Federal Reserve Bank of Dallas

Nonparametric Local Projections

Sílvia Gonçalves, Ana María Herrera, Lutz Kilian and Elena Pesavento

Working Paper 2414 **November 2024**

Research Department

https://doi.org/10.24149/wp2414

Nonparametric Local Projections*

Sílvia Gonçalves†, Ana María Herrera‡, Lutz Kilian§ and Elena Pesavento±

November 15, 2024

Abstract

Nonlinearities play an increasingly important role in applied work when studying the responses of macroeconomic aggregates to policy shocks. Seemingly natural adaptations of the popular local linear projection estimator to nonlinear settings may fail to recover the population responses of interest. In this paper we study the properties of an alternative nonparametric local projection estimator of the conditional and unconditional responses of an outcome variable to an observed identified shock. We discuss alternative ways of implementing this estimator and how to allow for data-dependent tuning parameters. Our results are based on data generating processes that involve, respectively, nonlinearly transformed regressors, state-dependent coefficients, and nonlinear interactions between shocks and state variables. Monte Carlo simulations show that a local-linear specification of the estimator tends to work well in reasonably large samples and is robust to nonlinearities of unknown form.

JEL Codes: C14, C32, E52

Keywords: impulse response, local projection, nonparametric estimation, nonlinear structural model, potential outcomes

---

*We thank seminar participants at Vanderbilt and Pittsburgh University as well as participants at the 2023 IAAE Conference for helpful comments. Jung Jae Kim and Iones Kelanemer Holban provided excellent research assistance. This paper is based on research supported by the NSF under Grants No. SES-2417534 and SES-2417535 and by a Natural Sciences and Engineering Research Council of Canada (NSERC) Grant No. RGPIN-2021-02663. The views expressed in this paper are those of the authors and do not necessarily represent the views of the Federal Reserve Bank of Dallas or the Federal Reserve System.

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

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

§ Lutz Kilian, Federal Reserve Bank of Dallas, Research Department, 2200 N. Pearl St., Dallas, TX 75201, USA. E-mail: lkilian2019@gmail.com.

± Elena Pesavento, Emory University, Economics Department, 1602 Fishburne Dr., Atlanta, GA 30322, USA. E-mail: epesave@emory.edu.

1 Introduction

Impulse response analysis is a cornerstone of empirical macroeconomics. Local projections have become a popular method for estimating impulse response functions (IRFs), especially when the shock of interest can be directly observed. In their simplest form, local projections consist of a sequence of OLS regressions, one for each horizon of interest. Their popularity in no small part stems from their simplicity. The impulse response of interest may be recovered from the estimated regressions without further transformations of the model coefficients or the need for Monte Carlo integration methods.

These advantages seem even more compelling when estimating nonlinear responses. For example, a large empirical literature has used generalizations of local linear projections to evaluate state-dependent impulse responses and other nonlinear responses. As shown in Gonçalves et al. (2021, 2024), however, seemingly natural generalizations of local linear projections to nonlinear models may fail to recover the population responses of interest.

For example, standard state-dependent local projection estimators of how the effect of monetary policy shocks changes with the level of government debt or of how the effect of fiscal spending shocks changes over the business cycle are inconsistent under empirically plausible conditions. An alternative approach that allows for state dependence of unknown form is the nonparametric local projection (LP) estimator recently proposed by Gonçalves et al. (2024).

The idea underlying this estimator is simple and intuitive. Consider a potential outcomes framework that allows for a general class of nonlinear structural dynamic models. Define the unconditional average impulse response function of an outcome variable y_{t+h} to a shock of size δ in the structural innovation of interest, ε_{1t}, as the expected value of the difference between the potential outcome with this shock and without the shock. Under the assumption that ε_{1t} is identified, it can be shown that the average structural impulse response function can be estimated by a two-step procedure. One first estimates nonparametrically the conditional mean function of y_{t+h} given ε_{1t}. One then takes the average over the sample of the difference between the conditional mean function evaluated at $\varepsilon_{1t} + \delta$ and ε_{1t}.

A similar two-step procedure can be applied to estimate the average response function conditional on the most recent state, with the difference that we condition on ε_{1t} and Ω_t in the first step, where Ω_t denotes the conditioning set of interest. Likewise, the second step has to be modified by replacing the sample average of the difference between the two conditional mean functions by a conditional average.

While this nonparametric LP estimator was originally proposed in the context of state-dependent models, in this paper we show that variations of this approach are valid more generally when modeling nonlinear response functions. We consider a range of nonlinear settings that are relevant for applied macroeconomists, including processes with nonlinearly transformed regressors (e.g., Herrera et al. (2015), Tenreyro and Thwaites (2016), Ben Zeev, Ramey and Zubairy (2023), Caravello and Martinez-Bruera (2024)), state-dependent coefficients (e.g., Ramey and Zubairy (2018), Alloza (2022)) and nonlinear interactions between shocks and state variables (e.g., Caramp and Feilich (2022), Cloyne et al. (2021), (2024)). We illustrate how to define the population impulse responses in these applications, how these responses may be interpreted within a potential outcomes framework, and how they can be recovered by a nonparametric LP estimator. We also provide guidance on how to implement this estimator in practice, and we examine its ability to recover the population responses.

The paper explores several alternative ways of implementing the nonparametric LP estimator, allowing for sample-size dependent tuning parameters. Monte Carlo simulations show that a local-linear specification of the nonparametric LP estimator tends to work well in reasonably large samples and is robust to nonlinearities of unknown form. We provide high-level conditions for the consistency of the estimator and show by simulation that it converges toward the population response in the RMSE sense in all our applications, as the sample size increases. We also examine how the specification of the nonlinear transformation affects the ability of the estimator to capture the size and sign asymmetries frequently discussed in applied work (e.g., Tenreyro and Thwaites (2016), Caravello and Martinez-Bruera (2024)).

The remainder of the paper is organized as follows. In Section 2, we introduce a general structural dynamic model that includes three examples of data generating processes commonly used in applied work. Section 3 defines the nonlinear population impulse response function of interest and contrasts our impulse response definition with an alternative definition that has been used in the literature (e.g., Koop et al. (1996)). Sections 4 and 5 discuss identification and estimation. The simulation results are in Section 6. Section 7 contains an empirical illustration focusing on possible nonlinearity in the response to monetary policy shocks. The concluding remarks are in Section 8.

2 Framework

Let $z_t = (x_t, Y_t')'$ denote an $n \times 1$ vector of observed time series, where $Y_t = (y_{2t}, \ldots, y_{nt})'$. When $n = 2$, we let $Y_t = y_t$. For instance, Y_t could be real GDP and x_t a measure of government spending shocks or monetary policy shocks. The data generating process (DGP) for z_t is described by the system of structural dynamic nonlinear equations:

$$\begin{aligned} x_t &= \varphi\left(\mathbf{z}_{t-1}\right) + \varepsilon_{1t} \qquad (1) \\ y_{it} &= \psi_i\left(x_t, Y_{-i,t}, \mathbf{z}_{t-1}, \varepsilon_{it}\right) \text{ for } \; i = 2, \ldots, n, \end{aligned}$$

where $\varepsilon_t \equiv (\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{nt})'$ denotes a vector of i.i.d. structural shocks with mean zero and covariance matrix Σ. φ and ψ_i are general nonlinear functions. Boldface is used to denote the history of a variable up to the time period in question. For instance, $\mathbf{z}_{t-1}$ is a vector containing $z_{t-1}, z_{t-2}, \dots, z_{t-p}$ for some lag order p which could be ∞. We let $Y_{-i,t}$ denote a $(n-2) \times 1$ vector that excludes y_{it} from the $(n-1) \times 1$ vector Y_t. We use a similar notation to define other vectors (e.g., $\varepsilon_{-i,t}$ corresponds to ε_t without the element ε_{it}).

Consistent with much of the literature, we exclude Y_t from the structural equation for x_t and assume that the model is additive in ε_{1t}. This exclusion restriction identifies the structural shock of interest ε_{1t}, whose causal effects on the elements of Y_t we are interested in estimating. An important special case arises when the structural shock of interest is observed such that $x_t = \varepsilon_{1t}$, as in the narrative approach to identification. We assume that ε_{1t} is independent of the other shocks, as is standard in the literature on nonlinear responses. We do not assume identification of $(\varepsilon_{2t}, \dots, \varepsilon_{nt})$. Several models commonly used in applied macroeconomics emerge as special cases of framework (1). Below we discuss four stylized examples that will be used to illustrate the implementation of the proposed estimation method and to asses its accuracy compared to alternative estimation methods.

To build intuition, first consider a simplified version of Angrist and Kuersteiner (2011)'s Example 1, which illustrates how potential outcomes can be computed in linear structural vector autoregressive (SVAR) models used to study the effects of monetary policy shocks (e.g., Bernanke and Blinder (1992); Bernanke and Mihov (1998)).

Example 2.1 (Linear model) *Let*

$$\begin{cases} x_t = \varepsilon_{1t}, \\ y_t = \beta x_t + \gamma y_{t-1} + \varepsilon_{2t}, \end{cases}$$

where $\varepsilon_t \equiv (\varepsilon_{1t}, \varepsilon_{2t})' \sim i.i.d.(0, \Sigma)$, *and* $\Sigma = diag\,(\sigma_1^2, \sigma_2^2)$.

Our second example is a model with nonlinearly transformed regressors. This model allows for a sign nonlinearity in the responses when the magnitude of the response depends on the sign of the shock (e.g., $f(x_t) = \max\{x_t, 0\}$) or a size nonlinearity when the magnitude of the response depends on the size of the shock (e.g., $f(x_t) = x_t^3$). Although the regression model is linear in the parameters, the impulse response function is nonlinear, requiring the use of nonstandard estimation methods (e.g., Kilian and Vigfusson (2011), Gonçalves et al. (2021)). Models with nonlinearly transformed regressors have been used extensively in applied macroeconomics. Examples include studies of the asymmetry in the responses to positive and negative oil price shocks (e.g., Herrera, Lagalo and Wada (2015)) as well as nonlinearities in the response of GDP to monetary policy shocks (e.g., Tenreyro and Thwaites (2016), Ascari and Haber (2022)), financial shocks (e.g., Forni et al. (2024)) and fiscal shocks (e.g., Ben Zeev et al. (2023)).

Example 2.2 (Model with Nonlinear Regressors) *Let*

$$\begin{cases} x_t = \varepsilon_{1t}, \\ y_t = \beta x_t + \rho y_{t-1} + cf(x_t) + \varepsilon_{2t}, \end{cases}$$

where $\varepsilon_t \equiv (\varepsilon_{1t}, \varepsilon_{2t})' \sim i.i.d.(0, \Sigma)$, $\Sigma = diag(\sigma_1^2, \sigma_2^2)$, *and* f *is a nonlinear function.*

A third example is the state-dependent model examined in Gonçalves et al. (2024) in which the response is allowed to differ between two observed states (e.g., expansion and recession) based on a dummy variable indicator S_{t-1}. Models of this type have been used extensively to study the magnitude of the fiscal multiplier, the effectiveness of monetary policy, and the impact of uncertainty shocks in expansions and recessions (e.g., Ramey and Zubairy (2018), Cacciatore and Ravenna 2021, Falck et al. 2021).

Example 2.3 (State-Dependent Model) *Let*

$$\begin{cases} x_t = \varepsilon_{1t} \\ y_t = \beta_{t-1} x_t + \gamma_{t-1} y_{t-1} + \varepsilon_{2t}, \end{cases}$$

where $\{\varepsilon_{1t}\}$ *is independent of* $\{\varepsilon_{2t}\}$ *and* $\varepsilon_t \equiv (\varepsilon_{1t}, \varepsilon_{2t})' \sim i.i.d.(0, \Sigma)$, $\Sigma = diag(\sigma_1^2, \sigma_2^2)$. *Let* $\beta_{t-1} = \beta_E S_{t-1} + \beta_R(1 - S_{t-1})$ *and* $\gamma_{t-1} = \gamma_E S_{t-1} + \gamma_R(1 - S_{t-1})$, *where* S_{t-1} *is a dummy variable indicating whether the economy is in an expansion or in a recession. Formally, we let* $S_t = \eta(w_r : r \leq t)$ *where* $\eta(\cdot)$ *is the composition of the indicator function and the function of* $\{w_r : r \leq t\}$ *used to indicate whether* S_t *equals 1 or 0. Here,* $\{w_r = (x_r, y_r, q_r)' : r \leq t\}$ *is a set which contains the random variables used to construct* S_t. *These potentially include the endogenous variables in the system* $z_t = (x_t, y_t)'$ *and their lags, as well as other exogenous variables* q_t *(and their lags).*

A final example is inspired by Cloyne et al. (2021, 2024) and Caramp and Feilich (2024) who consider a model in which the responses of y_{t+h} to ε_{1t} are allowed to be heterogeneous, with the heterogeneity being captured by an observable variable, say, r_t. For instance, imagine a situation in which monetary policy shocks, ε_{1t}, have an heterogeneous effect on GDP growth, y_t, that depends on the level of government debt, r_t. The level of debt, in turn, is a function of monetary policy in the previous period (x_{t-1}) through its effect on interest rates. The fact that the debt level is potentially correlated with the shock of interest induces a nonlinearity that needs to be taken into account when estimating the impulse response function. Note that this specification differs from the state-dependent model discussed earlier in that the model coefficients do not depend on the state, but the impulse response does.

Example 2.4 (Nonlinear Interaction Term of Shock with State) *Let*

$$\begin{cases} x_t = \varepsilon_{1t} \\ y_t = \beta_{21} x_t + \beta_{23} r_t + \alpha_{21} x_t r_t + \gamma_{21} y_{t-1} + \varepsilon_{2t} \\ r_t = f(x_{t-1}) + \varepsilon_{3t}, \end{cases}$$

where $\{\varepsilon_{1t}\}$ *is orthogonal to* $\{\varepsilon_{2t}\}$ *and* $\{\varepsilon_{3t}\}$, ε_{2t} *and* ε_{3t} *are potentially correlated,* x_t *is the shock of interest, and* r_t *is an observable variable that may change the effect of the policy shock. The form of the function* f *is unknown and can be linear or nonlinear.*

3 Population Impulse Response Functions

This section formally defines the average impulse response function and conditional average response function in a potential outcomes framework and illustrates the application of this framework in the context of the examples presented in the previous section.

3.1 Definitions of impulse response functions

Our goal is to identify the causal dynamic effect of a one-time perturbation in ε_{1t} on the outcome variable $y_{i,t+h}$, where $i = 2, \ldots, n$. For example, we may be interested in the response of GDP growth to a one-time exogenous shock to government spending.

To accomplish this goal, we rely on the notion of potential outcomes, which we describe next. First, note that one implication of model (1) is that we can write $Y_{t+h} = m_h(\varepsilon_{1t}, U_{t+h})$, where m_h is a (vector) function of the shock of interest ε_{1t} and $U_{t+h} \equiv \left(\varepsilon_{t+h}, \varepsilon_{t+h-1}, \ldots, \varepsilon_{t+1}, \varepsilon_{-1,t}, \mathbf{z}'_{t-1}\right)'$, with $\varepsilon_{-1,t}$ denoting ε_t without ε_{1t}. The mapping m_h can be linear or nonlinear. Given this mapping, the potential outcome associated with Y_t is given by

$$Y_{t+h}(e) = m_h(e, U_{t+h})$$

where $e \in E$ is any fixed value in the support of ε_{1t}. Given e, we have a collection of random variables given by

$$\{Y_{t+h}(e) : e \in E\}.$$

This is analogous to the treatment effect literature, where ε_{1t} is a binary treatment, $e \in \{0,1\}$, and we obtain $Y_{t+h}(0)$ and $Y_{t+h}(1)$. In the dynamic nonlinear model (1), ε_{1t} is a continuous treatment so that, given the potential outcome process $Y_{t+h}(e) = m_h(e, U_{t+h})$, the observed value of the target variable is $Y_{t+h} = m_h(\varepsilon_{1t}, U_{t+h})$. Put differently, it is the value that we observe when the treatment e is the random variable ε_{1t} that generated the observed data.

The fact that ε_{1t} and U_{t+h} are mutually independent random variables can be used to show that the potential outcomes are independent of ε_{1t}. This is the standard (conditional) independence assumption used in the treatment effects literature to link functionals of potential outcomes to functionals of observed data.

Our definition of the impulse response function is based on comparing the baseline value $Y_{t+h}(\varepsilon_{1t})$ with the counterfactual value of Y at $t+h$ that would have been observed if ε_{1t} had been subject to a shock of size δ, denoted $Y_{t+h}(\varepsilon_{1t}+\delta)$ (see, e.g. Potter 2000). In particular, following Gonçalves et al. (2024), we adopt the following definition:

Definition 1 *The average response function and the conditional average response function of Y_{t+h} to a shock of size δ in ε_{1t} are defined respectively as*

$$\begin{aligned} ARF_h(\delta) &\equiv E\left(Y_{t+h}(\varepsilon_{1t}+\delta) - Y_{t+h}(\varepsilon_{1t})\right) \\ CAR_h(\delta,\omega) &\equiv E\left(Y_{t+h}(\varepsilon_{1t}+\delta) - Y_{t+h}(\varepsilon_{1t})\,|\Omega_t = \omega\right), \end{aligned}$$

where Ω_t denotes the conditioning set.

Note that $ARF_h(\delta)$ defines the unconditional average response as in Gonçalves et al. (2021), whereas the $CAR_h(\delta,\omega)$ defines the conditional average response as, for example, in Gonçalves et al. (2024). Both definitions have been used in applied work.[1] Because our treatment is continuous, we compare $Y_{t+h}(\varepsilon_{1t}+\delta)$ against $Y_{t+h}(\varepsilon_{1t})$, where the latter corresponds to the observed value Y_{t+h}, whereas the former denotes the counterfactual value $Y_{t+h}(\varepsilon_{1t}+\delta)$ that is not observed. Since ε_{1t} is random, the conditional expectation in Definition 1 averages over all possible realizations of ε_{1t} (in addition to the other sources of randomness that enter into the potential outcomes), conditionally on Ω_t. The choice of ω in $CAR_h(\delta,\omega)$ is context-dependent. For instance, in Example 2.3 the conditioning set Ω_t is the state variable at time $t-1$, i.e. $\Omega_t = S_{t-1}$, implying that ω is either 0 or 1, while in Example 2.4 the conditioning set is $\Omega_t = r_t$, implying that ω can take on any value in the support of r_t.[2]

Definition 1 is not the only possible definition of the IRF. Other studies such as Koop, Pesaran, and Potter (1996) and Kilian and Vigfusson (2011), for example, have instead compared the two potential outcomes $Y_{t+h}(e')$ and $Y_{t+h}(e)$, setting $e'=\delta$ and $e=0$, which yields the alternative definition below.

[1]Alternatively, one could have considered the marginal and the conditional marginal response functions as discussed in Goncalves at al. (2024). In this paper, we focus on responses to a shock of finite magnitude δ, as is common in applied work (e.g., Ramey and Zubairy (2018)), rather than on responses to an infinitesimal shock. Thus, the natural object of interest is the average response functions.

[2]The appropriate conditioning set depends on the context. For example, Koop et al. (1996) condition on all the information at time $t-1$ that can be used to forecast y_t, while Gourieroux and Lee (2023) condition on y_t.

Definition 2 *The response function and the conditional response function of Y_{t+h} to a shock of size δ in ε_{1t} are defined respectively as*

$$ARF_h^*(\delta) = E(Y_{t+h}(e+\delta) - Y_{t+h}(e))$$
$$CAR_h^*(\delta,\omega) = E(Y_{t+h}(e+\delta) - Y_{t+h}(e)\,|\Omega_t = \omega),$$

where Ω_t denotes the conditioning set.

Whereas Definition 2 has been used widely in the literature, Definition 1 is more recent (see Goncalves et al. (2021), (2024)).These two definitions are equivalent in linear models and, more generally, when the the potential outcome is linear in e for all horizons, as would be the case for a linear model, as in Example 2.1, or in special cases of Example 2.3 and Example 2.4 when the conditioning sets (S_{t-1} and r_t respectively) are exogenous. In general, however, the two definitions differ.

Figure 1 illustrates these differences by example. Consider the nonlinear DGP:

$$\begin{aligned} x_t &= \varepsilon_{1t} \\ y_t &= 0.5y_{t-1} + 0.5x_t + 0.3x_{t-1} - 0.4f(x_t) - 0.3f(x_{t-1}) + \varepsilon_{2t} \end{aligned}$$

where ε_{1t} and ε_{2t} are independent and have a standard normal distribution. For illustrative purposes, let the magnitude of the shock be $\delta = 2$ and the functional forms $f(x_t) = max(x_t, 0)$ and $f(x_t) = x_t^3$, respectively. The solid red line in Figure 1 denotes the ARF obtained as the average over the y_t obtained for different realizations of ε_{1t}, whereas the dashed line denotes the value of ARF^* obtained by setting $\varepsilon_{1t} = e = 0$. It is readily apparent that in this example, the two definitions of the impulse response imply quite different measures of the conditional expectation of y_t in the absence of a perturbation.

Which approach is the more natural one? The only difference between these two approaches is the treatment of the impact period. The baseline in computing any impulse response is the conditional expectation of y_t in the absence of a perturbation δ (e.g., Potter 2000, p. 1430). In other words, the baseline is what we would have expected Y_t to be in the absence of a perturbation, possibly conditional on the history of the data. For example, consider the expectation $E_{t-1}(.)$ conditionally on the history up to time $t-1$. We have that

$$E_{t-1}(y_t) = 0.5y_{t-1} + 0.3x_{t-1} - 0.4E_{t-1}(f(x_t)) - 0.3f(x_{t-1})$$

where the predetermined values are known and we imposed $E_{t-1}(\varepsilon_{1t}) = E_{t-1}(\varepsilon_{2t}) = 0$. This expectation can only be evaluated by integrating $f(x_t)$ over all possible realizations of x_t, as in Definition 1. In contrast, Definition 2 evaluates this expression as $f(E(x_t)) = f(0)$. By Jensen's inequality, this will not yield the desired baseline for computing the population impulse response to a shock of magnitude δ because $E(f(x_t))$ is not $f(E(x_t))$. Thus, we mainly work with Definition 1 throughout this paper. It should be noted, however, that our approach is designed to accommodate alternative models in which the potential outcomes are 0 or δ, for example, as illustrated in the last empirical example.

3.2 Examples

We now illustrate the implications of the potential outcome framework in the context of the four stylized models in Examples 2.1-2.4. In linear models, the potential outcome $Y_{t+h}(e) = y_{t+h}(e)$ is linear in e for all horizons. This can be easily seen in Example 2.1, where the potential outcome associated with any fixed value e of ε_{1t} can be written as

$$y_{t+h}(e) = \beta\gamma^h e + \beta \sum_{k=0,k\neq h}^{\infty} \gamma^k \varepsilon_{1t+h-k} + \sum_{k=0}^{\infty} \gamma^k \varepsilon_{2t+h-k}, \text{ for any } e \in E,$$

where E is the support of ε_{1t}. It follows that

$$ARF_h(\delta) \equiv E\left(y_{t+h}(\varepsilon_{1t}+\delta) - y_{t+h}(\varepsilon_{1t})\right) = \left(\beta\gamma^h\right)\delta.$$

Clearly, for a linear model, the average response function is a linear function of the size of the shock δ and, in fact, all previous definitions of the impulse response function will coincide.

In contrast, for the model with nonlinearly transformed regressors in Example 2.2, the potential outcomes model is nonlinear in e for all h because of the nonlinearity of $f(e)$. In this case, the potential outcome is given by

$$y_{t+h}(e) = \theta_h e + \gamma_h f(e) + v_{t+h},$$

and

$$ARF_h(\delta) = \theta_h\delta + \gamma_h E\left[f(\varepsilon_{1t}+\delta) - f(\varepsilon_{1t})\right]. \tag{2}$$

It is easy to see that in this case the average response function depends nonlinearly on δ and on ε_{1t} through the function f which may or may not be known. Gonçalves et al. (2021) propose a plug-in estimator for ARF in general parametric models with nonlinearly transformed regressors. Section 5.1 explores an alternative nonparametric estimator of (2), which allows the researcher to be agnostic about the functional form of f.

State-dependent models such as Example 2.3 are discussed at length in Gonçalves et al. (2024). When the state is determined exogenously with respect to the system, the potential outcome is linear in e and the CAR can be estimated using local projections. However, when the state is endogeneous with respect to y_{it} and/or x_t, as is typically the case in practice, the potential outcome is only linear in e on impact, but not at longer horizons. For any response horizon $h > 0$, e enters the potential outcomes in a complicated nonlinear form, due to the effect of ε_{1t} on the states between t and $t + h$. In particular, Gonçalves et al. (2024) show that for a simplified version of Example 2.3 where $S_t = \eta(\varepsilon_{1t}) \equiv 1(\varepsilon_{1t} > c)$,

$$y_{t+1}(e) = \gamma(e)\beta_{t-1}e + V_{t+1}(e),$$

where $\gamma(e) = \gamma_R + (\gamma_E - \gamma_R)\eta(e)$, with $\eta(e) = 1(e > c)$. The fact that S_t depends on ε_{1t} implies that the counterfactual value of γ_t when $\varepsilon_{1t} = e$ is $\gamma(e)$, which introduces a nonlinearity in $y_{t+1}(e)$. This in turn implies that the conditional impulse response function $CAR_1(\delta, s) \equiv E\left(y_{t+1}\left(\varepsilon_{1t} + \delta\right) - y_{t+1}\left(\varepsilon_{1t}\right)|S_t = s\right)$ is a complicated nonlinear function which depends on the state $s = 0, 1$.

Finally, in models with nonlinear interactions between the shock and the state, where the state is a continuous variable, such as Example 2.4, we can show that the potential outcome is nonlinear in e for $h \geq 1$. For $h = 0$ the potential outcome

$$y_t(e) = \beta_{21}e + \beta_{23}r_t + \alpha_{21}er_t + \gamma_{21}y_{t-1} + \varepsilon_{2t}$$

is linear in e and a linear local projection of y_t on x_t, r_t, $x_t r_t$ and y_{t-1} recovers the impact effect, conditionally on $r_t = r$. However, this is no longer true at $h > 0$ if r_t depends nonlinearly on past values of x_t. In particular, if $r_t = f(x_{t-1}) + \varepsilon_{3t}$ and f is nonlinear, as we assume in Example 2.4, a linear LP does not recover $CAR_h(\delta, r) \equiv E\left(y_{t+1}\left(\varepsilon_{1t} + \delta\right) - y_{t+1}\left(\varepsilon_{1t}\right)|r_t = r\right)$ for any $h > 0$ because the potential outcome $y_{t+h}(e)$ is nonlinear in e. For example, for $h = 1$

$$y_{t+1}(e) = \beta_{21}\varepsilon_{1t+1} + \beta_{23}r_{t+1}(e) + \alpha_{21}\varepsilon_{1t+1}r_{t+1}(e) + \gamma_{21}y_t(e) + \varepsilon_{2t+1},$$

where

$$r_{t+1}(e) = f(e) + \varepsilon_{3t+1}.$$

The fact that $r_{t+1}(e)$ may depend nonlinearly on e introduces a nonlinearity in $y_{t+1}(e)$. For $h > 1$

$$y_{t+h}(e) = \gamma_{21}y_{t+h-1}(e) + \xi_{t+h}$$

where ξ_{t+h} does not depend on e under the assumption that x_t is i.i.d. As shown in the Appendix, in this setting

$$\begin{aligned}
CAR_0(\delta, r) &= (\beta_{21} + \alpha_{21}r)\delta \\
CAR_1(\delta, r) &= \beta_{23}E[f(\varepsilon_{1t} + \delta) - f(\varepsilon_{1t})] + \gamma_{21}CAR_0(\delta, r) \\
CAR_h(\delta, r) &= \gamma_{21}CAR_{h-1}(\delta, r) \text{ for } h > 1.
\end{aligned}$$

Similarly to the state-dependent model, a structural shock e in a model with interaction terms between the shock and the conditioning variable will have a direct effect on y_t through β_{21} and α_{21} and an indirect effect through β_{23} due to the endogeneity of the control variable r_t.

4 Identification

As expected from the literature on treatment effects, in order to give a causal interpretation to estimands involving observables, some form of conditional independence between potential outcomes $Y_{t+h}(e)$ and ε_{1t} needs to hold. It can be shown that, under our assumptions, ε_{1t} in model (1) is independent of $\{Y_{t+h}(e),\ e \in E\}$ (see Lemma A.1 in Gonçalves et al. (2024)). This result is instrumental in establishing identification. It is equivalent to a conditional independence assumption.

The following proposition summarizes our identification results.

Proposition 4.1 *Let $z_t = (x_t, Y_t')'$ be defined by (1) and assume that $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{nt})'$ is i.i.d. $(0, \Sigma)$, where $\Sigma = diag\left(\sigma_i^2\right)$. Then,*

(i) *$ARF_h(\delta) = E\left(g_h(\varepsilon_{1t} + \delta) - g_h(\varepsilon_{1t})\right)$, where $g_h(e) \equiv E\left(Y_{t+h}|\varepsilon_{1t} = e\right)$.*

(ii) *$CAR_h(\delta, \omega) = E\left(g_h(\varepsilon_{1t} + \delta, \omega) - g_h(\varepsilon_{1t}, \omega)|\Omega_t = \omega\right)$, where $g_h(e, \omega) \equiv E\left(Y_{t+h}|\varepsilon_{1t} = e, \Omega_t = \omega\right)$.*

The proof of Proposition 4.1 follows easily from Lemma A.1 of Gonçalves et al. (2024). Suppose that we observe a sample $\{Y_t, \varepsilon_{1t}\}$. For any fixed e,

$$
\begin{aligned}
& E\left[Y_{t+h}(e+\delta) - Y_{t+h}(e)\right] \\
=& E\left[Y_{t+h}(e+\delta)|\varepsilon_{1t} = e+\delta\right] - E\left[Y_{t+h}(e)|\varepsilon_{1t} = e\right] \\
=& E\left[Y_{t+h}|\varepsilon_{1t} = e+\delta\right] - E\left[Y_{t+h}|\varepsilon_{1t} = e\right] \\
\equiv& g_h(e+\delta) - g_h(e),
\end{aligned}
$$

where the first equality follows by the independence between the potential outcomes $Y_{t+h}(e)$ and ε_{1t}, and the second equality follows because $Y_{t+h}(e) = Y_{t+h}$ when $\varepsilon_{1t} = e$ and $Y_{t+h}(e+\delta) = Y_{t+h}$ when $\varepsilon_{1t} = e + \delta$. It follows that

$$
ARF_h(\delta) \equiv E\left(Y_{t+h}(\varepsilon_{1t}+\delta) - Y_{t+h}(\varepsilon_{1t})\right) = E\left(g_h(\varepsilon_{1t}+\delta) - g_h(\varepsilon_{1t})\right).
$$

The identification of $CAR(\delta, \omega)$ proceeds in a similar manner. Suppose that we observe a sample $\{Y_t, \varepsilon_{1t}, \Omega_t\}$. For any fixed e and ω, define $g_h(e, \omega) = E(Y_{t+h}|\varepsilon_{1t} = e, \Omega_t = \omega)$. We can write

$$\begin{aligned}
&E\left[Y_{t+h}(e+\delta)-Y_{t+h}(e) \mid \Omega_t=\omega\right] \\
=&E\left[Y_{t+h}(e+\delta) \mid \varepsilon_{1t}=e+\delta, \Omega_t=\omega\right]-E\left[Y_{t+h}(e) \mid \varepsilon_{1t}=e, \Omega_t=\omega\right] \\
=&E\left[Y_{t+h} \mid \varepsilon_{1t}=e+\delta, \Omega_t=\omega\right]-E\left[Y_{t+h} \mid \varepsilon_{1t}=e, \Omega_t=\omega\right].
\end{aligned}$$

The conditional independence between $Y_{t+h}(e)$ and ε_{1t}, given Ω_t, justifies the first equality above. This condition holds because $Y_{t+h}(e)$ depends on e and U_{t+h}, where U_{t+h} is independent of ε_{1t} by construction. Thus, the conditional independence assumption of potential outcomes and the shock of interest holds for any choice of Ω_t. It follows that

$$CAR_h(\delta,\omega)=E\left(g_h\left(\varepsilon_{1t}+\delta,\omega\right)-g_h\left(\varepsilon_{1t},\omega\right)|\Omega_t=\omega\right). \tag{3}$$

The conditional expectation in (3) simplifies to $CAR_h(\delta,\omega)=E\left(g_h\left(\varepsilon_{1t}+\delta\right)-g_h\left(\varepsilon_{1t}\right)\right)$ whenever ε_{1t} is independent of Ω_t. This occurs in Example 2.3 where $\Omega_t=S_{t-1}$ if S_{t-1} is determined based on $\mathbf{z}_{t-1}$, including past values of ε_{1t}. This is also true in Example 2.4 where $\Omega_t=r_t$ and r_t is a function of z_{t-1}. Note that this would no longer be true if r_t was a function of z_t (and hence of ε_{1t}).

These results suggest that it is feasible to use a nonparametric approach to estimating $ARF_h(\delta)$ and $CAR_h(\delta,\omega)$, as discussed in the next section.

5 Estimation

5.1 Nonparametric estimators of $ARF_h(\delta)$ and $CAR_h(\delta,\omega)$

Having defined the impulse response functions of interest and having derived them in our stylized examples, the next step is to discuss the proposed estimation method. Our approach is based on Proposition 4.1.

Throughout this section, we assume without loss of generality that Y_{t+h} is a scalar random variable so that we may write $Y_{t+h}=y_{t+h}$ (otherwise, the results that follow apply to $y_{i,t+h}$, a typical i^{th} element of Y_{t+h}). We also assume for simplicity that we observe the shock of interest ε_{1t}, which corresponds to the empirically relevant case when ε_{1t} is identified by a narrative approach. However, we discuss in Section 5.2 how to extend our estimation method to the more general case when ε_{1t} is estimated in a preliminary step.

Consider first $ARF_h\left(\delta\right)\equiv E\left(y_{t+h}\left(\varepsilon_{1t}+\delta\right)-y_{t+h}\left(\varepsilon_{1t}\right)\right)$. Given Proposition 4.1, we can write $ARF_h\left(\delta\right)=E\left(g_h\left(\varepsilon_{1t}+\delta\right)-g_h\left(\varepsilon_{1t}\right)\right)$, where $g_h\left(e\right)\equiv E\left(y_{t+h}|\varepsilon_{1t}=e\right)$ is the conditional expectation of y_{t+h}, given $\varepsilon_{1t}=e$. This result suggests the following estimator of $ARF_h\left(\delta\right)$.

Algorithm 5.1 (Unconditional Average Response) *Given a sample* $\{y_t,\varepsilon_{1t},:t=1,\ldots,T\}$,

1. *Obtain a nonparametric estimator* $\hat{g}_h(e)$ *of* $g_h(e) \equiv E(y_{t+h}|\varepsilon_{1t} = e)$.

2. *Estimate* $ARF_h(\delta)$ *as*

$$\widehat{ARF}_h(\delta) = \frac{1}{T}\sum_{t=1}^{T}\left(\hat{g}_h(\varepsilon_{1t} + \delta) - \hat{g}_h(\varepsilon_{1t})\right).$$

The estimator $\widehat{ARF}_h(\delta)$ is in fact a semiparametric two-step estimator of $ARF_h(\delta)$, where the first-step is based on nonparametric regression. Specifically, we can view $ARF_h(\delta)$ as the solution of a population moment condition given by

$$E\left[m\left(\varepsilon_{1t}, ARF_h(\delta), g_h\right)\right] = 0,$$

where $m(\varepsilon_{1t}, ARF_h(\delta), g_h) = ARF_h(\delta) - [g_h(\varepsilon_{1t} + \delta) - g_h(\varepsilon_{1t})]$, showing that the moment condition is linear in the parameter of interest $ARF_h(\delta)$ and the conditional mean function g_h. The estimator $\widehat{ARF}_h(\delta)$ is the solution of the empirical moment condition,

$$\frac{1}{T}\sum_{t=1}^{T} m\left(\varepsilon_{1t}, \widehat{ARF}_h(\delta), \hat{g}_h\right) = 0,$$

where $\hat{g}_h$ is a first-step estimate of g_h. Hence, $\widehat{ARF}_h(\delta)$ is a two-step M-estimator where the first step is a nonparametric regression.

The asymptotic properties of $\widehat{ARF}_h(\delta)$ can be derived using existing results in the semiparametrics literature (see e.g. Newey and McFadden (1994)). The following result provides a set of high-level conditions on g_h and $\hat{g}_h$ under which $\widehat{ARF}_h(\delta)$ is consistent for $ARF_h(\delta)$ as $T \to \infty$, for fixed δ and h.

Proposition 5.1 *Consider a bivariate version of model (1) with* $x_t = \varepsilon_{1t}$ *where* $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})'$ *is i.i.d.* $(0, \Sigma)$, *with* $\Sigma = diag(\sigma_i^2)$. *If (i)* $E|g_h(\varepsilon_{1t})| < \infty$ *and* $E|g_h(\varepsilon_{1t} + \delta)| < \infty$, *and (ii)* $\sup_{t=1,\ldots,T}|(\hat{g}_h(\varepsilon_{1t} + \delta) - \hat{g}_h(\varepsilon_{1t})) - (g_h(\varepsilon_{1t} + \delta) - g_h(\varepsilon_{1t}))| \to_p 0$, *then* $\widehat{ARF}_h(\delta) - ARF_h(\delta) \to_p 0$ *as* $T \to \infty$.

The proof of Proposition 5.1 follows easily by noting that we can decompose the difference $\widehat{ARF}_h(\delta) - ARF_h(\delta)$ as the sum of two averages,

$$\frac{1}{T}\sum_{t=1}^{T}\left[\hat{g}_h(\varepsilon_{1t} + \delta) - \hat{g}_h(\varepsilon_{1t})\right] - \left[g_h(\varepsilon_{1t} + \delta) - g_h(\varepsilon_{1t})\right] \tag{4}$$

and

$$\frac{1}{T}\sum_{t=1}^{T}\left[g_h(\varepsilon_{1t} + \delta) - g_h(\varepsilon_{1t})\right] - E\left[g_h(\varepsilon_{1t} + \delta) - g_h(\varepsilon_{1t})\right]. \tag{5}$$

Since ε_{1t} is i.i.d., (5) converges to zero in probability by a law of large numbers provided $E\left|g_h\left(\varepsilon_{1t}+\delta\right)-g_h\left(\varepsilon_{1t}\right)\right|<\infty$, which is implied by condition (i) in Proposition 5.1. Condition (ii) implies that (4) converges to zero in probability, yielding the consistency of $\widehat{ARF}_h(\delta)$ towards $ARF_h(\delta)$. This high-level condition is implied by the uniform convergence of $\hat{g}_h\left(e\right)-g_h\left(e\right)$ to zero over $e\in E$, when $E=\mathbb{R}$. Allowing for unbounded support is important here because our estimator involves evaluating $\hat{g}_h$ at ε_{1t} and $\varepsilon_{1t}+\delta$. Hence, uniform convergence of $\hat{g}_h\left(e\right)-g_h\left(e\right)$ over a bounded set E is not enough as the probability of $g_h\left(\varepsilon_{1t}+\delta\right)\notin E$ will be strictly positive no matter how large T is.

Providing primitive conditions for uniform convergence of $\hat{g}_h(e)$ over $e\in\mathbb{R}$ in our setting appears challenging and best left for future research. Most of the nonparametrics literature assumes bounded regressors, with a few exceptions. One is Hansen (2008), who derives uniform convergence rates for local constant and local linear regression estimators under the assumption that the data are stationary strong mixing time series without assuming a bounded support, see in particular his Theorems 8 and 10. In our notation, an application of these results would translate to bounds on $\sup_{|e|\leq c_T}\left|\hat{g}_h\left(e\right)-g_h\left(e\right)\right|$, where c_T is diverging slowly to infinity. Whether these results extend to uniform convergence over unrestricted Euclidean spaces (i.e., with $c_T=+\infty$) is unclear to us. Similarly, Chen and Christensen (2015) derive uniform convergence rates for general linear sieve estimators allowing for weakly dependent time series and potentially unbounded support. However, to obtain more refined (optimal) bounds for splines and wavelets series estimators, they assume bounded regressors (see e.g., their Theorem 2.1). More recently, Ballerin (2024) shows the consistency of a semiparametric sieve estimator of an average impulse response function for a special case of our model (1).[3] Ballerin (2014) assumes bounded support on the data, but changes the nature of the structural shock of interest to ensure that its values are always inside the bounded interval. Hence, his and our definitions of the impulse response function differ.

Next, we propose a two-step estimator of the conditional average response function. Given Proposition 4.1, we can write $CAR_h\left(\delta,\omega\right)=E\left(g_h\left(\varepsilon_{1t}+\delta,\omega\right)-g_h\left(\varepsilon_{1t},\omega\right)|\Omega_t=\omega\right)$, where $g_h\left(e,\omega\right)\equiv E\left(y_{t+h}|\varepsilon_{1t}=e,\Omega_t=\omega\right)$ is the conditional expectation of y_{t+h}, given $\varepsilon_{1t}=e$ and $\Omega_t=\omega$.

The following algorithm describes our estimator of $CAR_h\left(\delta,\omega\right)$ when Ω_t and ε_{1t} are mutually independent. This assumption holds whenever we choose the conditioning set Ω_t as a function of exogenous variables or lagged dependent variables variables $\mathbf{z}_{t-1}$, as is often the case in applications. One example is Example 2.3, where $\Omega_t=S_{t-1}$ and S_{t-1} is a function of $\mathbf{z}_{t-1}$. The independence between Ω_t and ε_{1t} also holds in Example 2.4, where $\Omega_t=r_t$ and $r_t=f(x_{t-1})+\varepsilon_{3t}$ and ε_{1t} and ε_{3t} are independent random shocks.

[3] In particular, Ballerin (2024) assumes that the function ψ_i in $y_{it}=\psi_i\left(x_t,Y_{-i,t},\mathbf{z}_{t-1},\varepsilon_{it}\right)$ is linear in $Y_{-i,t},\mathbf{y}_{t-1}$ and ε_{it}, only allowing for nonlinearity in x_t and its lags. A special case of this model is Example 2.2. Ballerin (2024)'s semiparametric estimator exploits the linearity and additivity of the function ψ_i and hence differs from our estimator.

Algorithm 5.2 (Conditional Average Response) *Given a sample* $\{y_t, \varepsilon_{1t}, \Omega_t, : t = 1, \ldots, T\}$,

1. *Obtain a nonparametric estimator* $\hat{g}_h(e, \omega)$ *of* $g_h(e, \omega) \equiv E(y_{t+h}|\varepsilon_{1t} = e, \Omega_t = \omega)$.

2. *Estimate* $CAR_h(\delta, \omega)$ *as*

$$\widehat{CAR}_h(\delta, \omega) = \frac{1}{T}\sum_{t=1}^{T}\left(\hat{g}_h(\varepsilon_{1t} + \delta, \omega) - \hat{g}_h(\varepsilon_{1t}, \omega)\right).$$

As in Algorithm 5.1, any nonparametric approach can be used to estimate $g_h(e, \omega)$. We will provide specific examples in Section 5.3. Consistency of $\widehat{CAR}_h(\delta, \omega)$ follows under a set of high-level conditions similar to those provided in Proposition 5.1. The only difference is that the conditional expectation function is now $g_h(e, \omega)$ rather the univariate regression $g_h(e)$.

The independence assumption between Ω_t and ε_{1t} is used to simplify the second step in Algorithm 5.2 since under this assumption we can write $CAR_h(\delta, \omega) = E(g_h(\varepsilon_{1t} + \delta, \omega) - g_h(\varepsilon_{1t}, \omega))$. Thus, $CAR_h(\delta, \omega)$ can be estimated as the sample average of a nonparametric regression over the sample values of ε_{1t}, holding $\Omega_t = \omega$ fixed. This is known as a partial means estimator in the nonparametrics literature (see e.g., Newey (1994), who derives the asymptotic variance of partial means M-estimators based on kernel regressions for i.i.d. data).

Remark 1 *When* Ω_t *and* ε_{1t} *are not independent, Proposition 4.1 suggests that we replace the sample average over* ε_{1t} *in step 2 of Algorithm 5.2 with the difference of two nonparametric regressions, one where we regress* $\hat{g}_h(\varepsilon_{1t} + \delta, \omega)$ *on* Ω_t, *and another where we regress* $\hat{g}_h(\varepsilon_{1t}, \omega)$ *on* Ω_t. *The new estimator of* $CAR_h(\delta, \omega)$ *is the difference between these two nonparametric regressions evaluated at* $\Omega_t = \omega$.

5.2 More general specifications for x_t

Algorithm 5.1 assumes that we observe ε_{1t}, which corresponds to the empirically relevant case where $x_t = \varepsilon_{1t}$. For example, often x_t is a fiscal or monetary policy shock constructed using the narrative approach to identification. However, nothing in our approach prevents us from allowing for serial correlation in x_t or from replacing the assumption of exogeneity by the weaker assumption that x_t is predetermined, as discussed in Gonçalves et al. (2021, 2024). Setting $x_t = \varphi(\mathbf{z}_{t-1}) + \varepsilon_{1t}$, as we do in model (1), is consistent with this more general specifications for x_t and allows for identification of ε_{1t}. Note that unlike in Gonçalves et al. (2021, 2024), we do not assume that the functional form of φ is known.

When ε_{1t} is not observed, we can apply the previous algorithm with ε_{1t} replaced with $\hat{\varepsilon}_{1t}$, an estimate of ε_{1t} provided this structural shock is identified. In this case, the algorithm becomes a three-step method, where the first step is the estimation of ε_{1t}. When φ is a parametric nonlinear function, we can estimate ε_{1t} by nonlinear least squares. When φ is a nonparametric function, $\hat{\varepsilon}_{1t}$ can be obtained by nonparametric regression. It can be shown that broadly similar simulation results hold for such less restrictive DGPs using this three-step nonparametric local projection estimator.

5.3 Possible choices of nonparametric estimators of g_h

5.3.1 Local linear regression

In this section, we briefly review how to obtain $\hat{g}_h(e)$ and $\hat{g}_h(e,\omega)$ using a local linear kernel regression estimator.

Starting with $\hat{g}_h(e)$, the local linear (LL) estimator approximates the regression function $g_h(e) \equiv E(y_{t+h}|\varepsilon_{1t}=e)$ using a local linear regression for ε_{1t} around e. More specifically, noting that $y_{t+h} = g_h(\varepsilon_{1t}) + \varepsilon_{1t}$, the approximating model is

$$y_{t+h} = g_h(e) + g_h'(e)(\varepsilon_{1t} - e) + v_{t+h}, \tag{6}$$

where $g_h'(e) = \partial g_h(e)/\partial e$ and v_{t+h} denotes an error term. We define the LL estimator of the intercept $\alpha_h(e) \equiv g_h(e)$ and the slope parameter $\beta_h(e) \equiv g_h'(e)$ in (6) as the solution to the following minimization problem:

$$\left(\hat{g}_h(e), \hat{g}_h'(e)\right) = \arg\min_{\alpha_h,\beta_h} \sum_{t=1}^{T-h} K\left(\frac{\varepsilon_{1t}-e}{b}\right)\left(y_{t+h} - \alpha_h - \beta_h(\varepsilon_{1t}-e)\right)^2, \tag{7}$$

where K denotes a kernel function and b is a bandwidth parameter. Next we rewrite the solution to this optimization problem in closed form as follows. For fixed e, let

$$z_t(e) = \begin{pmatrix} 1 \\ \varepsilon_{1t}-e \end{pmatrix} \text{ and } \phi_h(e) \equiv \begin{pmatrix} \alpha_h(e) \\ \beta_h(e) \end{pmatrix} = \begin{pmatrix} g_h(e) \\ g_h'(e) \end{pmatrix}.$$

With this notation, the local projection model underlying the LL estimator is

$$y_{t+h} = z_t(e)'\phi_h(e) + v_{t+h}.$$

The LL estimator of $\phi_h(e)$ is the weighted least squares estimator defined as

$$\begin{aligned} \hat{\phi}_h(e) &= \left(\sum_{t=1}^{T-h} K\left(\frac{\varepsilon_{1t}-e}{b}\right) z_t(e) z_t'(e)\right)^{-1} \sum_{t=1}^{T-h} K\left(\frac{\varepsilon_{1t}-e}{b}\right) z_t(e) y_{t+h} \\ &= (Z'KZ)^{-1} Z'KY, \end{aligned}$$

where Z is the $(T-h)\times 2$ matrix with typical row given by $z_t'(e) = (1, \varepsilon_{1t}-e)$, $Y =$

$(y_{1+h},\ldots,y_T)'$ and

$$\underset{(T-h)\times(T-h)}{K} = \begin{bmatrix} K\left(\frac{\varepsilon_{1,1}-e}{b}\right) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & K\left(\frac{\varepsilon_{1,T-h}-e}{b}\right) \end{bmatrix}.$$

Given $\hat{\phi}_h(e)$, our object of interest is the intercept $\hat{\alpha}_h(e)$ as this is the estimate of $g_h(e)$.[4] The LL estimator depends on two tuning parameters, the kernel function K and the bandwidth parameter. One example of K is the Gaussian kernel given by

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right).$$

Another popular example is the Epanechnikov kernel defined as

$$K(u) = \begin{cases} \frac{3}{4\sqrt{5}}\left(1-\frac{u^2}{5}\right) & \text{if } |u| < \sqrt{5} \\ 0 & \text{otherwise.} \end{cases}$$

In the simulations reported below, we use a Gaussian kernel. The choice of bandwidth b is important. Theoretically, $b \to 0$ as $T \to \infty$, but b needs to satisfy additional conditions. In the simulations we use the Fan and Gijbels (1996) rule of thumb for a univariate kernel (i.e. Example 2.1.

Remark 2 *Given $\hat{g}_h(e)$, we implement step 2 of Algorithm 5.1 by taking the average of $\hat{g}_h(\varepsilon_{1t}+\delta) - \hat{g}_h(\varepsilon_{1t})$ over $t = 1,\ldots,T$. Note that t starts at 1 rather than $1+h$ since we observe ε_{1t} for $t = 1,\ldots,T$.*

Remark 3 *If the object of interest is the average response function of y_{t+h} with respect to ε_{1t} given by Definition 2, only the first step of Algorithm 5.1 is needed since we do not need to integrate out the randomness of ε_{1t}. Rather we fix ε_{1t} at $e+\delta$ and e, respectively. Although this is a one-step estimator, its convergence rate is slower than that of the semiparametric two-step estimator proposed in Algorithm 5.1. Averaging over ε_{1t} in the second step yields a $\sqrt{T}$-convergent estimator whereas a nonparametric estimator of the average impulse response function given in Definition 2 will result in a slower rate of convergence (equal to $\sqrt{Tb}$ where $b \to 0$ as $T \to \infty$ when using a kernel regression to estimate $g_h(e)$).*

Remark 4 *A researcher interested in estimating the conditional average impulse response function can employ a LL estimator. In the state-dependent case (Example 2.3) the nonparametric regression involves a continuous variable, the shock ε_{1t}, and a discrete variable,*

[4] An alternative approach would be to use a local polynomial estimator of degree higher than one such as a local quadratic kernel estimator (see Fan and Gijbels (1996)).

the state S_{t-1}, which can be estimated using the frequency-based approach or the mixed kernel approach described in Li and Racine (2006). In the model with interaction (Example 2.4), we use the multivariate version of the LL estimator. Having obtained an estimate of $\hat{g}(e,\omega)$, the second step of Algorithm 5.2 algorithm amounts to taking the average of $\hat{g}(e+\delta,\omega)-\hat{g}(e,\omega)$ over $t=1,\ldots,T$.

In the following simulations, Gaussian kernels are used in the LL estimation. We use a frequency-based approach for the state-dependent model and in the model with interaction between the shock and the state, we select the bandwidths following Henderson and Parameter's (2012) rule of thumb for each variable for multivariate kernels

5.3.2 Power series (polynomial) estimators

An alternative to using kernel regression to estimate $g_h(e)$ is to use a sieve estimator. The method of sieves approximates $g_h(e)$ by a sequence of basis functions that become increasingly flexible as the sample size grows. Although many different basis functions can be used, we focus here on power series. Let $P^L(\varepsilon_{1t})$ be the first L terms of a sequence of approximating polynomial functions $P^L(e)=(e,e^2,...e^L)$. Then, an estimate of $g_h(e)$ can be constructed by regressing the observed values of y_{t+h} on $P^L(\varepsilon_{1t})$. Having obtained an estimate $\hat{g}_h(e)$, we can estimate $ARF_h(\delta)$ following the steps in Algorithm 5.1.

The resulting polynomial estimator resembles the nonlinear local projections estimator proposed in Jordà (2005) in the use of a power series, but differs in two dimensions. First, the polynomial terms are functions of e and not of z_t. This matters because the number of terms to be estimated in our approach increases one for one with the order of the approximation, regardless of the dimension of z_t, making our approach more parsimonious. This mitigates the curse of dimensionality that undermines the feasibility of nonparametric estimation in larger nonlinear models. Second, we compute $ARF_h(\delta)$ by averaging over $\hat{g}_h(\varepsilon_{1t}+\delta)-\hat{g}_h(\varepsilon_{1t})$ rather than computing the marginal response.

6 Simulation Results

This section studies the small-sample and large-sample accuracy of the nonparametric LP estimator. It presents simulation results for DGPs with nonlinearly transformed regressors, state-dependent coefficients, and nonlinear interactions between the shock and the state, respectively.

6.1 Model with Nonlinearly Transformed Regressors

We focus on the case where x_t is an observed i.i.d. shock, as in the narrative approach to identification. The order of the lag polynomials in the equation for y_t is set to $p=1$ as in

Examples 2.2-2.4. Specifically, our DGP is given by:

$$\begin{aligned} x_t &= \varepsilon_{1t} \\ y_t &= 0.5y_{t-1} + 0.5x_t + 0.3x_{t-1} - 0.4f(x_t) - 0.3f(x_{t-1}) + \varepsilon_{2t}. \end{aligned}$$

We consider two alternative nonlinear regressors: $f(x_t) = \max\{x_t, 0\}$ and $f(x_t) = x_t^3$, respectively.[5]

The population innovations ε_{1t} and ε_{2t} are assumed to be i.i.d. $N(0,1)$. The sample size is $T \in \{250, 500, 1000, 2000\}$. T=250 corresponds to sligthly more than 60 years of quarterly data or 20 years of monthly data, whereas T=500 can be thought of as slightly more than 40 years of monthly data. The larger values of T are used to illustrate the convergence properties of the estimator. The number of Monte Carlo trials is 5,000.

For each draw from the DGP, we estimate the unconditional impulse response function of y_{t+h}, $h = 0, 1, \ldots, H$ to a shock in ε_{1t} of magnitude $\delta = 2$. Results for a shock of magnitude $\delta = 1$ are qualitatively similar. An important question is how to choose the kernel and the bandwidth. We use a second-order Gaussian kernel for the local linear estimator and select the bandwidth using Fan and Gijbels (1996, section 4.2) rule-of-thumb (hereafter ROT). The order of the polynomial for the preliminary polynomial regression used in ROT is set to 2.[6] We also consider a power series estimator with the order of the series chosen as $L = round(0.5 * T^{1/3})$.

Figures 2 and 3 plot the bias, variance, and RMSE for various sample sizes when $f(x_t) = \max\{x_t, 0\}$ and $f(x_t) = x_t^3$, respectively. Estimation results for the parametric plug-in estimator of Gonçalves et al. (2021) are obtained under the assumption that the researcher knows the exact functional form of the nonlinearity. Both figures confirm that the plug-in estimator is consistent when the model is correctly specified and works extremely well even for small samples. Similarly, the local linear estimator, which does not require the researcher to select the form of nonlinearity a-priori, is consistent albeit at a slower rate.

For the power series estimator, the sign of the bias changes depending on the order of the polynomial which changes with T. The bias declines more slowly than the bias of the local linear estimator. Not surprisingly, when the order of the polynomial L is large (i.e., $L = 9$ when $T = 2000$), both the bias and the variance increase, resulting in a high RMSE. In brief, for this particular functional form and method of selecting L, the local

[5]These functional forms are motivated by the empirical macroeconomics literature. $f(x_t) = \max\{x_t, 0\}$ has been used extensively in studies of asymmetries in the effects of oil price shocks and other shocks, and $f(x_t) = x_t^3$ has been used in Tenreyro and Thwaites' (2016) analysis of monetary policy shocks, while $f(x_t) = x_t^2$ has been used by Ben Zeev et al. (2020, 2023) and Forni et al. (2024) to investigate nonlinear effects of fiscal and monetary policy shocks. The simulation results for the latter specification are very similar to those for $f(x_t) = x_t^3$ and, hence, are omitted.

[6]Given the rate of convergence of power series estimators discussed in Newey (1994, 1997), an alternative would have been to select $L = T^{1/3}$. However, this would have led to the selection of higher-order polynomials and an even higher variance for large T than reported in our simulations.

linear estimator appears to converge at a faster rate. In contrast, when the DGP involves $f(x_t) = x_t^3$, the power series estimator has a smaller RMSE than the local linear estimator, as long as the number of terms in the polynomial approximation $P^L(e)$ is not too large. This is not surprising, as the power series estimator includes the population model specification as a special case.

Clearly, in small samples, it is hard to beat the plug-in estimator in terms of the RMSE when the functional form is correctly specified. However, when the researcher wants to remain agnostic about the functional form of the nonlinear regressor, the nonparametric LP estimator is a good alternative. To illustrate why this is the case, as well as the cost of model misspecification, next we provide simulation results for the case when a polynomial of third degree exists in population, $f(x_t) = -2.29x_t^2 + 5.66x_t^3$, but the researcher estimates the model using either the second power (x^2) or third power (x^3) term alone in the plug-in estimator. For reference, we also include the results for the plug-in estimator when the functional form of $f(x_t)$ is correctly specified. The selection of the bandwidth, b, for the local linear estimator and the polynomial order, L, for the power series estimator is done using the sample size-dependent rules of thumb described before.

As Figure 4 illustrates, even for relatively small shocks, $\delta = 1$ and large samples, $T = 2000$, the bias of the misspecified plug-in estimator can be large. In fact, when the researcher misses the second-order term, the bias does not decrease with the sample size, resulting in a large RMSE for the plug-in estimator that uses $f(x_t) = x_t^3$ regardless of the magnitude of the shock. When the researcher ignores the third-order term and uses a plug-in estimator that assumes $f(x_t) = x_t^2$, the bias is greater than that of the nonparametric estimator for small shocks $\delta = 1$. However, the larger variance of the nonparametric local linear estimator results in a larger RMSE for small samples. As the sample size increases, the nonparametric local linear estimator exhibits a smaller RMSE at short horizons for $\delta = 1$ and comparable RMSE for $\delta = 2$. In summary, if the researcher is uncertain about the functional form of $f(x_t)$ and prefers to remain agnostic, the nonparametric estimator that uses a local linear kernel regression provides an alternative estimator that is consistent and has reasonably good small-sample properties.

6.2 State-dependent model

Next, we turn to the state-dependent model, as in Example 2.3. The DGP is of the form

$$\begin{cases} x_t = \varepsilon_{1t} \\ y_t = \beta_{t-1}x_t + \gamma_{t-1}y_{t-1} + \varepsilon_{2t}. \end{cases} \tag{8}$$

For simplicity and because this corresponds to the most common specification used in empirical work, we assume that there are only two states (e.g. expansions and recessions).

Thus, we have that $\beta_{t-1} = \beta_E S_{t-1} + \beta_R (1 - S_{t-1})$ and similarly for γ_{t-1}, where $S_{t-1} = 1(y_{t-1} > 0)$ is a binary stationary time series that takes the value $\omega = 1$ if the economy is in expansion, where $y_{t-1} > 0$, and $\omega = 0$ otherwise. For expository purposes, we set $\beta_E = 2.5$, $\beta_R = 3.5$, $\gamma_E = 0.9$, $\gamma_R = -0.1$ in the DGP. As previously, ε_{1t} and ε_{2t} are mutually independent i.i.d. $N(0,1)$ structural innovations. With these parameter values, we have about 70% of the data in expansion, which mimics reality.

Given that S_{t-1} is discrete while ε_{1t} is continuous, we estimate the local linear (LL) kernel regression using the frequency-based approach (see Li and Racine, 2006).[7] As before, we use a second-order Gaussian kernel for the continuous variable. The bandwidth for this variable is selected using the ROT bandwidth, as previously described, for each state. We also report results for the case when the bandwidth is set to twice the ROT bandwidth to illustrate the bias-variance trade-off. We compare this nonparametric LP estimator to the conventional state-dependent LP estimator commonly used by practitioners for two sample sizes, $T = 250$ and $T = 1000$. The latter estimator does not recover the average response function when the state is endogenous, except on impact (see Gonçalves et al. (2024)).

Figure 5 reports the population responses, the mean estimated response, bias, and the RMSE for horizon $h = 0, \ldots, 7$ when the magnitude of the shock is $\delta = 2$, which is equivalent to a shock of two standard deviations. We employ 5,000 Monte Carlo draws in the simulations. As the figure illustrates, for $\delta = 2$ the nonparametric LP estimator comes close to recovering the population CAR, while the state-dependent LP estimator does not. As the figure illustrates, the bias of the LP does not disappear as the sample size increases. The results for $\delta = 1$ (not reported here, but available from the authors upon request) are similar except that the differences between the state-dependent LP and the nonparametric LP estimator are smaller. That the difference between the two estimates is smaller for smaller shocks is to be expected as the state-dependent LP estimates a marginal response; hence, the two estimates will get closer, as the size of the shock decreases. Conversely, as the size of the shock increases, the asymptotic bias of the state-dependent LP estimator increases. As is typical for nonparametric estimators, in finite samples there is a trade-off between bias and variance. This trade-off depends on the value of the DGP parameters. For larger samples and with appropriate bandwidth selection, the RMSE of the nonparametric LP declines.

[7] Alternatively, the researcher could use the method proposed by Racine and Li (2004) for mixed data with the Aitchinson and Aitken (1976) kernel for the discrete variable. Since we only have two states and enough observations in each state, we follow the simpler frequency-based approach.

6.3 Model Interacting Shock and State

Finally, we provide simulation results for a model in which the shock interacts with a continuous state variable and the latter is endogenous. The DGP takes the form:

$$\begin{cases} x_t = \varepsilon_{1t} \\ y_t = \beta_{21} x_t + \beta_{23} r_t + \alpha_{21} x_t r_t + \gamma_{21} y_{t-1} + \varepsilon_{2t} \\ r_t = f(x_{t-1}) + \varepsilon_{3t}, \end{cases}$$

with $\beta_{21} = 0.5$, $\beta_{23} = 0.2$, $\alpha_{21} = 0.4$ and $\gamma_{21} = 0.6$. We consider three nonlinear functional forms: $f(x_t) = max(0, x_t)$ in Figure 6, $f(x_t) = x_t^2$ in Figure 7, and $f(x_t) = x_t^3$ in Figure 8. When computing the $CAR(\delta, r)$ we set $r = \bar{r}$, the sample average of r_t, but one could condition on any value of interest. As in the previous sections, the nonparametric estimator of $g(e, r)$ is obtained using a local linear (LL) kernel regression. We employ a second-order Gaussian kernel and present results for two bandwidths that illustrate the bias-variance trade-off faced by the researcher. The plots show the mean estimate of the responses, the bias, and the RMSE for two possible bandwidths. The blue line (LL1) illustrates the case when the normal reference rule-of-thumb bandwidth for a second-order bivariate Gaussian kernel is used in the nonparametric estimation of $g(e, r)$. Following Henderson and Parmeter (2012), we set the rule-of-thumb constant to one, and let the bandwidth for x_t and r_t equal to $b^x_{ROT} = \hat{\sigma}_x T^{-1/6}$ and $b^r_{ROT} = \hat{\sigma}_r T^{-1/6}$, respectively.[8] Here, $\hat{\sigma}_x$ denotes the sample standard deviation of x_t and $\hat{\sigma}_r$ is defined similarly. Then, to illustrate how the choice of bandwidth affects the bias, variance, and RMSE, we report simulation results for a larger bandwidth set to twice the ROT bandwidths defined above. The simulation results for these larger bandwidths are depicted by the red lines (LL2). We compare the performance of the nonparametric LP estimator with that of the nonlinear LP estimator (dashed line) used by Caramp and Feilich (2024). This local projection is given by

$$y_{t+h} = \alpha_h + \psi_h \varepsilon_t + \beta_h r_{t-1} + \gamma_h r_{t-1} \varepsilon_t + \omega_{t+h} \tag{9}$$

and the impulse response conditional on $r_{t-1} = \bar{r}$ is computed as $\psi_h + \gamma_h \bar{r}$, and $\bar{r}$ denotes the sample average.

As expected, given the results in Gonçalves et. al (2021, 2024), the nonlinear LP response estimator displays asymptotic bias for any $h > 0$. The intuition is similar to the state-dependent model: if the conditioning variable is endogenous, a local projection as defined in (9) will not take into account how the shock will affect the conditioning variable in the future. Unless the conditioning variable is exogenous, a nonlinear LP that estimates the

[8] Note that the order of the kernel (the first non-zero moment) is $\nu = 2$ and, given that we include two variables in the estimation, $q = 2$. Hence, the normal reference rule-of thumb constant equals 1 and the optimal bandwidth for the l^{th} variable is given by $\hat{\sigma}_l T^{\frac{-1}{2\nu+q}} = \hat{\sigma}_l T^{\frac{-1}{6}}$.

conditional impulse response function as $\psi_h + \gamma_h \bar{r}$ will only consistently estimate the impact effect. The size of the asymptotic bias at higher horizons will depend on the DGP. As in the previous examples, the nonparametric LP estimator incurs a cost in terms of variance (as evidenced by the RMSE), especially for small samples due to the slow convergence rate. However, the bias of the nonlinear LP estimator can be substantial and will not disappear even in very large samples.

7 Empirical Illustration

To illustrate how estimates obtained using a commonly used nonlinear LP and our nonparametric LP estimates may differ, we apply both methods to study the role of privately-held government debt in the transmission of monetary policy shocks. A generalization of the nonlinear variant of the LP method described in (9) has recently been employed by Caramp and Feilich (2024) to test the implications of a New Keynesian model that predicts that monetary policy is less effective when an economy has a higher level of government debt. Indeed, they find empirical evidence in support of their model and thus suggest that the common textbook idea that the level of government debt does not affect the efficacy of monetary policy should be revised.

We consider a nonlinear local projection specification that closely follows the specification in Caramp and Feilich (2024). Their LP specification is given by

$$\Delta y_{t+h} = \alpha_h + \psi_h \varepsilon_t + \beta_h r_{t-1} + \gamma_h r_{t-1} \varepsilon_t + \sum_{i=1}^{I} \boldsymbol{X}_{t-i} \boldsymbol{\theta_h} + \omega_{t+h} \tag{10}$$

where y_t is an outcome variable of interest, r_t is the measure of privately-held U.S. government debt provided by Hall et al. (2018) and standardized as in Caramp and Feilich (2024), ε_t is the narrative measure of monetary policy shocks derived in Wiedland and Yang (2020),[9] $h = 0, \ldots, 36$ and $I = 12$. The outcomes of interest, y_{t+h}, comprise the industrial production index, the consumer price index for all urban consumers (CPI), the unemployment rate, and the effective federal funds rate. The control variables, $\boldsymbol{X}_{t-i}$, include lags of the shocks, the log of industrial production (IP), the log of the CPI, the log of the PPI, the unemployment rate, and the federal funds rate. The data span the period between March 1969 and December 2007. The Caramp-Feilich LP estimator of the impulse response is then given by $\beta_h + \gamma_h r_{t-1}$, where r_{t-1} is equal to a given value, e.g., the mean value of r_t. As additional controls, we also include lags of the producer price index for all commodities (PPI).

Nonparametric local projection estimates of CAR are obtained following the steps described in Algorithm 5.2. That is, the nonparametric estimate of $g(e, r)$ corresponds to a

[9] Their measure extends Romer and Romer (2004) narrative monetary policy shocks.

local linear kernel estimator with a second-order Gaussian kernel and bandwidths selected using the rule-of-thumb bandwidths described in Section 6.3. The Caramp-Feilich LP and nonparametric LP estimates of the CAR are evaluated at the mean debt and at a 'high debt' level (defined as the mean plus one standard deviation). We report cumulative responses to a 25 basis point monetary policy shock.

Consistent with Caramp and Feilich (2024), the nonparametric LP estimates in Figure 9 (solid line) indicate that the sensitivity of industrial production and unemployment to monetary policy shocks is lower when government debt is high. The nonparametric LP estimates suggest a larger effect of monetary policy on industrial production and employment than the Caramp-Feilich LP estimator (dashed line) regardless of the size of the debt. We also find a smaller reduction in the effectiveness of monetary policy when government debt is high.

There are three key reasons why the magnitudes of the Caramp-Feilich estimates differ from those of the nonparametric LP estimates. One is the fact that the nonparametric LP estimates account for the endogeneity of government debt with respect to monetary policy shocks. Another reason is the fact that we examine the effect of a non-negligible policy shock of 25 basis points. Finally, these differences may also stem from the less parametric nature of our approach.

8 Concluding Remarks

In recent years, nonlinearities in the responses of macroeconomic aggregates to shocks have received increasing attention in applied work. In this paper, we examined in depth a recently proposed nonparametric LP estimator of the conditional and unconditional nonlinear responses of an outcome variable to a directly observed identified shock, as is common in applied work. We observed that this estimator may also be adapted to allow for richer dynamics and identified shocks subject to estimation uncertainty.

The nonparametric LP estimator has four advantages. First, it provides an alternative to existing adaptations of the linear LP estimator to nonlinear settings that have been shown to be invalid in many cases of practical interest. Second, it is more parsimonious than alternative estimators based on nonlinear or nonparametric structural VAR models. Third, as illustrated in this paper, it can be adapted to a wide range of nonlinear data generating processes used in applied work. Fourth, while economic theory may suggest specific parametric nonlinear specifications, alternative theories often imply different nonlinear specifications. The proposed estimator allows users to dispense with strong assumptions about the functional form of the nonlinearity. This point is important because the current practice of reporting results for alternative parametric functional forms by construction involves relying on one or more inconsistent estimators.

We formally defined the response functions of interest within a potential outcome framework, derived the nonparametric LP estimator and showed how it can be adapted to various nonlinear contexts, discussed how to construct data-dependent nonparametric approximations, illustrated how this estimator identifies the population response function, provided high-level conditions for its consistency, and studied the accuracy of the estimator in small and large samples by Monte Carlo simulation. We demonstrated that in three commonly used nonlinear settings the proposed estimator tends to work well in reasonably large samples and is robust to nonlinearities of unknown form. We also examined how the specification of the nonlinear transformation affects the ability of the estimator to capture the size and sign asymmetries frequently discussed in applied work. An empirical illustration focused on the question of how the level of government debt changes the effectiveness of monetary policy shocks.

9 References

Aitchinson, J., Aitken, C.G.G., 1976. Multivariate binary discrimination by the kernel method. Biometrika 63, 413-420.

Angrist, J.D., Kuersteiner, G.M., 2011. Causal effects of monetary policy shocks: Semi-parametric conditional independence tests with a multinomial propensity score. Review of Economics and Statistics 93, 725-747.

Ascari, G. and Haber, T., 2022. Non-linearities, state-dependent prices and the transmission mechanism of monetary policy. Economic Journal, 132, 37-57.

Ballerin, G., 2024. Impulse response analysis of structural nonlinear time series models. University of Mannheim, manuscript.

Ben Zeev, N., Ramey, V.A. and Zubairy, S., 2023. Do government spending multipliers depend on the sign of the shock? AEA Papers and Proceedings 113, 382-387.

Bernanke. B.S., Blinder, A.S., 1992. The federal funds rate and the channels of monetary transmission. American Economic Review 82, 901-921.

Bernanke, B.S., Mihov, I., 1998. Measuring monetary policy. Quarterly Journal of Economics 113, 869-902.

Cacciatore, M., Ravenna, F., 2021. Uncertainty, wages and the business cycle. Economic Journal 131, 2797-2823.

Caramp N., Feilich, E., 2024. Monetary policy and government debt. Manuscript, Journal of Money, Credit and Banking, forthcoming.

Caravello, T.E., Martinez-Bruera, P., 2024. Disentangling sign and size nonlinearities. Manuscript, MIT.

Chen, X. and T. M. Christensen, 2015. Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions, Journal of Econometrics, 188, 447-465.

Cloyne, J., Jorda, O., Taylor, A.M., 2021. Decomposing the fiscal multiplier. NBER

Working Paper No. 26939.

Cloyne, J, Jorda, O., Taylor, A.M., 2024. State-dependent local projections: Understanding impulse response heterogeneity. Manuscript, UC Davis.

Falck, E., Hoffmann, M., Hürtgen, P., 2021. Disagreement about inflation expectations and monetary policy transmission. Journal of Monetary Economics 118, 15-31.

Fan, J., Gijbels, I., 1996. Local polynomial modelling and its applications. New York: Chapman & Hall.

Forni, M., Gambetti, L., Maffei-Faccioli, N. and Sala, L., 2024. Nonlinear transmission of financial shocks: Some new evidence. Journal of Money, Credit and Banking 56, 5-33.

Gonçalves, S., Herrera, A.M., Kilian, L,. Pesavento, E., 2021. Impulse response analysis for structural dynamic models with nonlinear regressors. Journal of Econometrics 225, 107-130.

Gonçalves, S., Herrera, A.M., Kilian, L,. Pesavento, E., 2024. State-dependent local projections. Journal of Econometrics, forthcoming.

Gourieroux, C., Lee, Q., 2023. Nonlinear impulse response functions and local projections. Manuscript, University of Toronto.

Hall, G. J., Payne, J., and Sargent, T. J., 2018. US federal debt 1776-1960: Quantities and prices. NYU Stern, Department of Economics.

Hansen, B., 2008. Uniform convergence rates for kernel regression with dependent data, Econometric Theory, 24, 726–748.

Henderson, D. J., and Parmeter, C. F., 2012. Normal reference bandwidths for the general order, multivariate kernel density derivative estimator. Statistics and Probability Letters, 82(12), 2198-2205.

Herrera, A.M., Lagalo, L.G., Wada, T., 2015. Asymmetries in the response of economic activity to oil price increases and decreases? Journal of International Money and Finance 50, 108-133.

Jordá, O., 2005. Estimation and inference of impulse responses by local projections. American Economic Review 95, 161-182.

Kilian, L., Vigfusson, R.J., 2011. Are the responses of the U.S. economy asymmetric in energy price increases and decreases? Quantitative Economics 2, 419-453.

Koop, G., Pesaran, M.H., Potter, S.M., 1996. Impulse response analysis in nonlinear multivariate models. Journal of Econometrics 74, 119-147.

Li, Q., Racine, J.S., 2004. Cross-validated local linear nonparametric regression. Statistica Sinica 14, 485-512.

Li, Q., Racine, J.S., 2006. Nonparametric Econometrics: Theory and Practice, Princeton, N.J.: Princeton University Press.

Newey, W.K., 1994. Series estimation of regression functionals. Econometric Theory 10, 1-28.

Newey, W.K., 1997. Convergence rates and asymptotic normality for series estimators. Journal of Econometrics 79, 147-168.

Newey, W.K. and D. MacFadden, 1994. Large-sample estimation and hypothesis testing. Handbook of Econometrics, Volume IV, Edited by R.F. Engle and D.L. McFadden.

Potter, S.M., 2000. Nonlinear impulse response functions. Journal of Economic Dynamics and Control 24, 1425-1446.

Racine, J.S., Li, Q. 2004, Nonparametric estimation of regression functions with both categorical and continuous data. Journal of Econometrics 119, 99-130.

Ramey, V. A., Zubairy, S., 2018. Government spending multipliers in good times and in bad: Evidence from U.S. historical data. Journal of Political Economy 126, 850-901.

Romer, C. D., and Romer, D. H., 2004. A new measure of monetary shocks: Derivation and implications. American economic review, 94(4), 1055-1084.

Tenreyro, S., Thwaites, G., 2016. Pushing on a string: U.S. monetary policy is less powerful in recessions. American Economic Journal: Macroeconomics 8, 43-74.

Wieland, J. F., and Yang, M. J., 2020. Financial dampening. Journal of Money, Credit and Banking, 52(1), 79-113.

A Appendix

A.1 CAR for Model Interacting Shock and State

Recall the model that includes an interaction between the shock of interest and a continuous state variable given by:

$$\begin{cases} x_t = \varepsilon_{1t} \\ y_t = \beta_{21}x_t + \beta_{23}r_t + \alpha_{21}x_t r_t + \gamma_{21}y_{t-1} + \varepsilon_{2t} \\ r_t = f(x_{t-1}) + \varepsilon_{3t}. \end{cases}$$

For the sake of notational simplicity we will denote the potential outcome $y_t(e)$ and $y_t(e+\delta)$ with superscripts (i.e., y_t^e and $y_t^{e+\delta}$, respectively).

First, note that

$$\begin{aligned} y_t^e &= \beta_{21}e + \beta_{23}r_t + \alpha_{21}er_t + \gamma_{21}y_{t-1} + \varepsilon_{2t} \\ y_t^{e+\delta} &= \beta_{21}(e+\delta) + \beta_{23}r_t + \alpha_{21}(e+\delta)r_t + \gamma_{21}y_{t-1} + \varepsilon_{2t} \end{aligned}$$

where r_t is known at time t since it is a function of x_{t-1}. For $h = 0$ the potential outcome is linear in e.

For $h = 0$, we have that

$$y_t^{e+\delta} - y_t^e = \beta_{21}\delta + \alpha_{21}r_t\delta,$$

thus

$$CAR_0(\delta, r) = E[y_t(\varepsilon_{1t} + \delta) - y_t(\varepsilon_{1t})|r_t = r] = (\beta_{21} + \alpha_{21}r)\delta. \tag{11}$$

For $h = 1$,

$$\begin{aligned} y_{t+1}^e &= \beta_{21}\varepsilon_{1t+1} + \beta_{23}r_{t+1}^e + \alpha_{21}\varepsilon_{1t+1}r_{t+1}^e + \gamma_{21}y_t^e + \varepsilon_{2t+1} \\ y_{t+1}^{e+\delta} &= \beta_{21}\varepsilon_{1t+1} + \beta_{23}r_{t+1}^{e+\delta} + \alpha_{21}\varepsilon_{1t+1}r_{t+1}^{e+\delta} + \gamma_{21}y_t^{e+\delta} + \varepsilon_{2t+1} \\ y_{t+1}^{e+\delta} - y_{t+1}^e &= \beta_{23}(r_{t+1}^{e+\delta} - r_{t+1}^e) + \alpha_{21}\varepsilon_{1t+1}(r_{t+1}^{e+\delta} - r_{t+1}^e) + \gamma_{21}(y_t^{e+\delta} - y_t^e) \end{aligned}$$

with

$$r_{t+1}^{e+\delta} - r_{t+1}^e = f(e+\delta) - f(e).$$

Note that for $h = 1$ the potential outcome is not linear in e. A local projection that regresses y_{t+h} on x_t, r_t and the interaction term $x_t r_t$ will not be able to recover this conditional impulse response function. It follows that

$$E[y_{t+1}(\varepsilon_{1t}+\delta)-y_{t+1}(\varepsilon_{1t})|r_t = r] = \beta_{23}E[f(\varepsilon_{1t}+\delta)-f(\varepsilon_{1t})|r_t = r]+\gamma_{21}E[y_t(\varepsilon_{1t}+\delta)-y_t(\varepsilon_{1t})|r_t = r]$$

given that $E[\varepsilon_{1t+1}(f(\varepsilon_{1t}+\delta)-f(\varepsilon_{1t}))|r_t = r] = 0$ by the law of iterated expectation and the fact that ε_{1t} is i.i.d.

Then

$$CAR_1(\delta, r) = \beta_{23}E[f(\varepsilon_{1t}+\delta)-f(\varepsilon_{1t})|r_t = r] + \gamma_{21}CAR_0(\delta, r) \tag{12}$$

where $E[f(\varepsilon_{1t}+\delta)-f(\varepsilon_{1t})|r_t = r] = E[f(\varepsilon_{1t}+\delta)-f(\varepsilon_{1t})]$ under the assumptions of our model.

For $h > 1$, we have

$$CAR_h(\delta, r) = \gamma_{21}CAR_{h-1}(\delta, r) \tag{13}$$

where we use the fact that

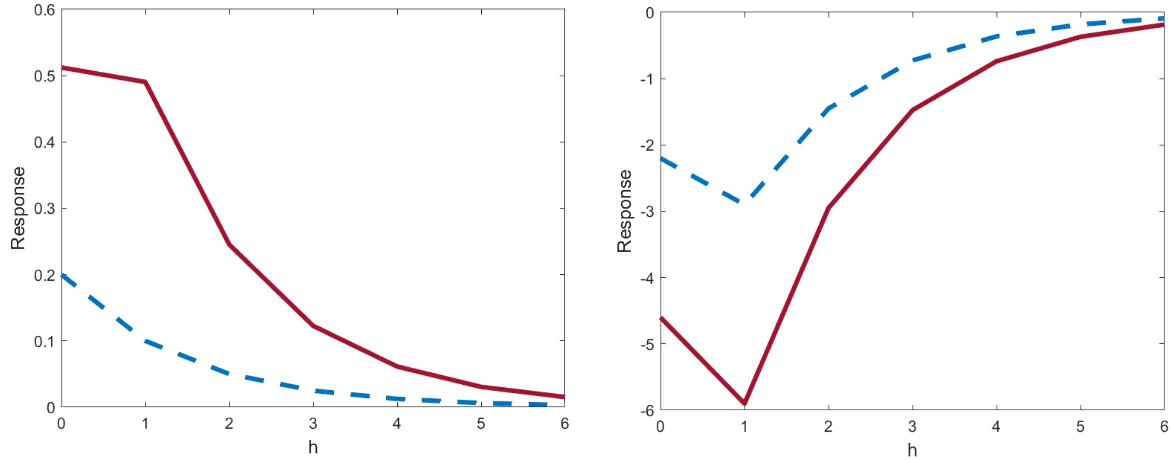
$$r^{e+\delta}_{t+h} - r^{e}_{t+h} = f(x^{e+\delta}_{t+h-1}) - f(x^{e}_{t+h-1}) = 0.$$

This follows because $x^{e+\delta}_{t+h-1} = x^{e}_{t+h-1}$ since $x_t = \varepsilon_{1t}$ and only ε_{1t} is subject to the δ shock.

In the special case where f is linear, $f(\varepsilon_{1t}+\delta)-f(\varepsilon_{1t})$ is a linear function of δ, the potential outcome is linear in e; thus, the CAR can be recovered from the usual local projection.

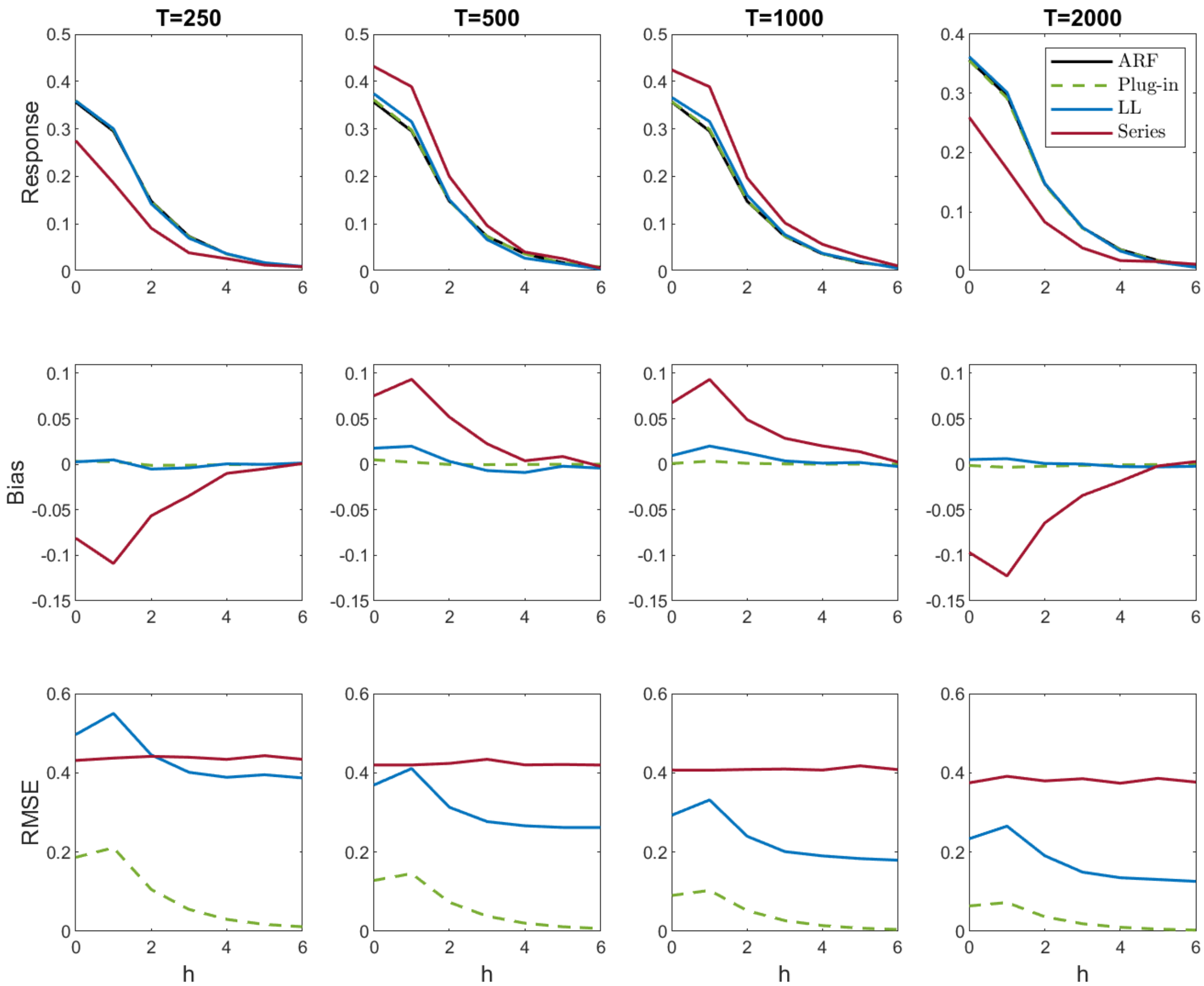
Figure 1: Alternative IRF definitions

(a) $f(x) = max(x_t, 0)$ (b) $f(x) = x_t^3$



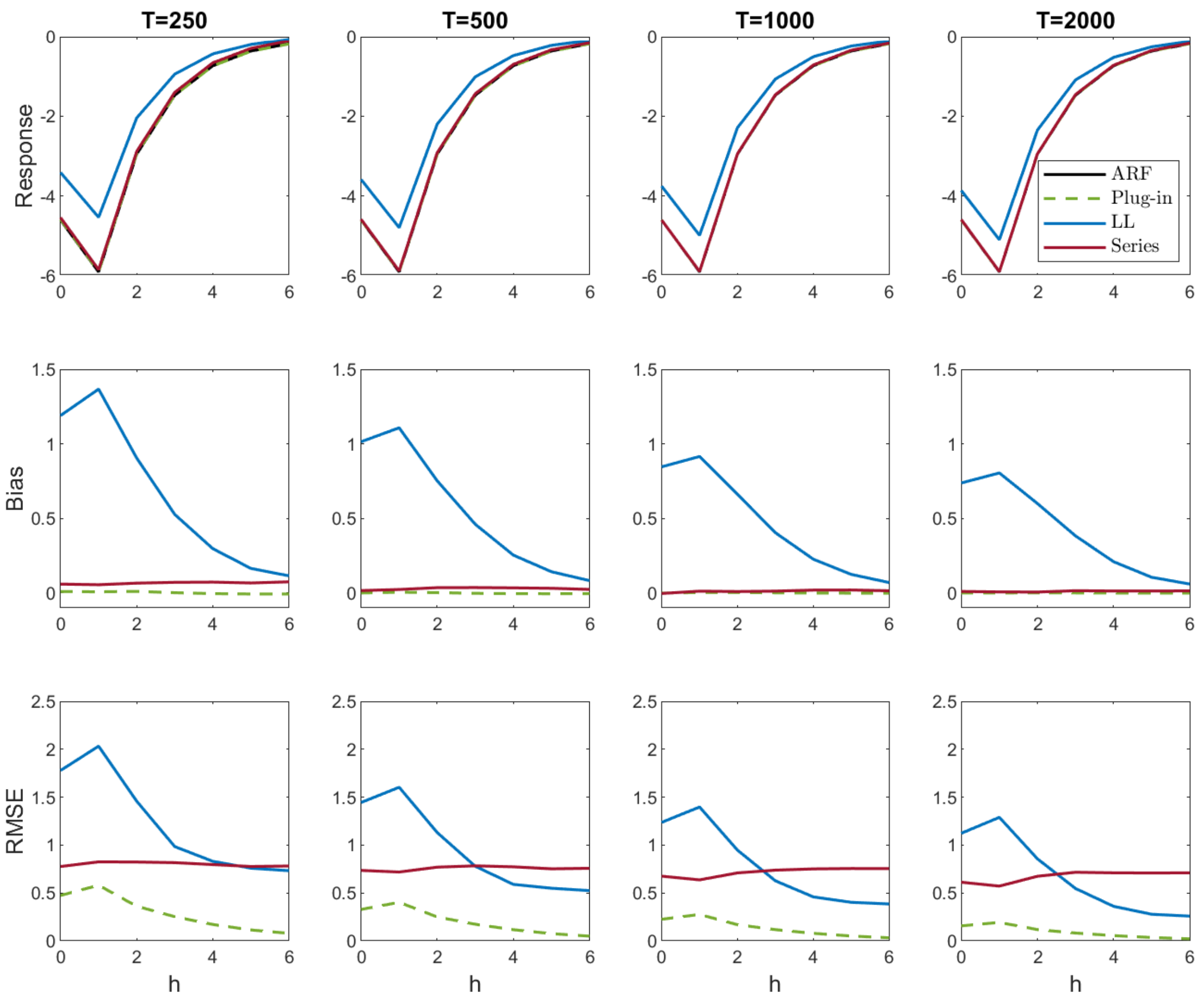
Notes: The solid red line and the dashed blue line correspond to the two definitions of the average response function ARF and ARF^* respectively, to a shock of size $\delta = 2$.

Figure 2: Nonlinear regressors model, Correctly specified, $f(x_t) = max(x_t, 0)$, $\delta = 2$



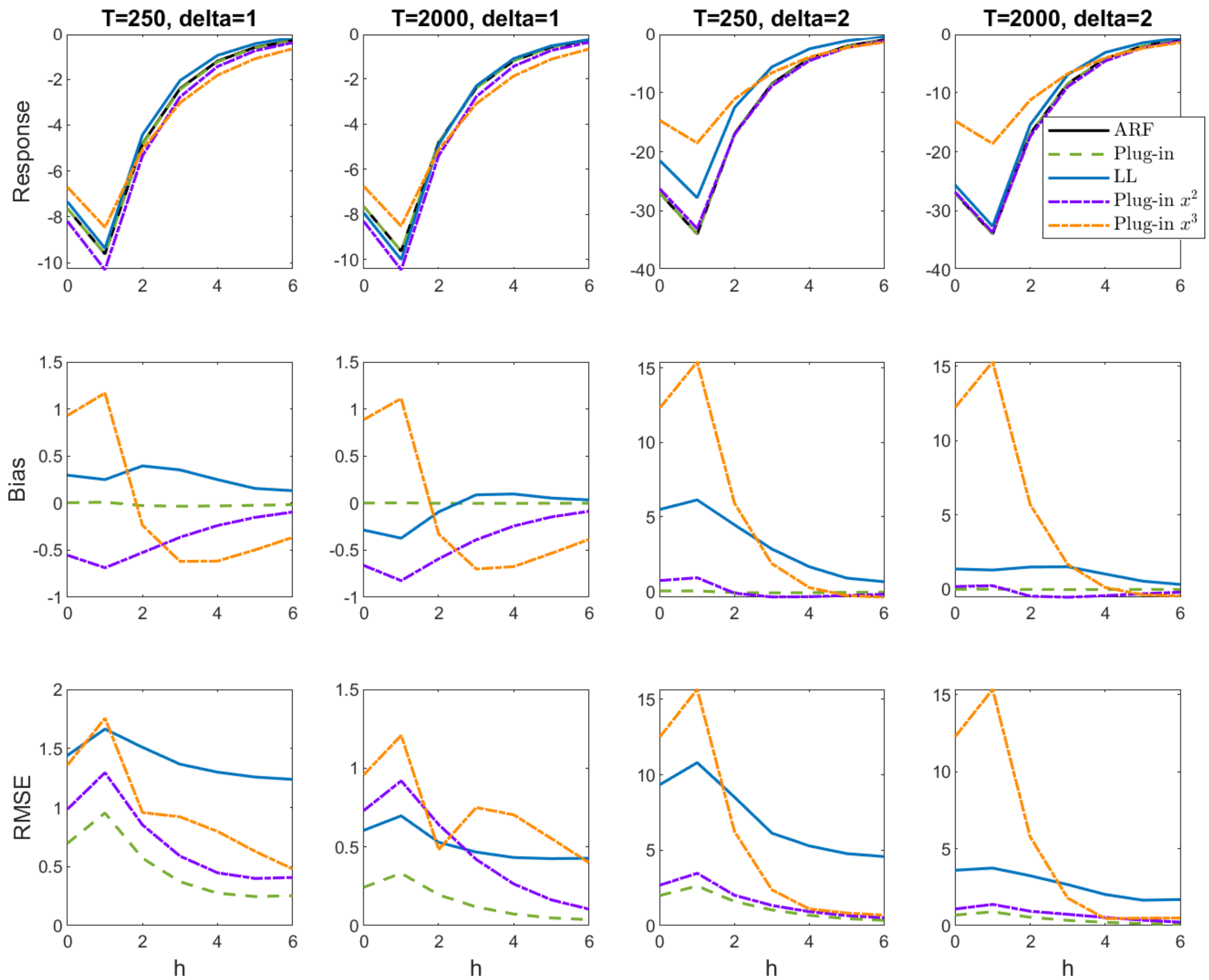
Notes: The black solid line reports the average response function, ARF, the green dashed line illustrates the plug-in estimator when the functional form of $f(x_t)$ is correctly specified, the blue line illustrates the local linear estimator, and the red line illustrates the power series estimator.

Figure 3: Nonlinear regressors model, Correctly specified, $f(x_t) = x_t^3$, $\delta = 2$



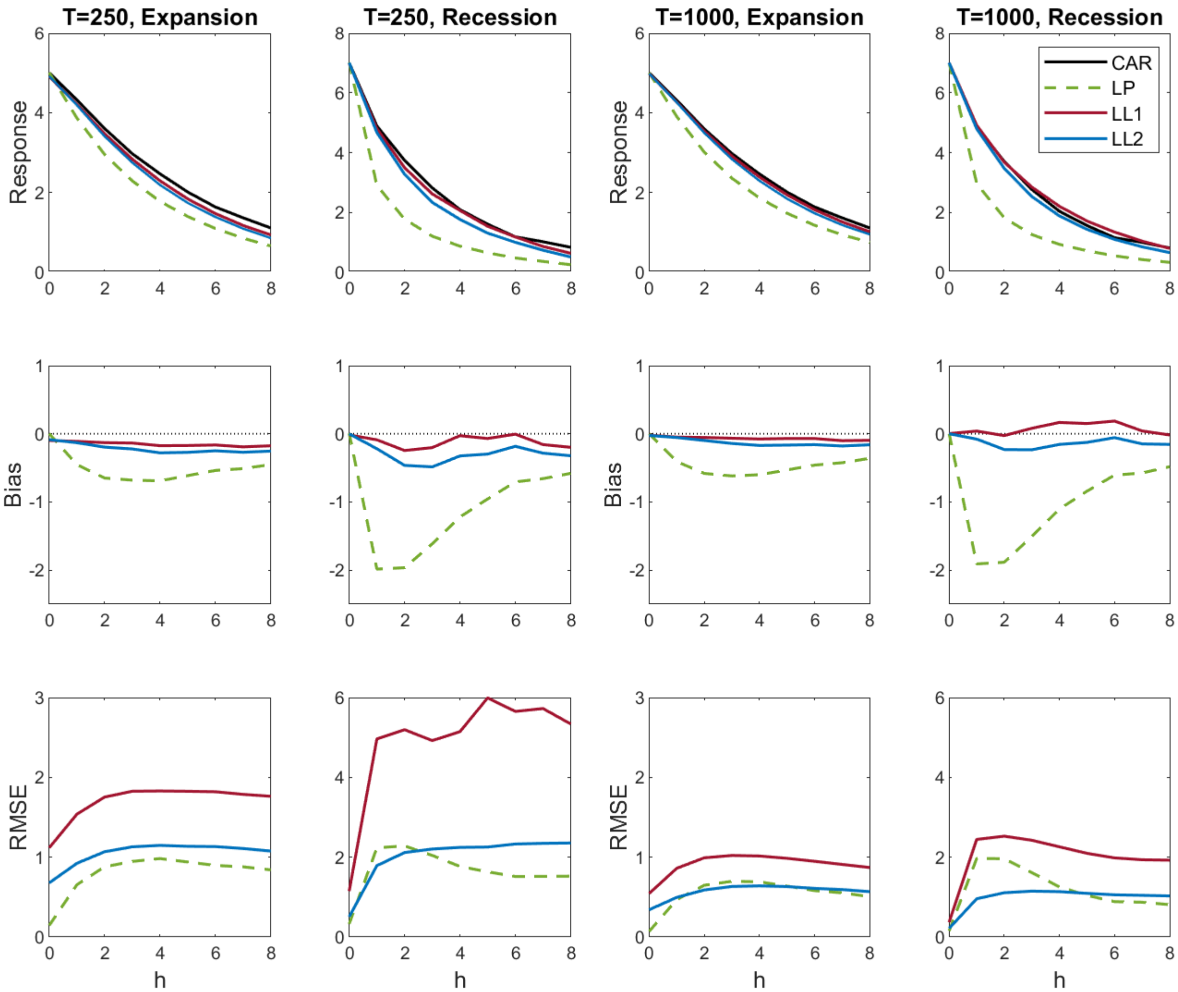
Notes: The black solid line reports the average response function, ARF, the green dashed line illustrates the plug-in estimator when the functional form of $f(x_t)$ is correctly specified, the blue line illustrates the local linear estimator, and the red line illustrates the power series estimator.

Figure 4: Misspecified nonlinear regressors model, $f(x_t) = -2.29x_t^2 + 5.66x_t^3$



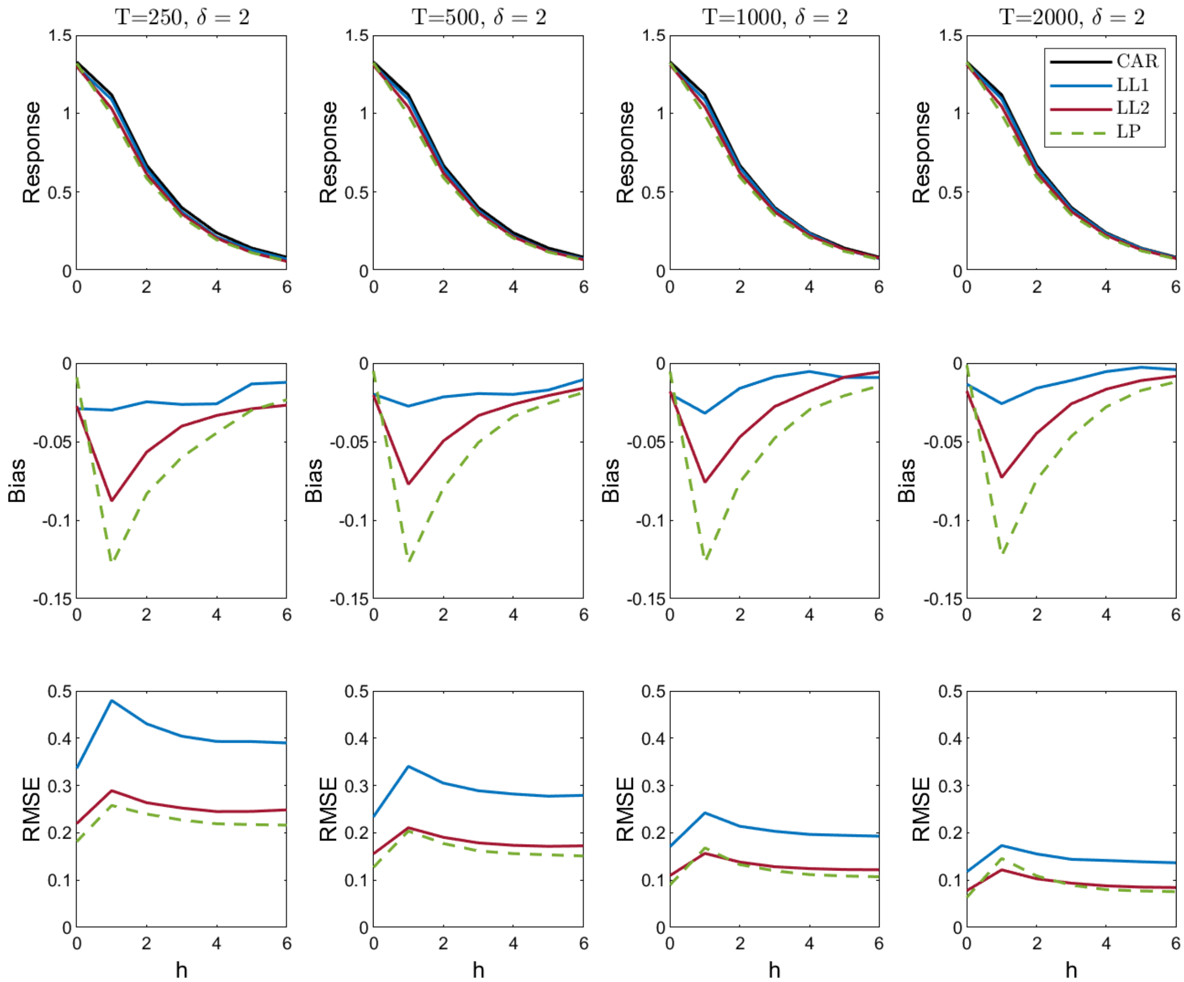
Notes: The black solid line reports the average response function, ARF, the green dashed line illustrates the plug-in estimator when the functional form of $f(x_t)$ is correctly specified, the blue line illustrates the local linear estimator, the yellow and purple lines illustrates the plug-in estimator in the presence of misspecification ($f(x_t) = x^2$ and $f(x_t) = x^3$, respectively).

Figure 5: State Dependent Model, $T = 240$, $\delta = 2$



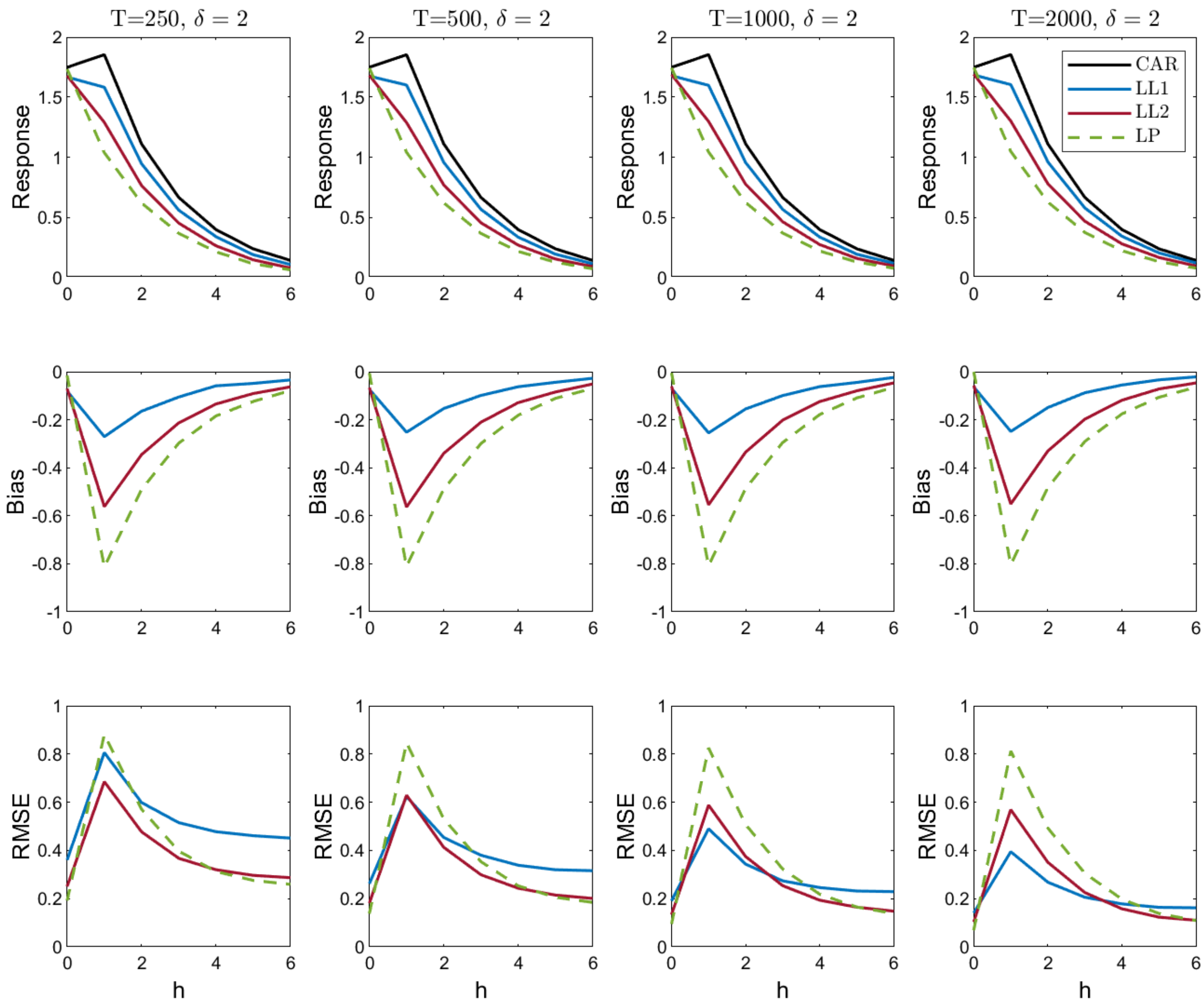
Notes: The black solid line reports the conditional average response function, CAR, the green dashed line illustrates the state-dependent LP estimator, the blue and red lines illustrate the local linear estimator with the rule-of-thumb bandwidth (LL1) and twice the rule-of-thumb bandwidth (LL2), respectively.

Figure 6: Model with Shock and State Interaction, $\delta = 2$, $f(x_t) = max(0, x_t)$



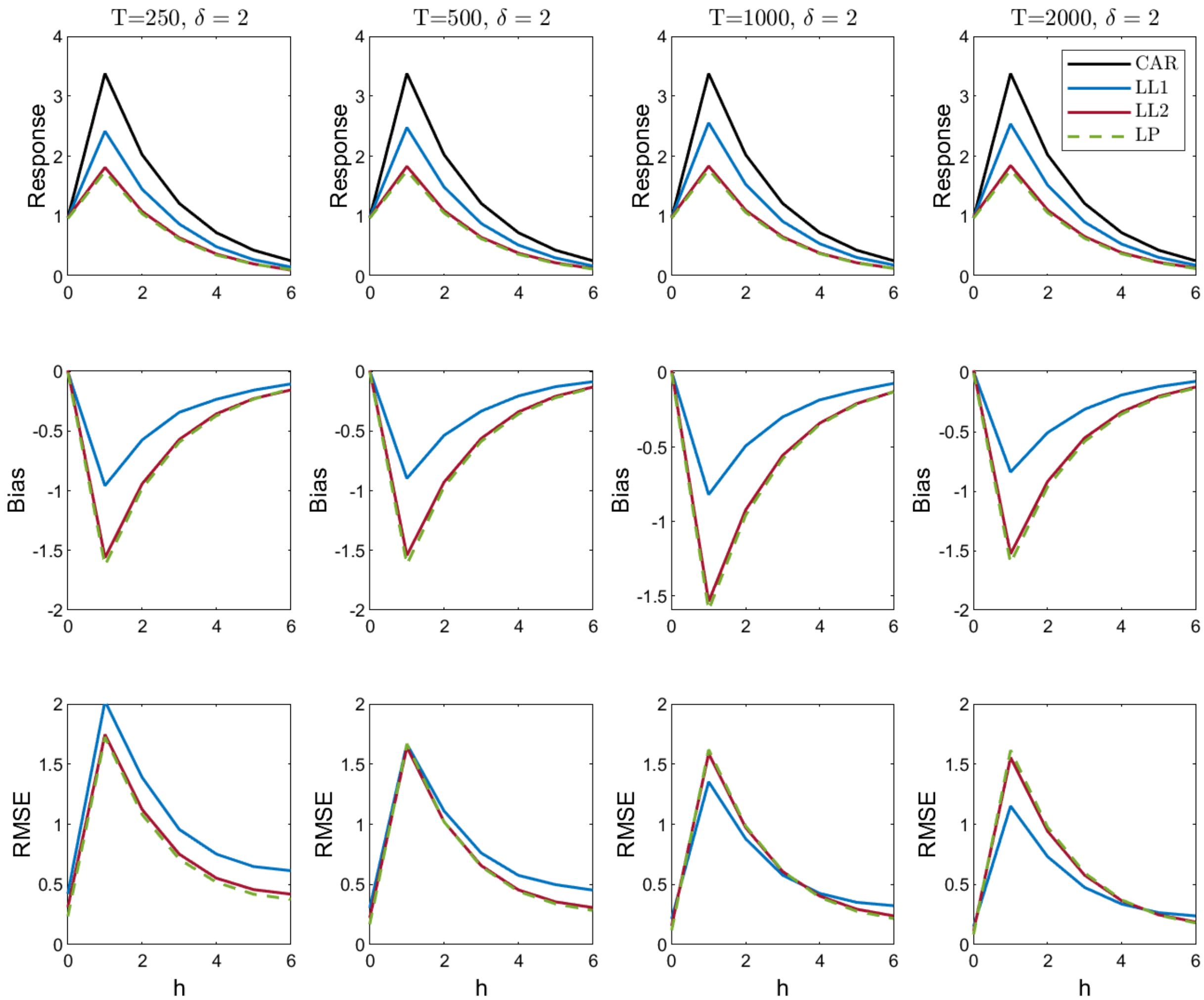
Notes: The black solid line reports the conditional average response function, CAR, the green dashed line illustrates the nonlinear LP estimator, the blue and red lines illustrate the local linear estimator with the rule-of-thumb bandwidth (LL1) and twice the rule-of-thumb bandwidth (LL2), respectively.

Figure 7: Model with Shock and State Interaction, $\delta = 2$, $f(x_t) = x_t^2$



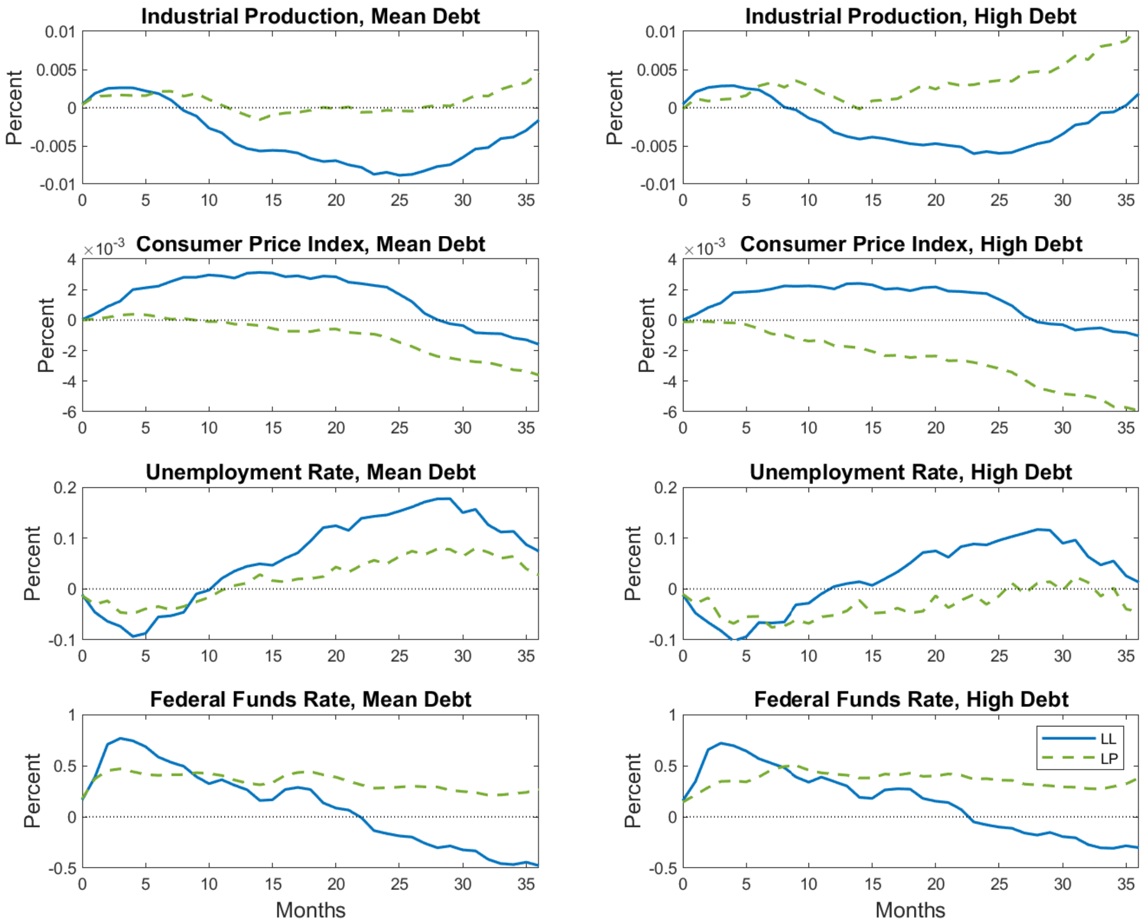
Notes: The black solid line reports the conditional average response function, CAR, the green dashed line illustrates the nonlinear LP estimator, the blue and red lines illustrate the local linear estimator with the rule-of-thumb bandwidth (LL1) and twice the rule-of-thumb bandwidth (LL2), respectively.

Figure 8: Model with Shock and State Interaction, $\delta = 2$, $f(x_t) = x_t^3$



Notes: The black solid line reports the conditional average response function, CAR, the green dashed line illustrates the nonlinear LP estimator, the blue and red lines illustrate the local linear estimator with the rule-of-thumb bandwidth (LL1) and twice the rule-of-thumb bandwidth (LL2), respectively.

Figure 9: Monetary Policy Effectiveness and Government Debt



Notes: The blue solid line plots to the local linear estimate and the green dashed line plots the nonlinear LP estimates.