds one standard deviation.

To explore how the bias of the linear LP evolves as the size of the shock increases relative to the standard deviation of x_t, we employ the same DGP as in the previous section and compute the relative bias of the linear LP to the population average response function as defined in Definition 1. We again examine the scenario wherein $h(x_t)$ corresponds to either x_t^3 or $|x_t|$. The distribution of ε_{1t} is $N(0,1)$, which, as we illustrated in the previous section, is the distribution most favorable to linear LP. Under this distribution, the standard deviation of X_t is $\sigma = 1$ and we can interpret δ as the ratio δ/σ.



(a) $f(x) = x^3$ (b) $f(x) = |x|$

Figure 6: Percentage bias of LP relative to population ARF, N(0,1) distribution.

As Figure 6 demonstrates, the relative bias increases with the ratio of $\frac{\delta}{\sigma}$. Indeed, even for shocks of size $\delta = 1$ (i.e., one standard deviation), the relative bias exceeds 30% at all horizons for x^3 and 50% after impact for $|x|$. These simulations suggest that even in cases where the shocks are normally distributed and the shock is identified through a narrative scheme, the "good" still has a downside: the linear LP results in a biased estimate of the ARF when the shock is "large".

6 Conclusions

Although linear LP has a causal interpretation under convex weights when the shock of interest is directly observed, interpreting the LP estimates may prove difficult as these depend on the shape of $g_h(x)$ and on the density of X_t. As the authors suggest, plotting the weights provides useful information to the researcher interested in estimating the marginal effect and identifying the sign of the response, but it might not suffice if the object of interest is the magnitude of the response. A natural benchmark in this case is the average marginal response function, whose weights are the density function of the shock. Whether linear LP provides a good approximation to this notion of IRF depends on the type of nonlinearity and on the density of the shock. Another potential limitation of the linear LP approach is the difficulty in measuring the impact of large shocks.

Ultimately, if we care about the magnitude of the IRFs or the impact of large shocks, we might not be able to escape some form of nonparametric estimation (which is in line with Section 6 of the paper). Alternatively, we may have to constrain ourselves to a parametric model that is rich enough to minimize the risk of misspecification and to justify the IRFs we get from that model.

References

Caravello, Tomás, and Pedro Martinez-Bruera. "Disentangling sign and size non-linearities." Available at SSRN 4704050 (2024).

Gonçalves, Sílvia, Ana María Herrera, Lutz Kilian, and Elena Pesavento. "Impulse response analysis for structural dynamic models with nonlinear regressors." *Journal of Econometrics* 225, no. 1 (2021): 107-130.

Gonçalves, Sílvia, Ana María Herrera, Lutz Kilian, and Elena Pesavento. "State-dependent local projections." *Journal of Econometrics* 244, no. 2 (2024): 105702.

Ramey, Valerie A. "Identifying government spending shocks: It's all in the timing" *The Quarterly Journal of Economics* 126, no. 1 (2011): 1-50.

Yitzhaki, Shlomo. "On using linear regressions in welfare economics" *Journal of Business & Economic Statistics* 14, no. 4 (1996): 478-486.

White, Halbert. "Using least squares to approximate unknown regression functions." *International Economic Review* (1980): 149-170.