Semiparametric Local Projections*

Sílvia Gonçalves† Ana María Herrera‡ Lutz Kilian§

Elena Pesavento¶ Iones Kelanemer Holban‖

June 16, 2026

Abstract

We propose a semiparametric local projection estimator of nonlinear impulse response functions for a broad class of structural dynamic models relevant for applied macroeconomics, including models with nonlinearly transformed regressors, state dependent coefficients, and nonlinear interactions between shocks and state variables. The estimator is based on a doubly robust moment condition that identifies the average response function as a linear functional of a nonparametric conditional mean, augmented by a density ratio that captures the effect of shifting the shock of interest. We combine this moment condition with cross-fitting that handles serial dependence. The resulting estimator is $\sqrt{T}$-consistent and asymptotically normal. We examine the finite-sample performance of the estimator across a range of nonlinear data generating processes and illustrate its use in two empirical examples.

JEL codes: C14, C32, E52, Q43

Keywords: impulse response, local projection, semiparametric estimation, double machine learning, nonlinear structural model, potential outcomes.

*We thank Jiaming Huang, Carlos Lamarche, Florian Gunsilius, seminar participants at Vanderbilt University, University of Kansas, Wake Forest University and Pittsburgh University as well as participants at the IAAE 2023, CIREQ 2023, MEG 2024, XV Time Series Workshop 2025, CEF-CM Statisics 2025, and ICEEE 2025 conferences for helpful comments. Jung Jae Kim and Cristhian Rosales-Castillo 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. We acknowledge the use of Anthropic's Claude (Opus 4.8 and Sonnet 4.6) to assist with coding and manuscript editing. The authors reviewed and verified all output and are solely responsible for the final content. Refine.ink was used to check the paper for consistency and clarity.

†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.

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

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

‖McGill University, Department of Economics, 855 Sherbrooke St. W., Montréal, Québec, H3A 2T7, Canada.

1 Introduction

Impulse response analysis is a cornerstone of empirical macroeconomics. Local projections have become a popular method for estimating impulse response functions (IRFs). In their simplest form, local projections consist of a sequence of OLS regressions, one for each horizon of interest. 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.

A large empirical literature has used generalizations of linear local projections to evaluate state-dependent impulse responses and other nonlinear responses. However, as shown in Gonçalves et al. (2021, 2024*a*,*b*), such state-dependent local projections fail to recover the population responses when the state is endogenous and the shock is large as in much of applied work (e.g., Ramey and Zubairy (2018)).

In this paper, we propose a semiparametric local projection estimator of nonlinear impulse response functions that is valid across a broad range of nonlinear settings relevant for applied macroeconomists. These include processes with nonlinearly transformed regressors (Herrera et al., 2015; Tenreyro and Thwaites, 2016; Ben Zeev et al., 2023; Caravello and Martinez-Bruera, 2024), state-dependent coefficients (Ramey and Zubairy, 2018), and nonlinear interactions between shocks and state variables (Caramp and Feilich, 2026; Cloyne et al., 2020, 2023).

As is common in macroeconomics, our object of interest is the impulse response function of an outcome variable y_{t+h} with respect to the primitive structural shock ε_{1t} in the equation for variable x_t. Specifically, we aim to identify and estimate the response of y_{t+h} to a shock of size δ in ε_{1t}. We assume that x_t is predetermined with respect to y_t, an exclusion restriction that encompasses situations in which $x_t = \varepsilon_{1t}$ is an observed i.i.d. shock, as in the narrative approach to identification. In this case, x_t is unconditionally independent of all other shocks driving the system between t and $t+h$. When x_t is not an observed shock ε_{1t}, the exclusion restriction together with the i.i.d. assumption on the structural shocks implies that x_t is conditionally independent of all other shocks between t and $t+h$, given control variables

$\mathbf{z}_{t-1}$ that include the history of the system up to $t-1$.

We show how this conditional independence condition, combined with the assumption that ε_{1t} enters x_t additively, can be used to identify the IRF of y_{t+h} with respect to ε_{1t} using only the observables $(y_{t+h}, x_t, \mathbf{z}_{t-1})$. In particular, the additive structure ensures that a δ-perturbation in ε_{1t} translates into a δ-perturbation in x_t, holding fixed the control variables. The structural equation for the outcome variable is left unrestricted, permitting arbitrary nonlinearities.

This identification strategy places our problem within the semiparametric literature on inference for linear functionals of regression functions (Newey (1994); Chernozhukov et al. (2018, 2022)). The key estimation challenge is that the conditional mean function $g_{0,h}(x,z) = E(y_{t+h}|x_t = x, \mathbf{z}_{t-1} = z)$ must be estimated nonparametrically, which can induce bias in the plug-in estimator. We address this concern by using the doubly robust moment condition of Chernozhukov et al. (2022), augmented by a density ratio reflecting the relative change in the conditional distribution of x_t given $\mathbf{z}_{t-1}$ when x_t is shifted by δ.

To handle the serial dependence of the data, we combine this moment condition with the NLO ("neighbors-left-out") cross-fitting approach of Semenova et al. (2023), which ensures approximate independence between training and evaluation sets. We derive the asymptotic distribution of the resulting estimator and show that it is $\sqrt{T}$-consistent and asymptotically normal, with the preliminary estimation of the nuisance functions having no effect on the first-order asymptotic distribution.

Our paper is related to a recent and growing literature on semiparametric and nonparametric inference for IRFs. The problem of estimating the average effects of policy interventions nonparametrically dates back at least to Stock (1989), but our focus is on the causal effect of structural macroeconomic shocks. Several recent papers have proposed nonparametric methods for estimating nonlinear IRFs. For example, Gourieroux and Lee (2023) propose a nonparametric local projection estimator for nonstructural IRFs identified from Gaussian shocks within a Markov process framework. Ballarin (2024) proposes a sieve-based nonparametric estimator for IRFs in models with nonlinearly transformed regressors, as in our Example 2.1 below, but does not cover the doubly robust approach, the more general nonlinear settings we consider, or inference. While our paper deals with shocks of finite magnitude δ, Kolesár and Plagborg-Møller (2025) focus on infinitesimally small shocks. They discuss the causal content of linear local projections when the data generating process is nonlinear and highlight challenges in applying doubly robust methods in the small samples typical of macroeconomics. Our paper builds on this work as well as two recent studies that apply doubly robust methods in macroeconometrics. Ballinari and Wehrli (2025) develop semiparametric inference for IRFs using double/debiased machine learning in a time series context, but focus on a binary treatment, so the adjustment term in their orthogonal moment condition is based on the propensity score rather than the density ratio we employ for continuous treatments. Huang et al. (2026) develop a two-step high-dimensional nonparametric local projection estimator combining Neyman-orthogonal pseudo-outcomes with cross-fitting. Because they focus on an IRF that shifts the policy variable from a fixed baseline, their second step requires a nonparametric regression and yields convergence rates slower than $\sqrt{T}$; in contrast, our estimand averages over the distribution of the shock and can be estimated at rate $\sqrt{T}$. Finally, and independently, Nikolaishvili (2026), building on an earlier version of our paper (Gonçalves et al., 2024a), proposes a closely related doubly robust estimator for nonparametric local projections under different assumptions and without allowing for covariates in the conditioning set.

While our main focus is on identifying unconditional responses, our analysis also extends to average response functions conditional on a state variable Ω_t. We provide identification conditions for this object and show how our semiparametric local projections estimator can be applied to each subsample $\{t : \Omega_t = \omega\}$ when Ω_t is discrete, as in state-dependent models.

The paper is organized as follows. Section 2 introduces the structural model and three leading examples. Section 3 defines the population IRFs of interest and contrasts our definition with alternative definitions used in the literature. Sections 4 and 5 discuss identification and estimation, and inference, respectively. Section 6 briefly discusses how to extend the analysis to conditional IRFs. The simulation results are presented in Section 7. Section 8 contains two empirical illustrations focusing on possible nonlinearities in the pass-through of gasoline price shocks to inflation and in the response of motor vehicle sales to real gasoline price shocks. We conclude in Section 9. The proofs are relegated to Appendices A and B.

2 Framework

Let $z_t = (x_t, y_t)'$ denote a vector of observed time series, where y_t is the outcome of interest and x_t is predetermined with respect to y_t. For example, y_t could be real GDP and x_t government spending. For simplicity, we assume that y_t is univariate, but extensions to multivariate outcomes could be easily accommodated. A general structural model for z_t is a triangular system of the form

$$x_t = \phi(\mathbf{z}_{t-1}) + \varepsilon_{1t}, \tag{1}$$

$$y_t = \mu(x_t,\, \mathbf{z}_{t-1}, \varepsilon_{2t}), \tag{2}$$

where ϕ and μ are (potentially unknown) nonlinear functions, and $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})'$ is a vector of mutually independent structural shocks. We assume ε_t to be i.i.d. with mean zero and diagonal covariance matrix $\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2)$. The vector $\mathbf{z}_{t-1} = (z_{t-1}, z_{t-2}, \ldots, z_{t-p})'$ contains lags of z_t; other variables can be included in $\mathbf{z}_{t-1}$ provided they are predetermined with respect to ε_{1t}. The exclusion of y_t from equation (1) is equivalent to the assumption of block recursiveness in linear structural VAR identification, where $x_t = \phi'\mathbf{z}_{t-1} + \varepsilon_{1t}$. An important special case is $x_t = \varepsilon_{1t}$, as in the narrative approach to identification.

Consistent with standard impulse response analysis in macroeconomics, our goal is to trace the effect over time of surprise changes in the variable x_t, as captured by a one-time change in the structural shock ε_{1t}, rather than changes in x_t itself. This is the main reason why we assume in (1) that x_t and ε_{1t} are separable, as this is crucial for identifying the impulse response function of y_{t+h} with respect to ε_{1t} from the observables $(y_{t+h}, x_t, \mathbf{z}_{t-1})$. A non-separable specification $x_t = \phi(\mathbf{z}_{t-1}, \varepsilon_{1t})$ could be considered if instead we targeted the impulse response function of y_{t+h} with respect to x_t, as in Kolesár and Plagborg-Møller (2025) and Huang et al. (2026).

Our framework accommodates a range of models of interest in applied work. One example is a model with nonlinearly transformed regressors. This model allows for a sign nonlinearity in the responses with the magnitude of the response depending on the sign of x_t (e.g., $f(x_t) = \max\{x_t, 0\}$) or a size nonlinearity with the magnitude of the response depending on the size of x_t (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 et al. (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 2.1 (Model with Nonlinear Regressors) *Let*

$$\begin{aligned} x_t &= \phi' \mathbf{z}_{t-1} + \varepsilon_{1t}, \\ y_t &= \beta x_t + \rho' \mathbf{z}_{t-1} + cf\left(x_t\right) + \varepsilon_{2t}, \end{aligned}$$

where f is a potentially unknown nonlinear function.

A second example is the state-dependent model examined in Gonçalves et al. (2024*b*) 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 2.2 (State-Dependent Model) *Let*

$$\begin{aligned} x_t &= \varepsilon_{1t} \\ y_t &= \beta_{t-1} x_t + \gamma_{t-1}' \mathbf{z}_{t-1} + \varepsilon_{2t}, \end{aligned}$$

with $\beta_{t-1} = \beta_E S_{t-1} + \beta_R (1 - S_{t-1})$ 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 expansion or in recession. When S_{t-1} depends on $\mathbf{z}_{t-1}$, the endogenous variables in the system, the IRF becomes nonlinear in ε_{1t}.*

A final example is inspired by Cloyne et al. (2020, 2023) and Caramp and Feilich (2026) 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 a 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 interaction between the debt level and the shock of interest induces a nonlinearity that needs to be taken into account when estimating the IRF. 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 2.3 (Nonlinear Interaction Term of Shock with State) *Let*

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

where ε_{1t} is mutually independent of ε_{2t} and ε_{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

Our main analysis focuses on unconditional versions of the IRF. Although the estimation and inference methods presented in Section 5 are tailored to estimating unconditional IRFs, they can also be applied to conditional IRFs in state-dependent models such as in Example 2.2, where the conditioning set is discrete. Section 6 discusses this application as well as the challenges one would face in estimating conditional IRFs in other examples such as Example 2.3.

Following the standard approach in macroeconomics, we care about the response of y_{t+h} with respect to the structural shock ε_{1t}. As in the recent macroeconometrics literature, we adopt a potential outcomes framework (see e.g., Gonçalves et al. (2021) and Gonçalves et al. (2024*b*)). One implication of the structural model (1)–(2) is that y_{t+h} can be written as $y_{t+h} = m_h(\varepsilon_{1t}, U_{t+h})$, where m_h is obtained by iterating (2) forward h steps and substituting (1), and $U_{t+h} \equiv (\varepsilon_{2t}, \varepsilon_{1,t+1}, \varepsilon_{2,t+1}, \ldots, \varepsilon_{1,t+h}, \varepsilon_{2,t+h}, \mathbf{z}'_{t-1})'$ collects all remaining determinants of y_{t+h}. Since ε_{1t} is i.i.d. and independent of $\{\varepsilon_{2t}\}$, ε_{1t} is independent of U_{t+h}, which we write as $\varepsilon_{1t} \perp U_{t+h}$. The potential outcome associated with fixing $\varepsilon_{1t} = e$ is $y_{t+h}(e) = m_h(e, U_{t+h})$, where e is any fixed value in the support of ε_{1t}. The observed outcome satisfies $y_{t+h} = y_{t+h}(\varepsilon_{1t})$, implying that it is the value that we observe when e takes the value ε_{1t} that generated the observed data. The fact that ε_{1t} and U_{t+h} are mutually independent implies that the potential outcomes are independent of ε_{1t}.

To define the response function of y_{t+h} with respect to ε_{1t}, we compare the (observed) baseline value $y_{t+h}(\varepsilon_{1t})$ with the counterfactual (unobserved) value of y at $t+h$ that would have been observed if ε_{1t} had been subject to a shock of size δ, denoted $y_{t+h}(\varepsilon_{1t}+\delta)$ (e.g., Potter (2000)). In particular, following Gonçalves et al. (2024*b*), we adopt the following definition:

Definition 1 *The average response function of y_{t+h} to a shock of size δ in ε_{1t} is defined as* $\mathrm{ARF}_h(\delta) \equiv E\left(y_{t+h}\left(\varepsilon_{1t}+\delta\right) - y_{t+h}\left(\varepsilon_{1t}\right)\right)$.

$\mathrm{ARF}_h(\delta)$ corresponds to the unconditional average response used in Gonçalves et al. (2021). A conditional average response function can also be defined as in Definition 3 in Section 6, following Gonçalves et al. (2024*a*).[1]

Definition 1 is not the only possible definition of an unconditional IRF. Other studies such as Koop et al. (1996), Rambachan and Shephard (2021), and Huang et al. (2026),

[1] Alternatively, one could also define versions of these IRFs where $\delta \to 0$, as discussed in Gonçalves et al. (2024*a*). 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)).

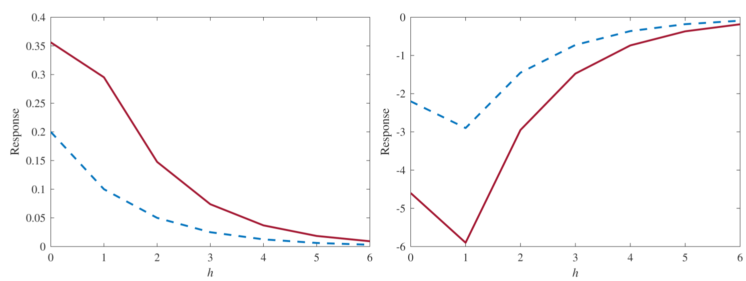
for example, have instead compared the two potential outcomes $y_{t+h}(e')$ and $y_{t+h}(e)$, often setting $e' = \delta$ and $e = 0$, which yields the alternative definition:

Definition 2 *The response function of y_{t+h} to a shock of size δ in $\varepsilon_{1t} = e$ is defined as* $\text{ARF}^*_h(\delta, e) = E(y_{t+h}(e + \delta) - y_{t+h}(e))$.

Whereas Definition 2 has been used widely in the literature, Definition 1 is more recent (Gonçalves et al. (2021, 2024*a*,*b*)). These two definitions are equivalent when the potential outcome is linear in e for all horizons, as would be the case for a linear model or in special cases of Example 2.2 and Example 2.3 when the conditioning sets (S_{t-1} and r_t, respectively) are exogenous. However, in general, the two definitions differ.

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 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 IRF 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, if $\mathcal{F}^{t-1}$ denotes the information available up to time $t-1$, a natural baseline is the conditional expectation $E_{t-1}(y_t) \equiv E(y_t \mid \mathcal{F}^{t-1})$. In this example, $E_{t-1}(y_t) = 0.5y_{t-1} + 0.3x_{t-1} - 0.4E_{t-1}(f(x_t)) - 0.3f(x_{t-1})$, 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 work with Definition 1 throughout this paper.

4 Identification

We discuss the identification of $\theta_{0,h} \equiv \mathrm{ARF}_h(\delta)$ in Definition 1. Motivated by Section 6 of Kolesár and Plagborg-Møller (2025), we first discuss identification based on a regression-based approach, where $\theta_{0,h}$ is identified using the conditional expectation function $g_{0,h}(x,z) \equiv E(y_{t+h}|x_t = x, \mathbf{z}_{t-1} = z)$, and then show how to obtain identification using a doubly robust approach. We use the subscript "0" to indicate true parameters and functions throughout.

Starting with the regression-based approach, note that the independence between ε_{1t} and U_{t+h} (which holds by (1) and (2) under the i.i.d. assumption on ε_t and the mutually independent shocks assumption) allows us to identify $\theta_{0,h}$ as $\theta_{0,h} = E[g_{\varepsilon,0,h}(\varepsilon_{1t} + \delta) - g_{\varepsilon,0,h}(\varepsilon_{1t})]$, where $g_{\varepsilon,0,h}(e) \equiv E(y_{t+h}|\varepsilon_{1t} = e)$ (e.g., Kolesár and Plagborg-Møller (2025)). This representation is useful for estimation when ε_{1t} is an observed shock, as in the narrative approach to identification. However, it does not directly apply when x_t is an observed variable and ε_{1t} is its underlying (unobserved) structural shock. In what follows, we show how the additive structure of (1) can be exploited to identify the average response function of y_{t+h}

to an impulse in ε_{1t} (i.e., $\text{ARF}_h(\delta)$ in Definition 1) in this more general context: since $x_t = \phi(\mathbf{z}_{t-1}) + \varepsilon_{1t}$, a δ-shift in ε_{1t} holding $\mathbf{z}_{t-1}$ fixed is identical to a δ-shift in x_t holding $\mathbf{z}_{t-1}$ fixed, yielding an identification result for $\theta_{0,h}$ in terms of the observables $(y_{t+h}, x_t, \mathbf{z}_{t-1})$ only.

Proposition 4.1 *Suppose that $z_t = (x_t, y_t)'$ satisfies (1) and (2) where $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})'$ contains mutually independent shocks and is i.i.d.$(0, \Sigma)$ with $\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2)$. It follows that*

$$\theta_{0,h} \equiv \text{ARF}_h(\delta) = E[g_{0,h}(x_t + \delta,\, \mathbf{z}_{t-1}) - g_{0,h}(x_t,\, \mathbf{z}_{t-1})]\,, \tag{3}$$

where $g_{0,h}(x, z) \equiv E(y_{t+h} \mid x_t = x,\, \mathbf{z}_{t-1} = z)$.

The representation of $\text{ARF}_h(\delta)$ given in (3) is identified from observables because $g_{0,h}(x, z)$ is the conditional mean of y_{t+h} given $(x_t, \mathbf{z}_{t-1})$, a functional of the joint distribution of $(y_{t+h}, x_t, \mathbf{z}_{t-1})$, and the outer expectation is taken over the marginal distribution of $(x_t, \mathbf{z}_{t-1})$, both of which are observable. No knowledge of ϕ or μ is required. Identification requires that $g_{0,h}$ be defined on the support of $(x_t + \delta, \mathbf{z}_{t-1})$ as well as on the support of $(x_t, \mathbf{z}_{t-1})$. When the support of ε_{1t} (and hence of x_t given $\mathbf{z}_{t-1}$) is bounded, the shifted support may fall outside the original one, in which case $g_{0,h}(x_t + \delta, \mathbf{z}_{t-1})$ is not identified for some $(x_t, \mathbf{z}_{t-1})$ pairs. For this reason, we do not impose bounded support conditions.

Remark 1 *The additive structure in (1) is important for identifying $\theta_{0,h} \equiv \text{ARF}_h(\delta)$ as stated in Definition 1. It implies that perturbing ε_{1t} by δ is equivalent to perturbing x_t by δ holding $\mathbf{z}_{t-1}$ fixed, so that $\theta_{0,h}$ can equivalently be written as*

$$\theta_{0,h} = E[m_h(x_t + \delta,\, U_{t+h}) - m_h(x_t,\, U_{t+h})] = E[g_{0,h}(x_t + \delta,\, \mathbf{z}_{t-1}) - g_{0,h}(x_t,\, \mathbf{z}_{t-1})]\,,$$

where m_h is a reparametrization of the structural function introduced in Section 3, now expressed as a function of x_t. Without the additive structure, the first equality fails. Nevertheless, the estimand $E[g_{0,h}(x_t + \delta,\, \mathbf{z}_{t-1}) - g_{0,h}(x_t,\, \mathbf{z}_{t-1})]$ retains a causal interpretation as the IRF of y_{t+h} to a δ-perturbation of x_t holding $\mathbf{z}_{t-1}$ fixed.

Remark 2 *If contemporaneous feedback from y_t to x_t were allowed, Definition 1 would still* *define the causal effect of a shock on the relevant potential outcome. However, the identification argument in Proposition 4.1 would no longer apply directly: it would require an additional instrument or identification strategy to isolate the exogenous component of x_t. We reserve this question for future research.*

Proposition 4.1 leads to a moment condition of the form

$$E\left[g_h(x_t + \delta, \mathbf{z}_{t-1}) - g_h(x_t, \mathbf{z}_{t-1}) - \theta_h\right] = 0, \tag{4}$$

which identifies $\theta_{0,h}$ when $g_h = g_{0,h}$. A natural approach is to replace $g_{0,h}$ by a first-step estimator $\hat{g}_h$, yielding $\widehat{\text{ARF}}_h(\delta) = T^{-1}\sum_{t=1}^{T}\left[\hat{g}_h(x_t + \delta, \mathbf{z}_{t-1}) - \hat{g}_h(x_t, \mathbf{z}_{t-1})\right]$. When $\hat{g}_h$ is estimated by machine learning or nonparametric methods, this regression-based estimator can suffer from first-order bias, and inference requires adjusting for estimation uncertainty in $g_{0,h}$. This motivates using the double/debiased machine learning approach of Chernozhukov et al. (2018, 2022, 2025) to augment the moment condition (4) using Neyman orthogonality. When combined with a form of cross-fitting that handles time series dependence, inference based on this estimator can proceed as if the nuisance functions were fully observed.

To describe the doubly robust approach, let $f_{0,x|z}(x|z)$ denote the conditional density of x_t given $\mathbf{z}_{t-1}$. It can be easily shown that for any function g_h,

$$E\left[g_h(x_t + \delta, \mathbf{z}_{t-1}) - g_h(x_t, \mathbf{z}_{t-1})\right] = E\left[\alpha_0(x_t, \mathbf{z}_{t-1}) g_h(x_t, \mathbf{z}_{t-1})\right], \tag{5}$$

where $\alpha_0(x, z) \equiv (f_{0,x|z}(x - \delta|z) - f_{0,x|z}(x|z))/f_{0,x|z}(x|z)$ is the Riesz representer (we omit the dependence on δ throughout). This Riesz representer is a density ratio that captures the relative change in the conditional density of x_t given $\mathbf{z}_{t-1}$ when x_t is shifted by δ (see Section 6 of Kolesár and Plagborg-Møller (2025), who report the form of the Riesz representer when $\delta \to 0$).

Proposition 4.2 *Suppose the conditions of Proposition 4.1 hold. Then $\theta_{0,h}$ solves*

$$E\left[g_{0,h}(x_t + \delta, \mathbf{z}_{t-1}) - g_{0,h}(x_t, \mathbf{z}_{t-1}) - \theta_h + \alpha_0(x_t, \mathbf{z}_{t-1})(y_{t+h} - g_{0,h}(x_t, \mathbf{z}_{t-1}))\right] = 0. \tag{6}$$

As it turns out, Equation (6) yields a doubly robust moment equation for $\theta_{0,h}$. Defining

$$\psi(y_{t+h}, x_t, \mathbf{z}_{t-1}, g_h, \alpha, \theta_h) = g_h(x_t+\delta, \mathbf{z}_{t-1}) - g_h(x_t, \mathbf{z}_{t-1}) - \theta_h + \alpha(x_t, \mathbf{z}_{t-1})(y_{t+h} - g_h(x_t, \mathbf{z}_{t-1})),$$

we can show that $E[\psi(y_{t+h}, x_t, \mathbf{z}_{t-1}, g_h, \alpha, \theta_{0,h})] = 0$ holds for any g_h when $\alpha = \alpha_0$ and for any α when $g_h = g_{0,h}$. This follows by an application of Theorem 5 of Chernozhukov et al. (2022). Hence, this moment condition yields an estimator that is insensitive to misspecification of g_h provided $\alpha = \alpha_0$, and insensitive to misspecification of α provided $g_h = g_{0,h}$.

5 Estimation and inference

We estimate $\theta_{0,h}$ by combining the doubly robust moment condition in (6) with cross-fitting. In the standard cross-fitting approach, the data are partitioned into non-overlapping blocks. For each block, the nuisance functions are estimated on the complement of that block and evaluated on the heldout observations. These out-of-sample nuisance estimates are then used to construct the moment condition for estimating $\theta_{0,h}$; see, for example, Chernozhukov et al. (2022). A crucial assumption that justifies this approach is random sampling, which implies that the blocks are mutually independent (and the data within blocks are i.i.d.).

In a time series context, the presence of serial dependence in $z_t = (x_t, y_t)'$ violates this assumption even when x_t is i.i.d., creating dependence among the blocks. Hence, we follow Semenova et al. (2023) and rely on NLO ("neighbors-left-out") cross-fitting.[2] We partition the index set $\{1, \ldots, T\}$ into $K \geq 4$ non-overlapping blocks $I_1, \ldots, I_K$ of contiguous time indices, each of size $T_\ell \equiv T/K$: $\{1, \ldots, T\} = I_1 \cup \cdots \cup I_K$.[3] We keep K fixed as $T \to \infty$ when deriving the asymptotic theory below. For each $\ell \in \{1, \ldots, K\}$, let $\mathcal{N}(\ell)$ denote the set containing ℓ and its immediate neighbors in $\{1, \ldots, K\}$, i.e. $\mathcal{N}(\ell) = \{\ell-1, \ell, \ell+1\} \cap$

[2]This approach leaves out not only the target block but also its immediate neighbors when estimating the nuisance functions. The resulting training and evaluation sets are then approximately independent, with the approximation error controlled by the speed of mixing of the underlying time series. Recent applications of the NLO cross-fitting approach to inference on impulse response functions in time series include Ballinari and Wehrli (2025) and Huang et al. (2026). Because their estimands are different than ours, their assumptions and asymptotic results also differ from ours.

[3]For simplicity we assume that T is divisible by K.

$\{1, \dots, K\}$, and define the quasi-complement of I_ℓ as $I_\ell^{\mathrm{qc}} = \bigcup_{j \notin \mathcal{N}(\ell)} I_j$, so that I_ℓ^{qc} is obtained from the full complement $I_{-\ell} = \bigcup_{j \neq \ell} I_j$ by additionally removing the two blocks adjacent to I_ℓ. The key feature of NLO cross-fitting is that I_ℓ and I_ℓ^{qc} are separated by at least T_ℓ time periods. Since K is fixed, as $T \to \infty$ we have $T_\ell \to \infty$, so that under mixing-type conditions I_ℓ is approximately independent of I_ℓ^{qc}.

Given horizon h, the NLO cross-fitting estimator of $\theta_{0,h}$, which we refer to as DR-NLO, is computed as follows. For notational simplicity, we drop the horizon index h in the nuisance functions and estimator that correspond to block ℓ, with the understanding that all objects depend on h. For each $\ell = 1, \dots, K$:

1. Estimate $\hat{g}_\ell$ and $\hat{\alpha}_\ell$ using some nonparametric or machine learning procedure on the quasi-complement I_ℓ^{qc}.

2. Set the average of the moment condition (6) over I_ℓ to zero to obtain

$$\hat{\theta}_\ell = \frac{1}{T_\ell} \sum_{t \in I_\ell} \left[\hat{g}_\ell(x_t + \delta, \mathbf{z}_{t-1}) - \hat{g}_\ell(x_t, \mathbf{z}_{t-1}) + \hat{\alpha}_\ell(x_t, \mathbf{z}_{t-1})(y_{t+h} - \hat{g}_\ell(x_t, \mathbf{z}_{t-1})) \right].$$

3. Compute the estimator of $\theta_{0,h}$ as

$$\hat{\theta}_h = \frac{1}{K} \sum_{\ell=1}^{K} \hat{\theta}_\ell = \frac{1}{T} \sum_{\ell=1}^{K} \sum_{t \in I_\ell} \left[\hat{g}_\ell(x_t + \delta, \mathbf{z}_{t-1}) - \hat{g}_\ell(x_t, \mathbf{z}_{t-1}) + \hat{\alpha}_\ell(x_t, \mathbf{z}_{t-1})(y_{t+h} - \hat{g}_\ell(x_t, \mathbf{z}_{t-1})) \right].$$

We next state the regularity conditions used to derive the asymptotic distribution of $\hat{\theta}_h$. Following Semenova et al. (2023) and Huang et al. (2026), we impose geometric β-mixing on $\{z_t\}$. β-mixing is strictly stronger than the more standard strong mixing assumption, but allows us to use the Strassen coupling result that underlies the NLO theory of Semenova et al. (2023). We define the β-mixing coefficients as $\beta(j) \equiv \sup_t \beta(\sigma(z_s, s \leq t), \sigma(z_s, s \geq t + j))$, where $\beta(\mathcal{A}, \mathcal{B}) = E[\sup_{B \in \mathcal{B}} |P(B \mid \mathcal{A}) - P(B)|]$ for any σ-algebras $\mathcal{A}$ and $\mathcal{B}$. The function $\psi(y_{t+h}, x_t, \mathbf{z}_{t-1}, g_0, \alpha_0, \theta_{0,h})$ in Assumption 2 below is the doubly robust moment function defined in (6) (see Proposition 4.2). We let $\mathcal{X}$ and $\mathcal{Z}$ denote the supports of x_t and $\mathbf{z}_{t-1}$.

Assumption 1 *$\{z_t = (x_t, y_t)'\}$ is stationary and geometrically β-mixing, i.e., $\beta(j) \leq C \exp(-c_\beta j)$ for some constants $C \geq 0$ and $c_\beta > 0$, and for all $j \geq 1$.*

Assumption 2 *For some finite constants $\bar{\alpha}$, $\bar{\sigma}_q$, and for some $q > 2$:*

(i) $\sup_{(x,z)\in\mathcal{X}\times\mathcal{Z}} |\alpha_0(x, z)| < \bar{\alpha}$.

(ii) $\sup_{(x,z)\in\mathcal{X}\times\mathcal{Z}} E[|e_{t+h}|^q | x_t = x, \mathbf{z}_{t-1} = z] < \bar{\sigma}_q^q$, *where* $e_{t+h} \equiv y_{t+h} - g_0(x_t, \mathbf{z}_{t-1})$.

(iii) $E[|\psi(y_{t+h}, x_t, \mathbf{z}_{t-1}, g_0, \alpha_0, \theta_{0,h})|^{2+\epsilon}] < \infty$ *for some* $\epsilon > 0$.

Assumption 2(i) requires the Riesz representer $\alpha_0(x, z)$ to be uniformly bounded over $\mathcal{X} \times \mathcal{Z}$, the support of $(x_t, \mathbf{z}_{t-1})$. This rules out distributions with thin-tails such as the Gaussian, but allows for Student-t distributions. Assumption 2(ii) imposes a conditional qth moment bound on the regression residual e_{t+h} for $q > 2$, which is needed to control autocovariance terms arising from serial dependence. Assumption 2(iii) is a standard moment condition on the influence function that, together with Assumption 1, ensures that a central limit theorem applies to the oracle estimator based on the true nuisance functions.

Assumption 3 *For each $\ell = 1, \ldots, K$, $\hat{g}_\ell \in \mathcal{G}_T$ and $\hat{\alpha}_\ell \in \mathcal{A}_T$ with probability converging to one, where $\mathcal{G}_T$ and $\mathcal{A}_T$ denote shrinking neighborhoods of $g_{0,h}$ and α_0, respectively, and q is as in Assumption 2. In addition,*

(i) $r_{g,q,T} \equiv \sup_{g\in\mathcal{G}_T} \left(E\left[|g(x_t, \mathbf{z}_{t-1}) - g_{0,h}(x_t, \mathbf{z}_{t-1})|^q\right]\right)^{1/q} = o(1)$;
$r_{\alpha,q,T} \equiv \sup_{\alpha\in\mathcal{A}_T} \left(E\left[|\alpha(x_t, \mathbf{z}_{t-1}) - \alpha_0(x_t, \mathbf{z}_{t-1})|^q\right]\right)^{1/q} = o(1)$.

(ii) $\sqrt{T}\, r_{g,T}\, r_{\alpha,T} = o(1)$, *where* $r_{g,T} \equiv r_{g,2,T}$ *and* $r_{\alpha,T} \equiv r_{\alpha,2,T}$ *denote the L_2 convergence rates.*

Assumption 3(i) imposes L_q consistency on both nuisance estimators, which implies L_2 consistency since $q > 2$. Assumption 3(ii) is the product rate condition standard in the double machine learning literature; see Assumption 2 of Chernozhukov et al. (2025). The L_q rates in (i) are stronger than the L_2 rates typically assumed in the i.i.d. case and are needed here to control autocovariance terms arising from serial dependence in e_{t+h} and $(x_t, \mathbf{z}_{t-1})$; see Remarks 3 and 4 below.

Theorem 5.1 *Suppose that Assumptions 1, 2, and 3 hold. Then*

$$\sqrt{T}(\hat{\theta}_h - \theta_{0,h}) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \psi(y_{t+h}, x_t, \mathbf{z}_{t-1}, g_{0,h}, \alpha_0, \theta_{0,h}) + o_p(1) \to_d N(0, V_h),$$

where V_h is the long-run variance of $\psi(y_{t+h}, x_t, \mathbf{z}_{t-1}, g_{0,h}, \alpha_0, \theta_{0,h})$.

The proof of Theorem 5.1 is in Appendix B. Letting $\tilde{\theta}_h$ denote the oracle estimator which assumes we know g_0 and α_0, we decompose

$$\sqrt{T}(\hat{\theta}_h - \theta_{0,h}) = \sqrt{T}(\tilde{\theta}_h - \theta_{0,h}) + \sqrt{T}(\hat{\theta}_h - \tilde{\theta}_h),$$

and show that $\sqrt{T}(\hat{\theta}_h - \tilde{\theta}_h) \equiv \sqrt{T}(R_1 + R_2 + R_3) = o_p(1)$, where R_1, R_2 and R_3 are remainder terms defined in Appendix B. Assumptions 1, 2 and 3 suffice for proving that these remainders are $o_p(T^{-1/2})$.

Remark 3 *Assumption 2(ii) and the L_q rate on $\hat{\alpha}_\ell$ in Assumption 3(i) can both be relaxed to $q = 2$, as is standard in the i.i.d. case, provided we impose a direct condition on the autocovariance structure of e_{t+h}. For instance, it suffices to assume $E(e_{t+h}|x_t, \mathcal{F}^{t-1}) = 0$, where $\mathcal{F}^{t-1}$ denotes the lagged history of z_t, together with uniform boundedness of the conditional autocovariances. Alternatively, one can drop the martingale difference condition and instead assume absolute summability of the conditional autocovariance sequence. The L_q rate on $\hat{g}_\ell$ in Assumption 3(i) is still required, as it controls autocovariance terms in R_1 arising from serial dependence in $(x_t, \mathbf{z}_{t-1})$ rather than in e_{t+h}.*

Remark 4 *For the special case where x_t is i.i.d. and $x_t \perp U_{t+h}$ without control variables $\mathbf{z}_{t-1}$, the L_q rate on $\hat{g}_\ell$ in Assumption 3(i) is not required. The main reason is that we can exploit the independence of x_t when showing that $\sqrt{T}R_1 = o_p(1)$. However, even when x_t is i.i.d., e_{t+h} may be serially dependent, so the L_q rate on $\hat{\alpha}_\ell$ in Assumption 3(i) is still needed for $\sqrt{T}R_2 = o_p(1)$. Alternatively, only L_2 rates on $\hat{\alpha}_\ell$ are required if further conditions on the autocovariance structure of e_{t+h} are imposed, as discussed in Remark 3.*

The main implication of Theorem 5.1 is that the preliminary estimation of $g_{0,h}$ and α_0

does not affect the first-order asymptotic distribution of $\hat{\theta}_h$: the estimator is asymptotically equivalent to the infeasible oracle estimator that uses the true nuisance functions. This is a consequence of the Neyman orthogonality of the moment condition (6) combined with the approximate independence between I_ℓ and I_ℓ^{qc} under NLO cross-fitting. A feasible confidence interval for $\theta_{0,h}$ can be constructed using a standard HAC estimator of V_h based on the estimated influence function $\psi(y_{t+h}, x_t, \mathbf{z}_{t-1}, \hat{g}_\ell, \hat{\alpha}_\ell, \hat{\theta}_h)$.

6 Conditional impulse response functions

Conditional impulse response functions are often of interest in applications such as in Examples 2.2 or 2.3. A generalization of Definition 1 to conditional IRFs is as follows.

Definition 3 *The conditional average response function of y_{t+h} to a shock of size δ in ε_{1t} is defined as $CAR_h(\delta, \omega) \equiv E\left(y_{t+h}\left(\varepsilon_{1t}+\delta\right) - y_{t+h}\left(\varepsilon_{1t}\right)|\Omega_t = \omega\right)$, where Ω_t denotes the conditioning set.*

Since ε_{1t} is random, the conditional expectation in Definition 3 averages over all possible realizations of ε_{1t} (in addition to the other sources of randomness that enter into the potential outcomes through U_{t+h}), conditionally on $\Omega_t = \omega$. The choice of ω in $CAR_h(\delta, \omega)$ is context-dependent. For instance, in Example 2.2 the conditioning set Ω_t is the state variable at time $t-1$, i.e. $\Omega_t = S_{t-1}$, so ω is either 0 or 1, while in Example 2.3 the conditioning set is $\Omega_t = r_t$, so ω can take on any value in the support of r_t.

The following proposition provides conditions under which $CAR_h(\delta, \omega)$ is identified.

Proposition 6.1 *Suppose that $z_t = (x_t, y_t)'$ satisfies* (1) *and* (2) *where $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})'$ is a vector of mutually independent shocks that is distributed i.i.d.$(0, \Sigma)$ with $\Sigma = \operatorname{diag}(\sigma_1^2, \sigma_2^2)$. Suppose further that $x_t \perp U_{t+h} \mid (\mathbf{z}_{t-1}, \Omega_t)$. It follows that*

$$\text{CAR}_h(\delta, \omega) = E[g_{0,h}(x_t + \delta,\, \mathbf{z}_{t-1}, \omega) - g_{0,h}(x_t,\, \mathbf{z}_{t-1}, \omega) \mid \Omega_t = \omega]\,, \tag{7}$$

where $g_{0,h}(x, z, \omega) \equiv E(y_{t+h} \mid x_t = x,\, \mathbf{z}_{t-1} = z,\, \Omega_t = \omega)$.

The key condition in Proposition 6.1 is $x_t \perp U_{t+h} \mid (\mathbf{z}_{t-1}, \Omega_t)$, which requires that conditionally on $(\mathbf{z}_{t-1}, \Omega_t)$, all remaining variation in x_t comes from ε_{1t} alone. This condition holds in two leading cases. First, if Ω_t is a function of $\mathbf{z}_{t-1}$ alone, then conditioning on $(\mathbf{z}_{t-1}, \Omega_t)$ is the same as conditioning on $\mathbf{z}_{t-1}$ alone, and conditional independence follows directly from (1) and (2), as in Proposition 4.1. This covers Example 2.2, where $\Omega_t = S_{t-1}$ is a function of $\mathbf{z}_{t-1}$.[4] Second, if Ω_t contains variables not in $\mathbf{z}_{t-1}$, conditional independence is satisfied if $\Omega_t \perp \varepsilon_{1t}$. This covers Example 2.3, where $\Omega_t = r_t = f(x_{t-1}) + \varepsilon_{3t}$: since x_{t-1} is dated $t-1$ and $\varepsilon_{3t} \perp \varepsilon_{1t}$ by assumption, $r_t \perp \varepsilon_{1t}$ and conditional independence holds. However, this would fail if r_t depended on x_t (and hence on ε_{1t}), for instance if $r_t = f(x_t) + \varepsilon_{3t}$.

In the special case where Ω_t is binary, as in Example 2.2 where $\Omega_t = S_{t-1} \in \{0, 1\}$, $CAR_h(\delta, \omega)$ takes two values $CAR_h(\delta, 0)$ and $CAR_h(\delta, 1)$, corresponding to the impulse responses in each state. Each can be estimated by applying the NLO cross-fitting estimator of Section 5 separately to each subsample $\{t : \Omega_t = \omega\}$, $\omega \in \{0, 1\}$.

When Ω_t is continuous, as in Example 2.3 where $\Omega_t = r_t$, $CAR_h(\delta, \omega)$ is a function of a continuous argument ω. From (7), it equals the conditional expectation of $g_{0,h}(x_t + \delta, \mathbf{z}_{t-1}, \omega) - g_{0,h}(x_t, \mathbf{z}_{t-1}, \omega)$ given $\Omega_t = \omega$, which must be estimated nonparametrically as a function of ω. This introduces an additional nonparametric estimation step beyond what is required for $\text{ARF}_h(\delta)$. We leave a formal treatment of this case for future work.

7 Simulation results

This section evaluates the finite sample performance of the semiparametric local projection estimator developed in Section 5. We focus on the nonlinear regressors model of Example 2.1 and the state-dependent design of Example 2.2. In both cases, the structural shocks ε_{1t} and ε_{2t} are mutually independent, each drawn i.i.d. from a Student-t distribution with $\nu = 10$ degrees of freedom rescaled to unit variance, so that the size of the shock can be interpreted as a one-standard-deviation shock.[5] Horizons range from $h = 0, 1, \ldots, 6$ for the first DGP

[4]This corresponds to the case where the state of the economy depends on lags of the outcome of interest as it is the case when studying state-dependent government spending multipliers.

[5]With $\delta = 1$ and the use of a t-distribution for generating ε_{1t}, the population Riesz representer $\alpha_0(x)$ is uniformly bounded, satisfying Assumption 2(i).

where the IRFs are less persistent, whereas we set $h = 0, 1, \ldots, 8$ for the second DGP. The true impulse responses are computed by simulation from the structural model using 5,000 counterfactual paths after a burn-in period of $T_0 = 500$. The DR-NLO estimator uses $K = 10$ ($K = 20$) contiguous blocks of equal length $\lfloor T/K \rfloor$, as described in Section 5 for the state-dependent (nonlinear regressor) model. In both cases, x_t is assumed predetermined and x_{t-1} and y_{t-1} are used as control variables. To estimate the conditional mean $g_{0,h}$, we use a nonparametric series estimator based on a Hermite polynomial with the total degree selected by the AIC up to a maximum value of 2.[6] The Riesz representer α_0 is estimated by the LASSO minimum-distance procedure of Chernozhukov et al. (2022), using the same Hermite dictionary at a fixed total degree of 2. HAC standard errors for the DR-NLO estimator are computed from the cross-fitted influence function as described in Section 5, using a Bartlett kernel and the Andrews (1991) automatic plug-in bandwidth. To prevent a small number of rare explosive Monte Carlo replications from dominating the results, we symmetrically trim the top and bottom 1% of the simulated IRF estimates.

7.1 Simulations for sign model

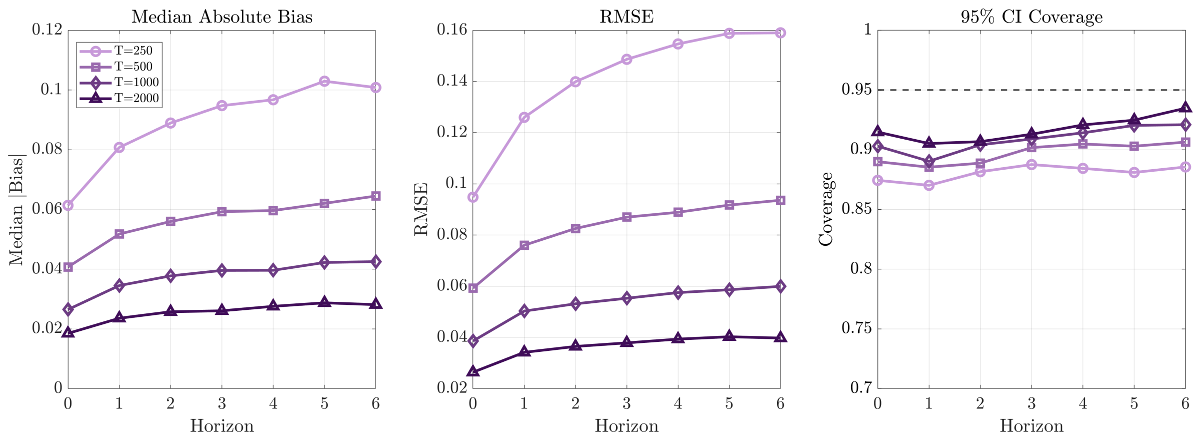
This section summarizes the simulation results for a design similar to Example 2.1 given by

$$x_t = 0.3x_{t-1} + 0.3y_{t-1} + \varepsilon_{1t}$$

$$y_t = 0.5x_t + 0.3x_{t-1} + 0.5y_{t-1} - 0.4\max(x_t, 0) - 0.3\max(x_{t-1}, 0) + \varepsilon_{2t}.$$

[6] While Theorem 5.1 is stated for general nonparametric estimators whose complexity grows to satisfy the rate conditions in Assumption 3, our numerical implementation utilizes a low-complexity approximation to manage the bias-variance tradeoff.

Figure 2: Performance of the DR-NLO estimator for the nonlinear regressor DGP



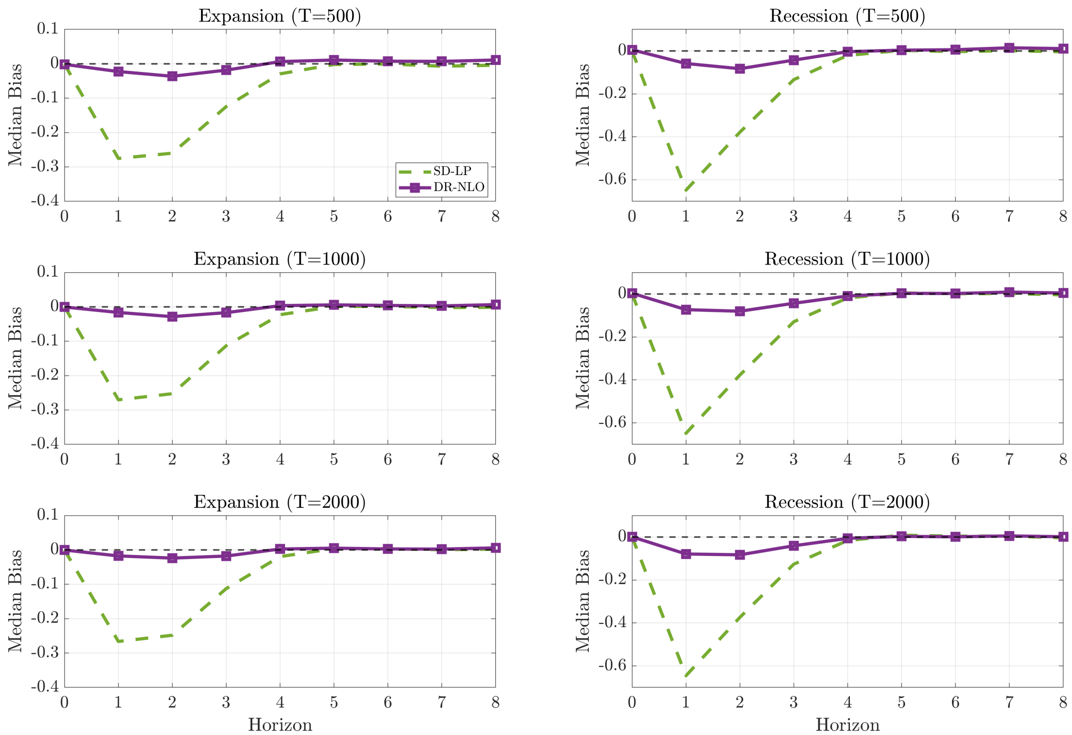
Note: Median Absolute Bias, RMSE and Coverage for different sample sizes.

For the sake of brevity and because simulations evaluating the performance of the plug-in and LP estimators in models with nonlinear regressors can be found in Gonçalves et al. (2021, 2024*a*), we focus on the DR-NLO estimator. We use $20,000$ replications per sample size, letting $T \in \{250, 500, 1000, 2000\}$. For each sample size, we report the median absolute bias, root mean squared error (RMSE) and nominal coverage of the 95% confidence intervals. The simulation results reported in Figure 2 indicate that DR-NLO performs well when the object of interest is the ARF in models with nonlinear regressors. The bias is small and close to zero across horizons and sample sizes, indicating that the estimator is approximately unbiased. RMSE decreases with T, as expected for a consistent estimator. Although coverage falls below the nominal 95%, it improves with increasing sample size. Overall, the results suggest that the estimator performs reasonably well in modestly large samples.

7.2 Simulations for state dependent model

We report simulation results for the state dependent model in Example 2.2, where x_t is predetermined so that $x_t = \rho_x x_{t-1} + \psi_x y_{t-1} + \varepsilon_{1t}$ and $S_{t-1} = \mathbf{1}\{y_{t-1} > 0\}$ classifies the previous period as an expansion ($S_{t-1} = 1$) or recession ($S_{t-1} = 0$). The parameters are set to $\rho_x = 0.3$, $\psi_x = -0.1$, $\beta_E = 2.5$, $\beta_R = 3.5$,$\gamma_E = 0.9$, and $\gamma_R = -0.1$. For each replication, we apply each estimator separately to the expansion ($S_{t-1} = 1$) and recession ($S_{t-1} = 0$) sub-samples. We compare our estimator to the widely used state-dependent local projection (SD-LP), implemented by OLS with state-by-covariate interactions and one lag of (x, y) as controls. The Monte Carlo design uses 30,000 replications per sample size. For each estimator, sample size, state (expansion and recession) and horizon, we report the empirical median bias, RMSE and the coverage of nominal 95% confidence intervals across replications.[7]

Figure 3: Median Bias

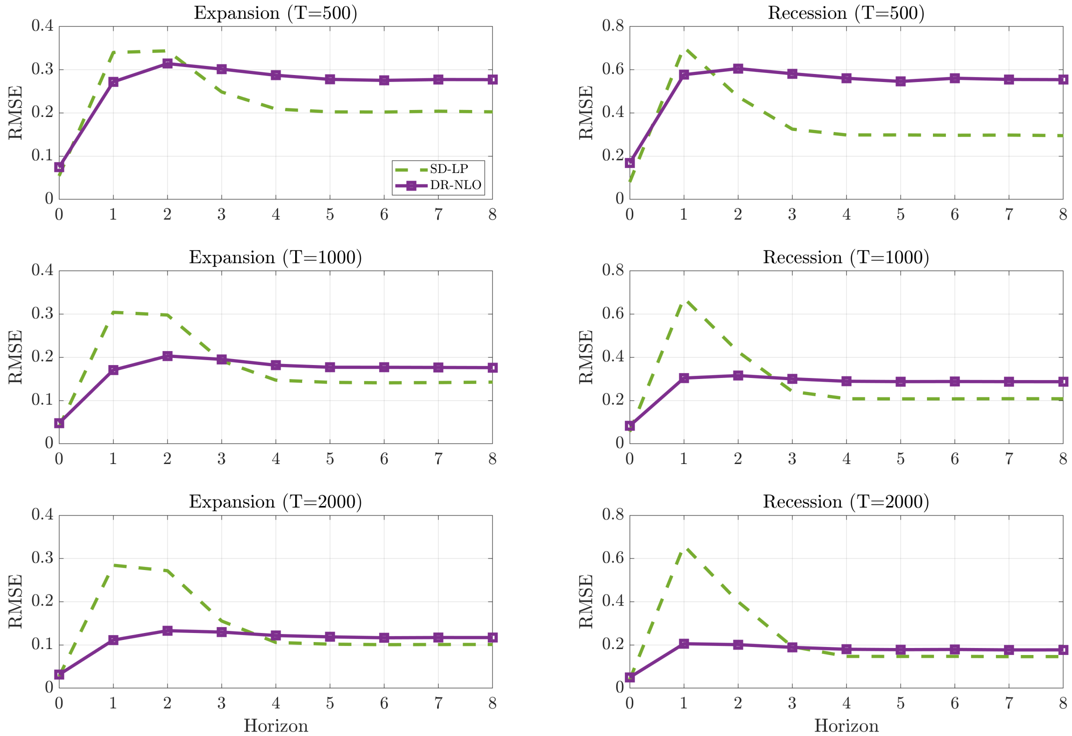


Note: Median bias of DR-NLO and LP estimators for different sample sizes and $\delta = 1$.

Figure 3 illustrates the systematic downward bias of the traditional state-dependent LP (SD-LP) estimator in both expansions and recessions, which does not vanish as the sample size grows. As explained in Gonçalves et al. (2024*b*), when the state is endogenous, the SD-LP specification does not recover the conditional average response to a fixed size shock because state-dependent dynamics in $(\beta_{t-1}, \gamma_{t-1})$ introduce nonlinearities that are not accounted for by the regression. This is not a small sample issue, but a feature of the SD-LP itself. By contrast, the semiparametric DR-NLO estimator has smaller median bias at every horizon and for every sample size considered, and its bias decreases with T, as predicted

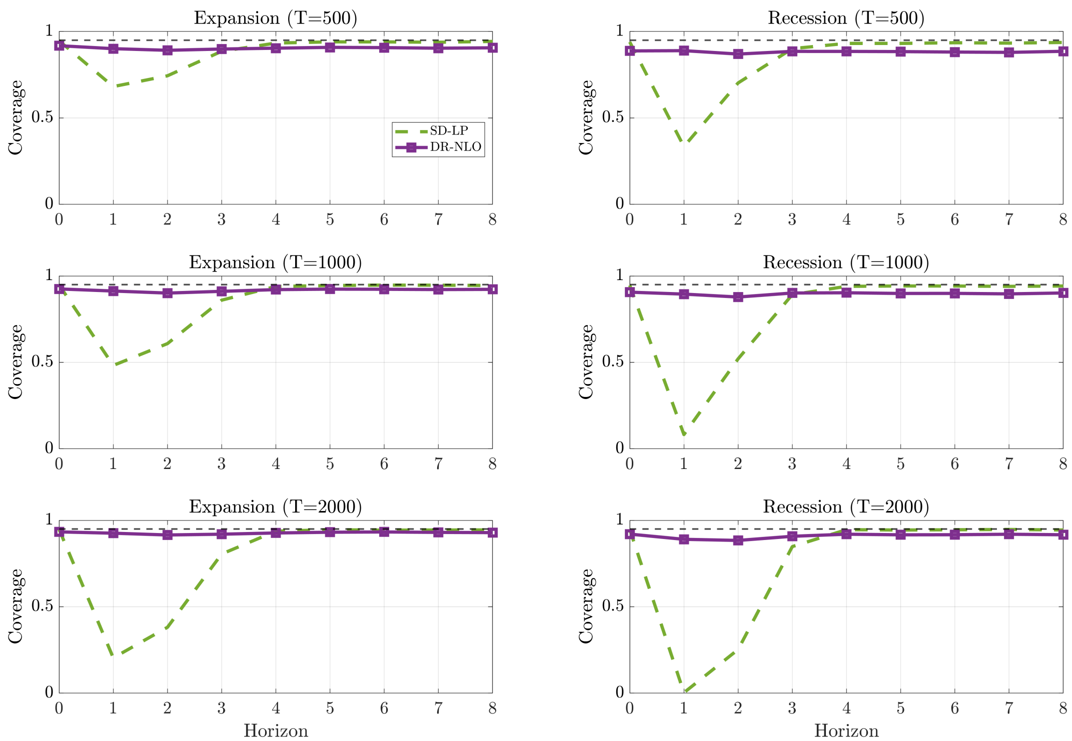
[7]In this example, we report median bias instead of median absolute bias in order to highlight the direction of the bias, namely that the LP estimator is negatively biased.

Figure 4: Root Mean Squared Error (RMSE)



Note: RMSE of DR-NLO and LP estimators for different sample sizes and $\delta = 1$.

Figure 5: Empirical coverage rates of 95% confidence intervals



Note: Coverage rates of 95% intervals based on DR-NLO and SD-LP for different sample sizes and $\delta = 1$.

by Theorem 5.1. The substantial bias of SD-LP at shorter horizons explains why its RMSE is greater than that of DR-NLO, despite the latter having slightly higher variance. At longer horizons, the bias of the SD-LP declines, resulting in a lower RMSE. As T increases, differences in RMSE between the two estimators vanish at longer horizons, with DR-NLO dominating SD-LP at shorter horizons.

Figure 5 shows the empirical coverage probabilities of 95% confidence intervals based on SD-LP and DR-NLO, both using HAC variance estimators with a Bartlett kernel and Andrews (1991)'s automatic bandwidth. The SD-LP intervals exhibit substantial undercoverage for all sample sizes, which worsens as T increases since the variance shrinks while the bias persists. DR-NLO slightly undercovers as well, but its coverage remains closer to the nominal 95% level and improves with T.

8 Empirical illustrations

Semiparametric LP estimators such as the DR-NLO estimator are useful not only for capturing nonlinearities when the functional form of the nonlinearity is unknown, as illustrated by our simulation evidence, but also as a diagnostic tool for judging the adequacy of linear approximations. To illustrate how our semiparametric LP estimator may be used to assess the adequacy of the linear LP estimator, we apply both methods to study the pass-through from retail gasoline price shocks to inflation in the United States.[8] This has been a question of continued policy interest, especially in recent years. There has been a proliferation of research addressing this question using linear VAR and distributed lag models (e.g., Chudik and Georgiadis (2022); Kilian and Zhou (2022, 2025)). The use of a semiparametric LP estimator is natural in this context since it has long been suspected that this pass-through may be nonlinear. A particular concern is that the pass-through may be stronger when the pre-existing level of inflation is higher. This concern has become particularly relevant in recent months, with the surge in headline inflation following the outbreak of the Iran War

[8]The reason the recent literature has focused on retail gasoline prices rather than the price of crude oil is that the relationship between crude oil and retail gasoline prices is unstable over time, reflecting large variation in the cost share of crude oil in the retail price of gasoline (see Kilian and Zhou (2022)).

in late February 2026, but similar concerns already arose during the 2022 surge in inflation.

While this question has been addressed by a number of studies, such as Clark and Terry (2010), Gründler (2024) or De Santis and Tornese (2025), these studies are based on parametric nonlinear models such as threshold VAR models, regime-switching models, or time-varying coefficient VAR models. Two obvious concerns are that these specifications are mutually exclusive and that they do not exhaust the range of possible nonlinear specifications. Our semiparametric approach allows us to dispense with these parametric restrictions. We allow the effect of gasoline price shocks (ε_{1t}) on inflation to depend on lagged headline and core inflation, according to the model:

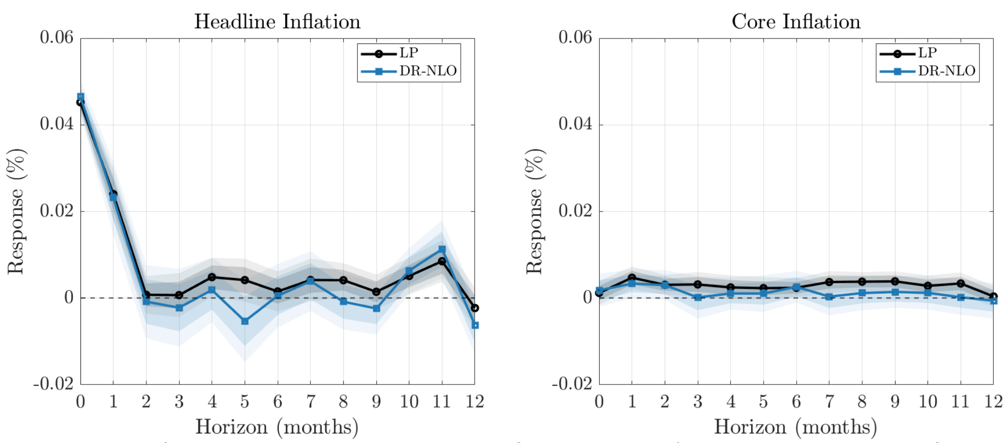
$$
\begin{aligned}
x_t &= \phi(\mathbf{z}_{t-1}) + \varepsilon_{1t} \\
y_{1t} &= \mu_1(x_t, \mathbf{z}_{t-1}, \varepsilon_{2t}) \\
y_{2t} &= \mu_2(x_t, \mathbf{z}_{t-1}, \varepsilon_{3t}),
\end{aligned}
$$

where $\mathbf{z}_{t-1}$ contains six lags of the percent change in retail gasoline prices (x_t), headline inflation (y_{1t}) and core CPI inflation for all urban consumers (y_{2t}). Core inflation is defined as inflation excluding food and energy. All data are monthly and seasonally adjusted. The data source is FRED. The model incorporates six lags consistent with other recent empirical studies. We are interested in the inflation responses at horizon $0, 1, \ldots, 12$. The estimation sample spans January 1974 through April 2026, which includes several high-inflation episodes.

We first compare the linear LP and DR-NLO estimates of the unconditional responses of headline and core inflation to a one percent gasoline price shock, using the same learners as in the Monte Carlo simulations and $K = 10$. Evidence that the semiparametric LP estimates are very different from the linear LP estimates would cast doubt on prior linear estimates of the pass-through to inflation. Evidence that the two estimates are close, in contrast, would reassure policymakers that existing LP and VAR estimates based on linear approximations are informative. Figure 6 illustrates that the choice of the estimator matters little. It shows point estimates of the unconditional impulse responses and 68% and 90%

pointwise confidence bands. Both estimates indicate that headline inflation jumps by close to 0.05 percentage points (not annualized) in response to a one percentage point gasoline price shock, but the response declines quickly. By horizon 2, it is close to zero. There is no evidence that headline inflation rises persistently in response to gasoline price shocks. The response of core inflation is an order of magnitude smaller and only slightly positive. Both the linear LP and the DR-NLO estimate are marginally statistically significant at horizons 1 and 2. Overall, this evidence suggests that linear approximations are adequate for assessing the unconditional response of inflation to gasoline price shocks.[9]

Figure 6: Average Responses of Headline and Core Inflation to Gasoline Price Shocks



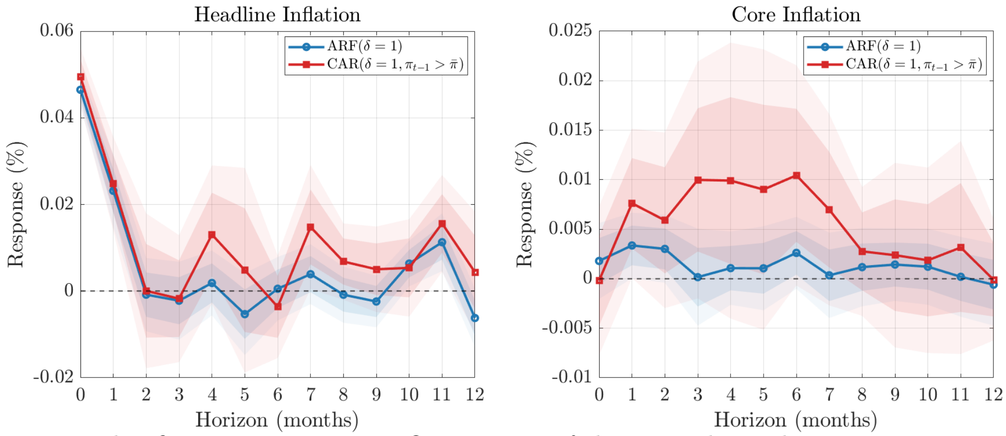
Notes: This figure reports LP and DR-NLO estimates of headline and core inflation to a one percentage point ($\delta = 1$) increase in the price of gasoline. The shaded areas represent 68% and 90% confidence intervals.

Next, we turn to the question of whether there is evidence of larger inflation responses conditional on the annualized inflation exceeding its historical average of 3.8% (i.e., 0.32% monthly) during the estimation period. Figure 7 shows the unconditional response and the response conditional on inflation being above average based on the DR-NLO estimator. There is no material change in the headline and core inflation response estimates at horizons 0, 1 and 2. At longer horizons, the conditional responses tend to be larger, but also less precisely estimated. Even a user of the conditional response estimate would be unable to conclude that there are statistically significant increases in inflation in response to gasoline price shocks, however. Nor are the conditional point estimates consistent with gasoline price

[9] Very similar results would have been obtained based on a bivariate model for gasoline prices and headline inflation.

shocks causing persistent increases in headline or core inflation of a material magnitude.

Figure 7: Comparison of Average and Conditional Average Responses of Headline and Core Inflation to Gasoline Price Shocks



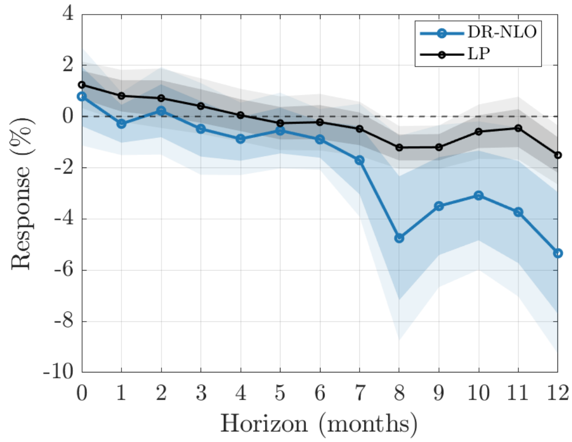
Notes: This figure reports DR-NLO estimates of the unconditional average responses of headline and core inflation to a one percentage point gasoline price shock ($ARF_h(\delta)$ with $\delta = 1$) and the corresponding average responses conditional on headline inflation being above its historical average in $t-1$ ($CAR_h(\delta, \pi_{t-1} > \bar{\pi})$ with $\delta = 1$). The shaded areas represent 68% and 90% confidence intervals.

We conclude with another example that illustrates that the impulse responses implied by the DR-NLO estimator may look materially different from those implied by the linear LP estimator and more economically plausible. The question of interest is how sales of U.S. motor vehicles respond to shocks to the real price of motor gasoline. This question is motivated by Edelstein and Kilian (2009) and Ramey and Vine (2006, 2011) who document potential nonlinearities in the response of motor vehicle sales. We rely on equations (1) and (2) with z_t containing the percent change in real U.S. retail gasoline prices and the percent change in U.S. sales of automobiles and light trucks. Retail gasoline prices are obtained from the CPI and deflated by the aggregate CPI. All data are monthly and seasonally adjusted. The data source is FRED. The model incorporates six lags. The estimation period is January 1974 through March 2026.

Figure 8 focuses on the responses of the cumulative growth rate to a one standard deviation shock in the real price of gasoline.[10] The DR-NLO impact response is positive but statistically insignificant. There is no statistically significant response in sales for the first three months, according to the DR-NLO estimator, followed by a persistent decline in auto

[10] The cumulative effect is estimated directly by applying the LP or DR-NLO to $\Delta^h y_{t+h} = y_{t+h} - y_{t-1}$.

Figure 8: Response of Motor Vehicle Sales to Real Gasoline Price Shocks ($\delta = 1$ s.d.)



Note: This figure reports LP and DR-NLO estimates of the cumulative response of motor vehicle sales to one standard deviation shock in the real price of gasoline. The shaded areas represent 68% and 90% confidence intervals.

sales starting at horizon 3. The linear LP estimate, in contrast, suggests a statistically significant but economically counterintuitive increase in auto sales at horizons 0, 1 and 2. More generally, the persistently positive linear LP responses for the first four months are difficult to reconcile with economic reasoning. The numerical differences between the estimates compound at longer horizons. At horizon 12, the linear LP estimate shows a decline in auto sales that is only 28% of the DR-NLO response. These differences suggest that linear approximations may not adequately capture the dynamics in question.

9 Concluding remarks

This paper developed a semiparametric local projection estimator of unconditional nonlinear impulse response functions for a broad class of structural dynamic models that are widely used in applied macroeconomics, including models with nonlinearly transformed regressors, state-dependent coefficients, and nonlinear interactions between shocks and state variables. Under standard mixing and rate conditions on the nuisance estimators, the resulting estimator is $\sqrt{T}$-consistent and asymptotically normal, with the preliminary estimation of the two nuisance functions having no effect on the first-order asymptotic distribution. Inference is conducted via a HAC long-run variance estimator applied to the estimated influence function. We also showed how our framework accommodates conditional impulse responses in state-dependent models, where the conditioning variable takes discrete values.

Our Monte Carlo evidence indicates that the proposed estimator delivers substantially smaller bias than standard state-dependent local projection methods, while preserving competitive root mean squared errors and confidence-interval coverage close to the nominal level. We considered two empirical illustrations. The first one focused on the potentially nonlinear pass-through from gasoline price shocks to inflation. The second example examined whether linear LP estimators miss nonlinearities in the transmission of gasoline price shocks to motor vehicle sales.

There are several avenues for further research. First, our identification result relies on the additive separability of ε_{1t} in equation (1). One possible extension would be to allow the shock to interact nonlinearly with lagged state variables. Second, our treatment of conditional impulse responses assumes a discrete conditioning variable, leaving the continuous-state case as discussed in Example 2.3 for future work.

References

Andrews, D. W. K. (1991), 'Heteroskedasticity and autocorrelation consistent covariance matrix estimation', *Econometrica* **59**, 817–858.

Ascari, G. and Haber, T. (2022), 'Non-linearities, state-dependent prices and the transmission mechanism of monetary policy', *Economic Journal* **132**, 37–57.

Ballarin, G. (2024), Impulse response analysis of structural nonlinear time series models. Manuscript, University of Mannheim.

Ballinari, D. and Wehrli, A. (2025), Semiparametric inference for impulse response functions using double/debiased machine learning. Manuscript, arXiv:2411.10009.

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

Cacciatore, M. and Ravenna, F. (2021), 'Uncertainty, wages and the business cycle', *Economic Journal* **131**, 2797–2823.

Caramp, N. and Feilich, E. (2026), 'Monetary policy and government debt', *Journal of Money, Credit and Banking* **58**, 389–420.

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

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W. and Robins, J. (2018), 'Double/debiased machine learning for treatment and structural parameters', *Econometrics Journal* **21**, C1–C68.

Chernozhukov, V., Newey, M., Newey, W. K., Singh, R. and Syrgkanis, V. (2025), 'Automatic debiased machine learning for covariate shifts'.
URL: *https://arxiv.org/abs/2307.04527*

Chernozhukov, V., Newey, W. and Singh, R. (2022), 'Automatic debiased machine learning of causal and structural effects', *Econometrica* **90**, 967–1027.

Chudik, A. and Georgiadis, G. (2022), 'Estimation of impulse response functions when shocks are observed at a higher frequency than outcome variables', *Journal of Business & Economic Statistics* **40**, 965–979.

Clark, T. E. and Terry, S. J. (2010), 'Time variation in the inflation passthrough of energy prices', *Journal of Money, Credit and Banking* **42**, 1419–1433.

Cloyne, J., Jordà, Ò. and Taylor, A. M. (2020), Decomposing the fiscal multiplier, Working Paper 26939, National Bureau of Economic Research.

Cloyne, J., Jordà, Ò. and Taylor, A. M. (2023), State-dependent local projections: Understanding impulse response heterogeneity, Working Paper 30971, National Bureau of Economic Research.

Davidson, J. (1994), *Stochastic Limit Theory*, Oxford University Press, Oxford.

De Santis, R. A. and Tornese, T. (2025), 'Energy supply shocks' nonlinearities on output and prices', *European Economic Review* **176**, 105037.

Edelstein, P. and Kilian, L. (2009), 'How sensitive are consumer expenditures to retail energy prices?', *Journal of Monetary Economics* **56**, 766–779.

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

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., Kilian, L. and Pesavento, E. (2021), 'Impulse response analysis for structural dynamic models with nonlinear regressors', *Journal of Econometrics* **225**, 107–130.

Gonçalves, S., Herrera, A., Kilian, L. and Pesavento, E. (2024*a*), Nonparametric local projections. Manuscript, Federal Reserve Bank of Dallas.

Gonçalves, S., Herrera, A., Kilian, L. and Pesavento, E. (2024*b*), 'State-dependent local projections', *Journal of Econometrics* **244**.

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

Gründler, D. (2024), 'Does the inflation pass-through of gasoline price shocks depend on the level of inflation?', *Economics Letters* **243**, 111928.

Herrera, A., Lagalo, L. and 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.

Huang, J., Jordà, O., Marcellino, M. and Tornese, T. (2026), High-dimensional nonparametric local projections. Manuscript, Bocconi University.

Kilian, L. and Vigfusson, R. (2011), 'Are the responses of the U.S. economy asymmetric in energy price increases and decreases?', *Quantitative Economics* **2**, 419–453.

Kilian, L. and Zhou, X. (2022), 'Oil prices, gasoline prices, and inflation expectations', *Journal of Applied Econometrics* **37**, 867–881.

Kilian, L. and Zhou, X. (2025), Oil price shocks and inflation, *in* 'Research Handbook on Inflation', Edward Elgar Publishing, pp. 420–438.

Kolesár, M. and Plagborg-Møller, M. (2025), 'Dynamic causal effects in a nonlinear world: The good, the bad, and the ugly', *Journal of Business & Economic Statistics* **43**, 737–754.

Koop, G., Pesaran, M. and Potter, S. (1996), 'Impulse response analysis in nonlinear multivariate models', *Journal of Econometrics* **74**, 119–147.

Newey, W. (1994), 'Series estimation of regression functionals', *Econometric Theory* **10**, 1–28.

Nikolaishvili, G. (2026), Doubly robust nonparametric local projections. Manuscript, Wake Forest University.

Potter, S. (2000), 'Nonlinear impulse response functions', *Journal of Economic Dynamics and Control* **24**, 1425–1446.

Rambachan, A. and Shephard, N. (2021), 'When do common time series estimands have nonparametric causal meaning?', *Manuscript, arXiv:1903.01637* .

Ramey, V. A. and Vine, D. J. (2006), 'Declining volatility in the U.S. automobile industry', *American Economic Review* **96**, 1876–1889.

Ramey, V. A. and Vine, D. J. (2011), 'Oil, automobiles, and the U.S. economy: How much have things really changed?', *NBER Macroeconomics Annual* **25**, 333–368.

Ramey, V. and 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.

Semenova, V., Goldman, M., Chernozhukov, V. and Taddy, M. (2023), 'Inference on heterogeneous treatment effects in high-dimensional dynamic panels under weak dependence', *Quantitative Economics* **14**, 471–510.

Stock, J. (1989), 'Nonparametric policy analysis', *Journal of the American Statistical Association* **84**, 567–575.

Tenreyro, S. and Thwaites, G. (2016), 'Pushing on a string: U.S. monetary policy is less powerful in recessions', *American Economic Journal: Macroeconomics* **8**, 43–74.

Appendix

A Proof of results in Section 4

Proof of Proposition 4.1. By (1) and (2), we can rewrite $y_{t+h} = m_h(x_t, U_{t+h})$, for some function m_h that maps x_t (which we observe) and U_{t+h} (defined as in the text) into y_{t+h}. Although x_t is not independent of U_{t+h} (unless $x_t = \varepsilon_{1t}$), the independence condition $\varepsilon_{1t} \perp U_{t+h}$ implies the conditional independence assumption $x_t \perp U_{t+h} \mid \mathbf{z}_{t-1}$. This is the key identifying condition: conditionally on $\mathbf{z}_{t-1}$, all remaining variation in x_t comes from ε_{1t} alone, which is independent of U_{t+h}. Because the model for x_t given in (1) is additive in ε_{1t}, a δ-shift in x_t holding $\mathbf{z}_{t-1}$ fixed is identical to a δ-shift in ε_{1t}, so that $\theta_{0,h} \equiv \text{ARF}_h(\delta)$ in Definition 1 can equivalently be written as $\theta_{0,h} = E[m_h(x_t + \delta, U_{t+h}) - m_h(x_t, U_{t+h})]$. But for any fixed x and z, $g_{0,h}(x, z) \equiv E(y_{t+h} \mid x_t = x, \mathbf{z}_{t-1} = z) = E[m_h(x, U_{t+h}) \mid \mathbf{z}_{t-1} = z]$ where the second equality uses the fact that U_{t+h} is independent of x_t, conditionally on $\mathbf{z}_{t-1}$. The desired result follows by applying the law of iterated expectations (LIE). ■

Proof of Proposition 4.2. By (5) with $g_h = g_{0,h}$, we have $E[g_{0,h}(x_t + \delta, \mathbf{z}_{t-1}) - g_{0,h}(x_t, \mathbf{z}_{t-1})] = E[\alpha_0(x_t, \mathbf{z}_{t-1}) g_{0,h}(x_t, \mathbf{z}_{t-1})] = \theta_{0,h}$. Substituting $\theta_h = \theta_{0,h}$ in (6) and using $E[\alpha_0(x_t, \mathbf{z}_{t-1})(y_{t+h} - g_{0,h}(x_t, \mathbf{z}_{t-1}))] = 0$, which follows by LIE since $E[y_{t+h} - g_{0,h}(x_t, \mathbf{z}_{t-1}) | x_t, \mathbf{z}_{t-1}] = 0$, gives the result. ■

Proof of Proposition 6.1. The proof follows by the same arguments as the proof of Proposition 4.1, with $\mathbf{z}_{t-1}$ replaced by $(\mathbf{z}_{t-1}, \Omega_t)$ throughout, and applying the law of iterated expectations (LIE) conditioning additionally on $\Omega_t = \omega$. ■

B Proofs of results in Section 5

To prove our results, we rely on Lemma A.6 of Semenova et al. (2023). For completeness, we describe this result next.

Lemma B.1 *Let $A(z_t, \eta)$ be some generic function of data z_t and nuisance function $\eta \in \mathcal{H}_T$, where $\mathcal{H}_T$ is a shrinking neighborhood of η_0 which contains the estimator $\hat{\eta}_\ell$ with probability converging to one for all $\ell = 1, \ldots, K$. Suppose that $\{z_t\}$ is geometrically β-mixing and define*

$$B_\ell(\eta) \equiv \frac{1}{T_\ell} \sum_{t \in I_\ell} E[A(z_t, \eta)] \quad \textit{and} \quad V_\ell(\eta) \equiv \frac{1}{T_\ell} \sum_{t \in I_\ell} [A(z_t, \eta) - E(A(z_t, \eta))].$$

If for any non-stochastic sequence $\eta_T \in \mathcal{H}_T$ and any norm $\|\cdot\|$,

$$\|B_\ell(\eta_T)\| = O(\xi_{1T}) \tag{A.1}$$

$$\|V_\ell(\eta_T)\| = O_p(\xi_{2T}), \tag{A.2}$$

then $\left\| T^{-1} \sum_{\ell=1}^{K} \sum_{t \in I_\ell} A(z_t, \hat{\eta}_\ell) \right\| = O_p(\xi_{1T} + \xi_{2T})$.

Proof of Theorem 5.1. Let $\tilde{\theta}_h$ denote the oracle estimator that uses the true conditional mean function $g_{0,h}(x_t, \mathbf{z}_{t-1})$ and the true Riesz representer $\alpha_0(x_t, \mathbf{z}_{t-1})$. We decompose

$$\sqrt{T}(\hat{\theta}_h - \theta_{0,h}) = \sqrt{T}(\tilde{\theta}_h - \theta_{0,h}) + \sqrt{T}(\hat{\theta}_h - \tilde{\theta}_h).$$

The first term satisfies $\sqrt{T}(\tilde{\theta}_h - \theta_{0,h}) \to_d N(0, V_h)$ by a CLT for weakly dependent time series (e.g., Corollary 24.7 of Davidson (1994), since β-mixing implies strong mixing), using Assumptions 1 and 2(iii). It remains to show that $\sqrt{T}(\hat{\theta}_h - \tilde{\theta}_h) = o_p(1)$.

Writing $\hat{\theta}_h - \tilde{\theta}_h = R_1 + R_2 + R_3$ with

$$R_1 \equiv \frac{1}{T}\sum_{\ell=1}^{K}\sum_{t\in I_\ell}\big(m(x_t, \mathbf{z}_{t-1}, \hat{g}_\ell - g_0) + \alpha_0(x_t, \mathbf{z}_{t-1})(g_0(x_t, \mathbf{z}_{t-1}) - \hat{g}_\ell(x_t, \mathbf{z}_{t-1}))\big),$$
$$R_2 \equiv \frac{1}{T}\sum_{\ell=1}^{K}\sum_{t\in I_\ell}(\hat{\alpha}_\ell(x_t, \mathbf{z}_{t-1}) - \alpha_0(x_t, \mathbf{z}_{t-1}))(y_{t+h} - g_0(x_t, \mathbf{z}_{t-1})),$$
$$R_3 \equiv \frac{1}{T}\sum_{\ell=1}^{K}\sum_{t\in I_\ell}(\hat{\alpha}_\ell(x_t, \mathbf{z}_{t-1}) - \alpha_0(x_t, \mathbf{z}_{t-1}))(g_0(x_t, \mathbf{z}_{t-1}) - \hat{g}_\ell(x_t, \mathbf{z}_{t-1})),$$

where $m(x_t, \mathbf{z}_{t-1}, \hat{g}_\ell - g_0) \equiv (\hat{g}_\ell(x_t+\delta, \mathbf{z}_{t-1}) - \hat{g}_\ell(x_t, \mathbf{z}_{t-1})) - (g_0(x_t+\delta, \mathbf{z}_{t-1}) - g_0(x_t, \mathbf{z}_{t-1}))$, we show that $\sqrt{T}R_j = o_p(1)$ for $j = 1, 2, 3$. Since $\{z_t\}$ is geometrically β-mixing (Assumption 1) and I_ℓ and I_ℓ^{qc} are separated by at least T_ℓ time periods, the hypothesis of Lemma B.1 is satisfied, and it suffices to verify conditions (A.1) and (A.2) for non-stochastic sequences $\eta_T \in \mathcal{G}_T \times \mathcal{A}_T$. Without loss of generality we take $\ell = 1$, so that $t \in I_\ell$ corresponds to $t = 1, \ldots, T_1$; we nonetheless retain the generic index ℓ in the sums below. We define $\xi_t \equiv (x_t, \mathbf{z}_{t-1})$ throughout.

We start with R_1 and apply Lemma B.1 with

$$A(\xi_t, \eta) = m(x_t, \mathbf{z}_{t-1}, g - g_0) + \alpha_0(x_t, \mathbf{z}_{t-1})(g_0(x_t, \mathbf{z}_{t-1}) - g(x_t, \mathbf{z}_{t-1})),$$

where $\eta = g$. By Neyman orthogonality, $E[A(\xi_t, \eta_T)] = 0$ for any $g_T \in \mathcal{G}_T$, so $B_\ell(\eta_T) = 0$ and condition (A.1) holds trivially. For condition (A.2), we show $\mathrm{Var}(V_\ell(\eta_T)) = o(T_\ell^{-1})$. By stationarity of z_t (and hence of ξ_t),

$$\mathrm{Var}(V_\ell(\eta_T)) = \frac{1}{T_\ell}E[A(\xi_t, \eta_T)^2] + \frac{2}{T_\ell^2}\sum_{j=1}^{T_\ell-1}(T_\ell - j)\,\mathrm{Cov}(A(\xi_t, \eta_T), A(\xi_{t-j}, \eta_T)).$$

For the first term, we have

$$E[A(\xi_t, \eta_T)^2] \le 2\left(E[m(x_t, \mathbf{z}_{t-1}, g_T - g_0)^2] + E[\alpha_0(x_t, \mathbf{z}_{t-1})^2(g_T(x_t, \mathbf{z}_{t-1}) - g_0(x_t, \mathbf{z}_{t-1}))^2]\right).$$

Using Assumption 2(i) and a change of variables,

$$
\begin{aligned}
&E[m(x_t, \mathbf{z}_{t-1}, g_T - g_0)^2] \\
&\le 2\left(\int \left(\int (g_T(x_t, \mathbf{z}_{t-1}) - g_0(x_t, \mathbf{z}_{t-1}))^2 \left(\frac{f_0(x_t - \delta|\mathbf{z}_{t-1})}{f_0(x_t|\mathbf{z}_{t-1})} + 1\right) f_0(x_t|\mathbf{z}_{t-1}) dx_t\right) f_0(\mathbf{z}_{t-1}) d\mathbf{z}_{t-1}\right) \\
&\le 2(\bar{\alpha} + 2) r_{g,T}^2 = o(1),
\end{aligned}
$$

where $f_0(\mathbf{z}_{t-1})$ denotes the marginal density of $\mathbf{z}_{t-1}$, the last inequality uses $\sup_{x,z} |\alpha_0(x, z)| < \bar{\alpha}$ and the last equality uses Assumption 3(i). An analogous argument gives $E[\alpha_0(x_t, \mathbf{z}_{t-1})^2 (g_T(x_t, \mathbf{z}_{t-1}) - g_0(x_t, \mathbf{z}_{t-1}))^2] \le \bar{\alpha}^2 r_{g,T}^2 = o(1)$, so $T_\ell^{-1} E[A(\xi_t, \eta_T)^2] = o(T_\ell^{-1})$. For the second term, $A(\xi_t, \eta_T)$ inherits the geometric β-mixing of $\{z_t\}$, and since β-mixing implies strong mixing, by Corollary 14.3 of Davidson (1994), for $q > 2$ as in Assumption 2(ii),

$$
|\mathrm{Cov}(A(\xi_t, \eta_T), A(\xi_{t-j}, \eta_T))| \le C\, \beta(j)^{1-2/q} \|A(\xi_t, \eta_T)\|_q^2.
$$

By the triangle inequality and Assumption 2(i), $|A(\xi_t, \eta_T)| \le |m(x_t, \mathbf{z}_{t-1}, g_T - g_0)| + \bar{\alpha}\, |g_T(x_t, \mathbf{z}_{t-1}) - g_0(x_t, \mathbf{z}_{t-1})|$, so $\|A(\xi_t, \eta_T)\|_q \le C(\bar{\alpha})\, r_{g,q,T} = o(1)$ by Assumption 3(i). Since $\beta(j) \le C e^{-c_\beta j}$ with $c_\beta > 0$, summing over j gives

$$
\frac{2}{T_\ell^2} \sum_{j=1}^{T_\ell - 1} (T_\ell - j) |\mathrm{Cov}(A(\xi_t, \eta_T), A(\xi_{t-j}, \eta_T))| \le \frac{C(\bar{\alpha})\, r_{g,q,T}^2}{T_\ell} \sum_{j=1}^{\infty} e^{-c_\beta j (1-2/q)} = o(T_\ell^{-1}),
$$

where $\sum_{j=1}^{\infty} e^{-c_\beta j(1-2/q)} < \infty$ because $c_\beta(1 - 2/q) > 0$ for $c_\beta > 0$ and $q > 2$, and $r_{g,q,T}^2 = o(1)$ by Assumption 3(i). Combining the two bounds gives $\mathrm{Var}(V_\ell(\eta_T)) = o(T_\ell^{-1})$, so $\sqrt{T} R_1 = o_p(1)$.

We next consider R_2. We apply Lemma B.1 with $A(\xi_t, \eta_T) = (\alpha_T(x_t, \mathbf{z}_{t-1}) - \alpha_0(x_t, \mathbf{z}_{t-1})) e_{t+h}$, where $e_{t+h} \equiv y_{t+h} - g_0(x_t, \mathbf{z}_{t-1})$. Since $E(e_{t+h}|x_t, \mathbf{z}_{t-1}) = 0$, an application of the LIE gives $E[A(\xi_t, \eta_T)] = 0$, so $B_\ell(\eta_T) = 0$ and condition (A.1) holds trivially. For condition (A.2), let $f_t \equiv (\alpha_T(x_t, \mathbf{z}_{t-1}) - \alpha_0(x_t, \mathbf{z}_{t-1})) e_{t+h}$. By stationarity, $\mathrm{Var}(V_\ell(\eta_T)) = \frac{1}{T_\ell} E(f_t^2) + \frac{2}{T_\ell^2} \sum_{j=1}^{T_\ell - 1} (T_\ell -$

$j)\,\mathrm{Cov}(f_t, f_{t-j})$. For the variance, by LIE and Assumption 2(ii),

$$E(f_t^2) = E\left[(\alpha_T(x_t, \mathbf{z}_{t-1}) - \alpha_0(x_t, \mathbf{z}_{t-1}))^2 E(e_{t+h}^2 | x_t, \mathbf{z}_{t-1})\right] \leq \bar{\sigma}_q^2\, r_{\alpha,T}^2 = o(1),$$

by Assumption 3(i) (since $r_{\alpha,T} = r_{\alpha,2,T} \leq r_{\alpha,q,T}$ by Hölder's inequality), so $T_\ell^{-1} E(f_t^2) = o(T_\ell^{-1})$. For the autocovariance terms, f_t is a function of (y_{t+h}, ξ_t) and hence inherits the geometric β-mixing of $\{z_t\}$. By the same covariance inequality as above, $|\mathrm{Cov}(f_t, f_{t-j})| \leq C\,\beta(j)^{1-2/q} \|f_t\|_q^2$. By LIE and Assumption 2(ii), $E[|f_t|^q] \leq \bar{\sigma}_q^q\, r_{\alpha,q,T}^q$, so $\|f_t\|_q \leq \bar{\sigma}_q\, r_{\alpha,q,T} = o(1)$ by Assumption 3(i). Summing over lags gives

$$\frac{2}{T_\ell^2} \sum_{j=1}^{T_\ell - 1} (T_\ell - j) |\mathrm{Cov}(f_t, f_{t-j})| \leq \frac{C\,\bar{\sigma}_q^2\, r_{\alpha,q,T}^2}{T_\ell} \sum_{j=1}^{\infty} e^{-c_\beta j (1-2/q)} = o(T_\ell^{-1}),$$

by Assumption 3(i). Combining both bounds gives $\mathrm{Var}(V_\ell(\eta_T)) = o(T_\ell^{-1})$, so $\sqrt{T} R_2 = o_p(1)$.

Finally, we consider R_3. We apply Lemma B.1 with

$$A(\xi_t, \eta) = (\alpha(x_t, \mathbf{z}_{t-1}) - \alpha_0(x_t, \mathbf{z}_{t-1}))(g(x_t, \mathbf{z}_{t-1}) - g_0(x_t, \mathbf{z}_{t-1}))$$

where $\eta = (\alpha, g)$. For the bias term, by Cauchy-Schwarz,

$$|E[A(\xi_t, \eta_T)]| \leq \big(E[(\alpha_T(x_t, \mathbf{z}_{t-1}) - \alpha_0(x_t, \mathbf{z}_{t-1}))^2]\big)^{1/2} \big(E[(g_T(x_t, \mathbf{z}_{t-1}) - g_0(x_t, \mathbf{z}_{t-1}))^2]\big)^{1/2},$$

which is bounded by $r_{\alpha,T}\, r_{g,T} = o(T^{-1/2})$, given Assumption 3(ii). So $B_\ell(\eta_T) = o(T^{-1/2})$ and condition (A.1) is satisfied. For the stochastic term, by stationarity, the triangle inequality, and Cauchy-Schwarz, $E|V_\ell(\eta_T)| \leq 2E|A(\xi_t, \eta_T)| \leq 2r_{\alpha,T}\, r_{g,T} = o(T^{-1/2})$, by Assumption 3(ii), so condition (A.2) holds by Markov's inequality. This completes the proof that $\sqrt{T} R_3 = o_p(1)$. ■