High-Dimensional Nonparametric Local Projections*

Jiaming Huang[1], Òscar Jordà[2], Massimiliano Marcellino[3], and Tommaso Tornese[4]

[1]The Hong Kong University of Science and Technology

[2]Federal Reserve Bank of San Francisco, UC Davis, and CEPR

[3]Bocconi University, Baffi, IGIER and CEPR

[4]Università Cattolica del Sacro Cuore

March 25, 2026

Abstract. This paper develops a flexible two-step high-dimensional nonparametric local projections estimator (HDNLP) by constructing Neyman-orthogonal pseudo-outcomes whose expectation conditional only on low-dimensional policy variables identifies the average structural function. In the first step, we construct pseudo-outcomes with high-dimensional nuisance parameters using flexible machine learning tools and cross-fitting. In the second step, we estimate the conditional expectation using local linear estimation, allowing for flexible inference on potentially nonlinear impulse responses. We establish asymptotic properties of the HDNLP estimator at standard nonparametric rates and demonstrate superior finite-sample performance relative to existing methods in simulations and in an application on the nonlinear effects of financial conditions on the economy.

Keywords: Impulse responses, high-dimensional data, debiased/orthogonal estimation, machine learning, nonparametric estimation, local projections.
JEL: C14, C32, C55

*Contact: jiaminghuang@ust.hk. We would like to thank Elena Pesavento, Silvia Gonçalves, Karin Klieber, and Juan Carlos Escanciano for their helpful comments. The views expressed in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Federal Reserve Bank of San Francisco, or the Federal Reserve System.

1 Introduction

This paper proposes a two-step nonparametric approach for estimating impulse response functions under unconfoundedness (Angrist and Kuersteiner, 2011; Angrist et al., 2018). In the first step, we reduce the dimension of the problem by constructing pseudo-outcomes whose conditional expectation—conditional only on low-dimensional policy variables—identifies the *average structural function* (ASF) (Imbens and Newey, 2009). In the second step, the ASF is estimated using standard nonparametric methods such as local polynomial or sieve estimators, yielding flexible estimates of nonlinear impulse responses. The impulse response function is then flexibly obtained as the difference in the ASF evaluated at two distinct policy levels.

The estimation of dynamic causal effects is often formulated in high-dimensional settings. For example, to address endogeneity concerns, researchers often condition on a large information set arising from disaggregate macro variables (McCracken and and Ng, 2016), high sampling frequency (Babii et al., 2022), or various sources of private information (Romer and Romer, 2004). Moreover, the dynamic causal effects—more commonly referred to as impulse response functions—are often themselves high-dimensional to accommodate nonlinearities (Barnichon and Matthes, 2018).

To avoid the curse of dimensionality, existing approaches often impose restrictions such as linearity, sparsity, or particular parametric forms of nonlinearity. For instance, Adamek et al. (2024) estimate a linear local projection model with ℓ_1 regularization. When modeling nonlinear impulse responses, Barnichon and Matthes (2018) consider Gaussian basis approximations, whereas Gonçalves et al. (2021) assume knowledge of the exact functional form of nonlinearity. However, such practices risk introducing bias under functional form misspecification.[1]

The dimension reduction strategy that we propose is particularly valuable in macroeco-

[1]See Giannone et al. (2021) for a discussion of the limitations of sparsity assumptions and Kolesár and Plagborg-Møller (2025) for critiques of parametric nonlinear models.

nomic applications, where even carefully designed shock measures are rarely exogenous (see Ramey, 2016, for a review). For example, the fed funds futures surprises around FOMC announcements constructed by Gertler and Karadi (2015) exhibit autocorrelation and are predictable by a range of macroeconomic variables (Nakamura and Steinsson, 2018; Miranda-Agrippino and Ricco, 2021). Addressing this endogeneity typically requires conditioning on an extensive collection of macroeconomic controls, which quickly render fully nonparametric estimation computationally infeasible. Even when feasible, slow convergence rates due to the curse of dimensionality often induce large bias in the final impulse response estimator.

Our construction of pseudo-outcomes builds on recent advances in double/debiased machine learning, which combines Neyman-orthogonal moment conditions with cross-fitting to achieve robustness and control for first-step estimation bias (Chernozhukov et al., 2018; Semenova and Chernozhukov, 2021; Colangelo and Lee, 2026). The pseudo-outcomes depend on two key components, both of which can be flexibly estimated using modern machine learning methods such as random forests, boosting, and neural networks.

The first component is the conditional expectation function, whose estimation turns out to be a standard forecasting task: predicting h-period ahead outcomes using both policy and control variables as predictors. Machine learning algorithms are particularly well-suited to capture complex nonlinear interactions between policy variables and high-dimensional controls, and have demonstrated superior forecasting performance in various macro and financial applications (e.g., Medeiros et al., 2021; Babii et al., 2022; Goulet Coulombe et al., 2022; Hauzenberger et al., 2025).

The second component is the stabilized weights (Robins et al., 2000), defined as the ratio of the product of marginal densities to the joint density of the policy variable and controls. Following Gutmann and Hyvärinen (2012), we recast the stabilized weight estimation as a probabilistic classification task. Specifically, we generate an artificial sample in which policy and control variables are drawn independently from their empirical marginal distributions and train a classifier to distinguish real data from this artificial data. The stabilized weights

emerge as the limiting odds ratio (Hastie et al., 2009, Chapter 14.2.4). This formulation bypasses direct density estimation, which is prone to misspecification bias and numerical instability in the parametric case, and plagued by the curse of dimensionality in the nonparametric case (Ai et al., 2021; Matsushita et al., 2023).

Recent work by Ballinari and Wehrli (2024) develops a related double/debiased machine learning approach for estimating nonlinear impulse responses. However, their framework applies only to discrete treatments, which introduces two limitations. First, discretizing a continuous treatment, e.g., splitting shocks into "large" and "small," imposes arbitrary binning decisions that render the responses threshold-dependent. Second, discretizing treatments restricts the class of policy counterfactuals one can consider. In contrast, our continuous treatment framework naturally accommodates the full distribution of policy interventions, enabling a wide range of policy counterfactuals through the average structural function, such as stochastic policy shifts.

An important practical advantage of our two-step approach is computational feasibility. In particular, the first-step estimation of nuisance parameters involves standard prediction tasks that scale well with modern machine learning methods. In the second-step, nonparametric regressions, e.g., local linear regression, operate only on low-dimensional policy variables, avoiding the curse of dimensionality. In our empirical application with 87 control variables, the proposed method completes estimation in minutes, while standard nonparametric local projections (e.g., Gonçalves et al., 2024a) would be computationally prohibitive.

We evaluate the finite-sample performance of the proposed method through simulations and benchmark it against nonparametric local projections as in Gonçalves et al. (2024a). Given the computational burden of nonparametric estimation, we consider two data generating processes (DGPs) with five and ten variables, respectively. The first DGP is a linear VAR with lower-triangular contemporaneous impact matrix, under which the true structural shocks can be recovered by properly selecting the control sets. The second DGP extends this scenario to a nonlinear setting via transformed variables in the spirit of Gonçalves et al.

(2021). Even in the linear case—where our method is expected to yield higher variance due to the introduction of the stabilized weights (Wu et al., 2024)—the proposed method still outperforms nonparametric local projections in terms of mean integrated squared error (MISE). Moreover, the estimated impulse response functions align closely with linear local projections when the true DGP is linear, while successfully capturing nonlinearities when present.

We apply the proposed method to estimate the dynamic causal effects of financial shocks on the US economy using a 7-variable local projections model with 12 lags. Our approach successfully recovers theoretically consistent average impulse response functions that differ quantitatively from their linear local projections counterparts. Moreover, it reveals rich nonlinearities of the effects of financial shocks, including the well-documented sign asymmetry—where adverse financial shocks have pronounced economic effects, but favorable shocks are more muted (e.g., Forni et al., 2024; Alessandri et al., 2025)—as well as size-dependency, where large adverse shocks produce disproportionately larger real effects than more modest ones, a pattern that linear models and plug-in machine learning estimators fail to capture.

The remainder of the paper is organized as follows. Section 2 illustrates the proposed framework through a simple example. Section 3 formalizes the model and provides details on the identification and estimation. Section 4 discusses the asymptotic properties and Section 5 evaluates the finite sample performance through simulation evidence. Section 6 presents the empirical application and Section 7 concludes.

2 An illustrative example

In this section, we illustrate the intuition of the proposed method through a simple example. Let X_t be a policy or treatment variable that is *randomly assigned* at time t. The dynamic causal effect of X_t on the h-period ahead outcome Y_{t+h}, of shifting the policy variable from

x_0 to x_1, known as the impulse response function (IRF) is defined as

$$\theta_h(x_1, x_0) = \mathbb{E}\left[Y_{t+h}|X_t = x_1\right] - \mathbb{E}\left[Y_{t+h}|X_t = x_0\right].$$

In practice, however, the policy variable is far from randomly assigned since it responds to macroeconomic conditions. This holds true even for carefully constructed macroeconomic shock measures.[2] Instead, we assume that the treatment X_t is *as good as randomly assigned* (Angrist and Kuersteiner, 2011; Angrist et al., 2018) conditional on a set of control variables, under which the conditional impulse response functions can be recovered as:

$$\theta_h(x_1, x_0, \boldsymbol{W}_t) = \mathbb{E}\left[Y_{t+h}|X_t = x_1, \boldsymbol{W}_t\right] - \mathbb{E}\left[Y_{t+h}|X_t = x_0, \boldsymbol{W}_t\right] \tag{1}$$

where $\boldsymbol{W}_t$ is the set of control variables typically corresponding to the information set available at time t. The unconditional IRF can then be recovered by averaging the conditional IRF over the distribution of $\boldsymbol{W}_t$, i.e.,

$$\theta_h(x_1, x_0) = \mathbb{E}_W\left[\theta_h(x_1, x_0, \boldsymbol{W}_t)\right] \ , \tag{2}$$

where $\mathbb{E}_W[\cdot]$ denotes expectation under the marginal distribution of $\boldsymbol{W}_t$.

The unconditional responses in Equation 2 allow us to construct a flexible class of policy counterfactuals, at the cost of a slower (nonparametric) convergence rate.[3] A natural plug-in estimator, also known as the *regression adjustment* estimator, is given by (Gonçalves et al., 2024a):

$$\hat{\theta}_h^{\mathrm{RA}}(x_1, x_0) = \frac{1}{T}\sum_{t=1}^{T}\left[\hat{\mu}_h(x_1, \boldsymbol{W}_t) - \hat{\mu}_h(x_0, \boldsymbol{W}_t)\right] \tag{3}$$

[2] See Miranda-Agrippino and Ricco (2021) for an example of the monetary policy shock measures, which can be predicted by various macro variables, and exhibit autocorrelation, and Ramey (2016) for an overview.

[3] The definition of impulse responses in the presence of nonlinearity is not unique. For example, (2) coincides with Definition 2 of Gonçalves et al. (2024a). Alternatively, their Definition 1 uses the realized series of shocks as the benchmark and examines how outcomes would change if we perturbed each realized shock by δ. This allows for averaging over the entire sample of shocks, leading to a parametric $\sqrt{T}$ convergence rate.

where $\hat{\mu}_h(x, \boldsymbol{w})$ is some consistent estimator for the conditional expectation $\mathbb{E}_W[Y_{t+h}|X_t = x, \boldsymbol{W}_t = \boldsymbol{w}]$. As a result, the estimation of causal IRFs reduces to a forecasting task, i.e., predicting the h-period ahead outcome Y_t using the policy X_t and controls $\boldsymbol{W}_t$ as predictors.

While conceptually attractive, such a plug-in estimator faces a sharp bias-variance trade-off, especially when the control variables $\boldsymbol{W}_t$ are high-dimensional. First, parametric SVAR and local projections counterparts with linearity or particular functional form restrictions are prone to misspecification bias (e.g., Barnichon and Matthes, 2018; Ramey and Zubairy, 2018; Gonçalves et al., 2024b). Second, flexible nonparametric methods easily suffer from the curse of dimensionality when the dimension of $\boldsymbol{W}_t$ is large. Meanwhile, machine learning approaches can lead to a first-order overfitting bias in the plug-in step (Chernozhukov et al., 2018).

To overcome these limitations, we propose an alternative approach that maintains nonparametric flexibility while mitigating the curse of dimensionality. The key insight is to construct a "pseudo-outcome" that remains valid for inference even when nuisance components are estimated with error (Kennedy et al., 2017; Semenova and Chernozhukov, 2021).

To fix ideas, consider a researcher who estimates linear local projections:

$$Y_{t+h} = X_t\beta_h + \boldsymbol{W}_t'\boldsymbol{\gamma}_h + u_{t+h}; \quad h = 0, 1, \ldots, H; \tag{4}$$

whereas the true data generating process contains nonlinear interactions:

$$Y_{t+h} = X_t\beta_h + \boldsymbol{W}_t'\boldsymbol{\gamma}_h + m_h(X_t, \boldsymbol{W}_t) + \varepsilon_{t+h}; \quad h = 0, 1, \ldots, H. \tag{5}$$

The linear local projections in Equation 4 provide the best linear approximation of the true conditional expectation function $\mu_h^0(X_t, \boldsymbol{W}_t) = X_t\beta_h + \boldsymbol{W}_t'\boldsymbol{\gamma}_h + m_h(X_t, \boldsymbol{W}_t)$, with coefficients $(\tilde{\beta}_h, \tilde{\boldsymbol{\gamma}}_h) = \underset{\beta,\boldsymbol{\gamma}}{\operatorname{argmin}} \mathbb{E}[(Y_{t+h} - X_t\beta_h - \boldsymbol{W}_t'\boldsymbol{\gamma}_h)^2]$. The plug-in estimand $\tilde{\theta}_h^{\mathrm{RA}} = \tilde{\beta}_h(x_1 - x_0)$ will usually be biased as it ignores the nonlinear interactions contained in $m_h(X_t, \boldsymbol{W}_t)$.

The pseudo-outcome approach automatically adjusts for such misspecification bias. In

particular, the h-period ahead pseudo-outcome is constructed as

$$Y^*_{t+h} = \underbrace{\mathbb{E}_W\left[X_t\tilde{\beta}_h + \boldsymbol{W}_t'\tilde{\gamma}_h\right]}_{\substack{\text{marginal}\\ \text{policy contribution}}} + \underbrace{\left(Y_{t+h} - X_t\tilde{\beta}_h - \boldsymbol{W}_t'\tilde{\gamma}_h\right)\pi(X_t, \boldsymbol{W}_t)}_{\text{robustness adjustment}} \tag{6}$$

where $\pi(X_t, \boldsymbol{W}_t) = \frac{f_X(X_t)}{f_{X|W}(X_t|\boldsymbol{W}_t)}$ are the stabilized weights (Robins et al., 2000).

The pseudo-outcome consists of two components. The first component, $\mathbb{E}_W\left[X_t\tilde{\beta}_h + \boldsymbol{W}_t'\tilde{\boldsymbol{\gamma}}_h\right]$, measures the marginal contribution of the policy variable X_t to the outcome Y_{t+h}, which can be used to estimate impulse responses as in Equation 2. However, due to misspecification, this term alone leads to biased estimates. This is reflected in the residual $Y_{t+h} - X_t\tilde{\beta}_h - \boldsymbol{W}_t'\tilde{\boldsymbol{\gamma}}_h$, which can be further decomposed into

$$\underbrace{Y_{t+h} - \mu^0_h(X_t, \boldsymbol{W}_t)}_{\text{intrinsic forecast error}} + \underbrace{\mu^0_h(X_t, \boldsymbol{W}_t) - X_t\tilde{\beta}_h - \boldsymbol{W}_t'\tilde{\gamma}_h}_{\text{misspecification error}}.$$

The first term of this expression is orthogonal to the policy variable X_t and thus does not affect the impulse response estimation asymptotically. The second term captures misspecification error of the conditional expectation function.

Stabilized weights play a role analogous to inverse propensity score weighting in the treatment effects literature. Here we use stabilized weights for continuous treatments. Observations featuring a level of the policy variable that is more likely to be "randomly assigned" receive a larger weight. For instance, conditional on recessionary macroeconomic conditions as measured by $\boldsymbol{W}_t$, the central bank typically reduces interest rate X_t to, say, x to stimulate the economy. When the marginal density $f_X(x)$—the unconditional probability of setting the policy rate to x—exceeds the conditional density $f_{X|W}(x \mid \boldsymbol{W}_t)$—implied by prevailing economic conditions—the stabilized weight $\pi(X_t, \boldsymbol{W}_t)$ increases accordingly. Intuitively, the stabilized weights down-weigh observations where the policy variable is highly predictable given the controls, and up-weigh those closer to quasi-experimental variation—i.e., where the realized policy deviates from its systematic component.

Given these adjustments, the conditional expectation of the pseudo-outcome identifies the average structural function and thus the impulse responses:

$$\mathbb{E}\left[Y^*_{t+h}|X_t = x_1\right] - \mathbb{E}\left[Y^*_{t+h}|X_t = x_0\right] = \theta_h(x_1, x_0).$$

Moreover, the pseudo-outcome enjoys a double robustness property: identification holds as long as one of $\mu_h(X_t, \boldsymbol{W}_t)$ or $\pi(X_t, \boldsymbol{W}_t)$ is correctly specified. If the conditional expectation function is correctly specified, the residual term is orthogonal to X_t and misspecification of $\pi(X_t, \boldsymbol{W}_t)$ does not bias the estimator. Conversely, if the stabilized weights are correctly specified, misspecification in $\mu_h(X_t, \boldsymbol{W}_t)$ is properly corrected through the weighting.

3 High-dimensional non-parametric local projections

In this section, we introduce the modeling framework for constructing dynamic policy counterfactuals. We then discuss the identification and estimation of the average structural function using double/debiased machine learning.

Assume that we observe a vector of macroeconomic variables $\boldsymbol{Z}_t = (Y_t, \boldsymbol{X}_t', \boldsymbol{W}_t')'$ where $Y_t \in \mathbb{R}$ is the outcome of interest, and $\boldsymbol{X}_t \in \mathcal{X} \subset \mathbb{R}^d$ denotes the vector of policy or shock variables with $d \geq 1$. Note that we expand the definition from a single policy variable to potentially a vector of them. This generality will allow us to think of conditional impulse responses more generally. Finally, $\boldsymbol{W}_t \in \mathcal{W} \subset \mathbb{R}^p$ is a vector of additional controls that may include lags of Y_t and $\boldsymbol{X}_t$. We consider a non-separable structural model (e.g., Imbens and Newey, 2009):

$$Y_{t+h} = g_h\left(\boldsymbol{X}_t, \boldsymbol{U}_{t,h}\right), \quad h = 0, 1, \ldots, H, \tag{7}$$

where $\boldsymbol{U}_{t,h} = (\boldsymbol{W}_t', V_{t+h})'$ denotes general disturbances that may be correlated with the policy variables $\boldsymbol{X}_t$. The dynamic counterfactual outcome is defined as $Y_{t+h}(\boldsymbol{x}) = g_h(\boldsymbol{x}, \boldsymbol{U}_{t,h})$ for a *fixed* policy vector $\boldsymbol{X}_t = \boldsymbol{x}$. Thus, $Y_{t+h}(\boldsymbol{x})$ is a random variable of the potential outcomes

given $\boldsymbol{x}$. The corresponding *average structural function* (ASF) is

$$\phi_h^0(\boldsymbol{x}) \equiv \mathbb{E}\left[Y_{t+h}(\boldsymbol{x})\right] = \mathbb{E}\left[g_h(\boldsymbol{x}, \boldsymbol{U}_{t,h})\right] \tag{8}$$

where the expectation is taken over the marginal distribution of $\boldsymbol{U}_{t,h}$. Hereafter, we use the superscript 0 to denote the true parameter.

The general nonseparable structural model in Equation 7 encompasses SVAR and local projections and their parametric nonlinear variants as special cases. For example, typical local projections (Jordà, 2005) model the h-period ahead outcome as a linear projection onto $\boldsymbol{X}_t$ and $\boldsymbol{W}_t$:

$$Y_{t+h} = g_h(\boldsymbol{X}_t, \boldsymbol{U}_{t,h}) = \boldsymbol{X}_t\boldsymbol{\beta}_h + \boldsymbol{W}_t'\boldsymbol{\gamma}_h + V_{t+h}, \quad h = 0, 1, \ldots, H, \tag{9}$$

where identification relies on $\boldsymbol{X}_t$ being conditionally uncorrelated with V_{t+h} given $\boldsymbol{W}_t$. In this case, local projections provide the best linear approximation to Equation 7, and the impulse response coefficients $\boldsymbol{\beta}_h$ can be interpreted as weighted average derivative effects (Kolesár and Plagborg-Møller, 2025):

$$\boldsymbol{\beta}_h = \int \omega(\boldsymbol{x}) \frac{\partial \phi_h(\boldsymbol{x})}{\partial \boldsymbol{x}} d\boldsymbol{x},$$

for some non-negative weight function $\omega(\boldsymbol{x})$.

Other special cases of Equation 7 include local projections with transformed regressors (Gonçalves et al., 2021), high-dimensional local projections with ℓ_1 regularization (Adamek et al., 2024), state-dependent local projections (Gonçalves et al., 2024b), and semi-parametric local projections (e.g., Paranhos, 2025). Following Gonçalves et al. (2024a), we impose no functional form restrictions on $\phi_h^0(\boldsymbol{x})$ beyond regularity conditions required for identification.

The ASF allows for flexible computation of impulse responses and policy counterfactuals, as illustrated by the examples below.

Example 1 (Average nonlinear impulse responses)**.** For any given levels of policy variable $\boldsymbol{x}_0$ and $\boldsymbol{x}_1$, the average *nonlinear* impulse response at horizon h is defined as the difference in the average structural functions, i.e.,

$$\theta_h^{\text{nonlin},0}(\boldsymbol{x}_1, \boldsymbol{x}_0) = \phi_h^0(\boldsymbol{x}_1) - \phi_h^0(\boldsymbol{x}_0). \tag{10}$$

Example 2 (Average linear impulse responses)**.** For a small perturbation $\Delta\boldsymbol{x}$ of the policy variable, the average *linear* impulse response at horizon h is defined as the average differences in the ASF when the policy variable is shifted by $\Delta\boldsymbol{x}$, i.e.,

$$\theta_h^{\text{lin},0}(\Delta\boldsymbol{x}) = \int_{\mathcal{X}} [\phi_h^0(\boldsymbol{x} + \Delta\boldsymbol{x}) - \phi_h^0(\boldsymbol{x})] \mathrm{d}\boldsymbol{x}. \tag{11}$$

Example 3 (Asymmetric impulse responses)**.** For any partition $\mathcal{X}^+$, $\mathcal{X}^- \subset \mathcal{X}$, we define region-specific impulse responses as:

$$\theta_h^{+,0} = \int_{\mathcal{X}^+} \left[\phi_h^0(\boldsymbol{x} + \Delta\boldsymbol{x}) - \phi_h^0(\boldsymbol{x})\right] \mathrm{d}\boldsymbol{x}, \quad \theta_h^{-,0} = \int_{\mathcal{X}^-} \left[\phi_h^0(\boldsymbol{x} + \Delta\boldsymbol{x}) - \phi_h^0(\boldsymbol{x})\right] \mathrm{d}\boldsymbol{x}. \tag{12}$$

Example 4 (Stochastic policy counterfactuals)**.** Consider evaluating the effects of a policy regime shift. Let $f_1(\boldsymbol{x})$ and $f_2(\boldsymbol{x})$ represent the marginal distributions of the policy rates under regimes 1 and 2, respectively. The expected effect of switching regimes is:

$$\theta_h^{f_1 \to f_2,0} = \int_{\mathcal{X}} \phi_h^0(\boldsymbol{x})[f_1(\boldsymbol{x}) - f_2(\boldsymbol{x})] \mathrm{d}\boldsymbol{x} \ . \tag{13}$$

As these examples show, identification of popular counterfactual policy effects follows directly from the identification of the ASF. We thus focus on estimating the entire function $\phi_h^0(\boldsymbol{x})$ in the following subsections.

3.1 Identification

To identify the average structural function ϕ_h^0, we work under the following assumptions.

Assumption 1 (Identification)**.**

(a) For all $\boldsymbol{x} \in \mathcal{X}$ and h, $Y_{t+h}(\boldsymbol{x}) \perp \boldsymbol{X}_t \mid \boldsymbol{W}_t$.

(b) For all $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{w} \in \mathcal{W}$, $f_{X|W}(\boldsymbol{x}|\boldsymbol{w})$ is bounded away from zero.

(c) For all t, whenever $X_t = x$, the observed outcome satisfies $Y_t = Y_t(x)$.

(d) The parameters of $f_{X|W}(\boldsymbol{x}|\boldsymbol{w})$ are constant for all t.

Assumption 1 is standard in the causal inference literature (Angrist and Kuersteiner, 2011; Angrist et al., 2018). Specifically, 1(a) is a weak unconfoundedness assumption that justifies the use of the conditional expectation function $\mu_h(\boldsymbol{x}, \boldsymbol{w})$ to identify the structural function. In the context of monetary policy, this requires that once we control for $\boldsymbol{W}_t$—which summarizes the information set at time t–policy rates are as good as randomly assigned (i.e., "monetary policy shocks"). Therefore, this assumption is plausible only if $\boldsymbol{W}_t$ is sufficiently rich.[4] Assumption 1(b) is a positivity assumption that requires the policy variables $\boldsymbol{X}_t$ can be set at any level in the support, conditional on the information set $\boldsymbol{W}_t$. This is commonly assumed in propensity score-based methods (Imbens, 2000). Assumption 1(c) is a standard assumption in the potential outcome framework, which indicates that the observed outcome is the potential outcome evaluated under the realized policy level. Finally, Assumption 1(d) ensures the policy reaction function parameters are constant.

Under Assumption 1, the average structural function is identified via

$$\phi_h^0(\boldsymbol{x}) = \mathbb{E}_W\left[\mu_h^0(\boldsymbol{x}, \boldsymbol{W}_t)\right], \quad \mu_h^0(\boldsymbol{X}_t, \boldsymbol{W}_t) \equiv \mathbb{E}\left[Y_{t+h} \mid \boldsymbol{X}_t, \boldsymbol{W}_t\right], \tag{14}$$

[4] A special case of the current setup is a triangular system (Imbens and Newey, 2009):

$$Y_{t+h} = g_h(X_t, \boldsymbol{W}_t, \boldsymbol{U}_{t,h}), \quad X_t = \psi(\boldsymbol{W}_t, V_t)$$

where $\psi(\cdot,\cdot)$ is the policy function and V_t is the scalar "policy shock" that is independent of $\boldsymbol{W}_t$ and $\boldsymbol{U}_{t,h}$. The potential outcome is then defined as $Y_{t+h}(x) = g_h(x, \boldsymbol{W}_t, \boldsymbol{U}_{t,h})$, which is still a random variable due to variation in $\boldsymbol{W}_t$ and $\boldsymbol{U}_{t,h}$.

which motivates the regression adjustment estimators (e.g., Gonçalves et al., 2024a). In parallel fashion to the example in Equation 6, we alternatively identify $\phi_h^0(\boldsymbol{x})$ using Neyman-orthogonal moments, following Chernozhukov et al. (2018) and Semenova et al. (2023). In particular, we construct the following pseudo-outcome (Kennedy et al., 2017):

$$Y_{t+h}^*(\eta_h) = \mathbb{E}_W[\mu_h(\boldsymbol{X}_t, \boldsymbol{W}_t)] + (Y_{t+h} - \mu_h(\boldsymbol{X}_t, \boldsymbol{W}_t))\,\pi(\boldsymbol{X}_t, \boldsymbol{W}_t), \tag{15}$$

where $\eta_h = (\mu_h, \pi)$ is the collection of nuisance functions, $\mu_h(\boldsymbol{X}_t, \boldsymbol{W}_t)$ is an estimand for the conditional expectation $\mathbb{E}[Y_{t+h}|\boldsymbol{X}_t, \boldsymbol{W}_t]$, and $\pi(\boldsymbol{X}_t, \boldsymbol{W}_t)$ are the stabilized weights (Robins et al., 2000).

The pseudo-outcome Y_{t+h}^* enjoys the following properties:

Proposition 1. Under Assumption 1, we have

(a) For all $\boldsymbol{x} \in \mathcal{X}$ and h, $\phi_h^0(\boldsymbol{x}) = \mathbb{E}[Y_{t+h}(\eta_h^0)|\boldsymbol{X}_t = \boldsymbol{x}]$.

(b) If either $\mu_h = \mu_h^0$ or $\pi = \pi^0$ or both,

$$\mathbb{E}\left[Y_{t+h}^*(\eta_h) - Y_{t+h}^*(\eta_h^0)|\boldsymbol{X}_t = \boldsymbol{x}\right] = 0\ . \tag{16}$$

(c) For all $\boldsymbol{x} \in \mathcal{X}$ and η_h, the Gâteaux derivative is zero

$$\partial_r \mathbb{E}\left[Y_{t+h}(\eta_h^0 + r(\eta_h - \eta_h^0))|\boldsymbol{X}_t = \boldsymbol{x}\right] = 0\ . \tag{17}$$

First, Proposition 1(a) shows that estimating the ASF reduces to a nonparametric estimation problem with low-dimensional policy vector $\boldsymbol{X}_t$. Second, Proposition 1(b) shows that the resulting estimator is doubly robust in the sense that if either μ_h or π is correctly specified, the conditional expectation identifies the counterfactuals of interest. Third, Proposition 1(c) shows that the moment condition $\phi_h(\boldsymbol{x}) = \mathbb{E}\left[Y_{t+h}^*(\eta_h)|\boldsymbol{X}_t = \boldsymbol{x}\right]$ is also *Neyman-orthogonal*,

i.e., the plug-in estimator $\mathbb{E}\left[Y^*_{t+h}(\hat{\eta}_h)|\boldsymbol{X}_t\right]$ is insensitive to the first-order bias in the nuisance parameters $\hat{\eta}_h$.

3.2 Estimation

Motivated by Proposition 1, we estimate the average structural function via a two-step procedure. First, we construct the pseudo-outcome $Y^*_{t+h}(\hat{\eta}_h)$, where the nuisance parameters $\hat{\eta}_h$ are estimated using flexible machine learning methods combined with cross-fitting. Second, we estimate the conditional expectation of the pseudo-outcome given the policy variables $\boldsymbol{X}_t$ using standard nonparametric regression techniques.

Specifically, we partition the data $\{(\boldsymbol{Y}_{t,0:H},\, \boldsymbol{X}_t,\, \boldsymbol{W}_t)\}_{t=1}^T$ into K adjacent blocks $\{\mathcal{N}_k\}_{k=1}^K$, each containing $T_K = \lfloor T/K \rfloor$ observations.[5] Let the quasi-complement $\mathcal{N}_k^{\mathrm{c}}$ be the union of all blocks excluding block k and its immediate neighbors, that is,

$$\mathcal{N}_k^{\mathrm{c}} = \{\mathcal{N}_1, \ldots, \mathcal{N}_K\} \setminus \{\mathcal{N}_{k-1}, \mathcal{N}_k, \mathcal{N}_{k+1}\}. \tag{18}$$

This construction preserves the dependence structure while mitigating data leakage due to serial dependence. Under mixing conditions (formalized in the next section), block $\mathcal{N}_k$ is approximately independent of its quasi-complement $\mathcal{N}_k^{\mathrm{c}}$.

The estimation proceeds as follows:

Algorithm: High-Dimensional Nonparametric Local Projections

1. **Construct pseudo-outcome**.

   (a) For each $k \in \{1, \ldots, K\}$, estimate nuisance parameters $\hat{\eta}_{h,k}$ using data in $\mathcal{N}_k^{\mathrm{c}}$.

   (b) For each $k \in \{1, \ldots, K\}$, compute the pseudo-outcome $Y^*_{t+h}(\hat{\eta}_{h,k})$ for each $t \in \mathcal{N}_k$:

$$Y^*_{t+h}(\hat{\eta}_{h,k}) = \frac{1}{|\mathcal{N}_{k,t}^{\mathrm{c}}|} \sum_{s \in \mathcal{N}_{k,t}^{\mathrm{c}}} \hat{\mu}_{h,k}(\boldsymbol{X}_t,\, \boldsymbol{W}_s) + (Y_{t+h} - \hat{\mu}_{h,k}(\boldsymbol{X}_t, \boldsymbol{W}_t))\, \hat{\pi}_k(\boldsymbol{X}_t, \boldsymbol{W}_t) \tag{19}$$

[5] $\boldsymbol{Y}_{t,0:H} = (Y_t, \ldots, Y_{t+H})'$ is a collection of outcomes in forecast horizons, and the kth block includes observations $Y_{(k-1)T_K+1}, \ldots, Y_{kT_K+H}$. For notational simplicity, we assume equal-length blocks.

where with slight abuse of notation, $\mathcal{N}_{k,t}^{c}$ is the quasi-complement of observation t in block k.[6]

2. **Estimate average structural function**. For any $\boldsymbol{x} \in \mathcal{X}$ in the support of $\boldsymbol{X}_t$, define

$$\left(\hat{\phi}_h(\boldsymbol{x}), \hat{\psi}_h(\boldsymbol{x})\right) = \underset{\phi,\psi}{\operatorname{argmin}} \sum_{k=1}^{K} \sum_{t \in \mathcal{N}_k} \left[Y_{t+h}^{*}(\hat{\eta}_{h,k}) - \phi - (\boldsymbol{X}_t - \boldsymbol{x})'\psi\right]^2 \mathcal{K}_b(\boldsymbol{X}_t - \boldsymbol{x}) \quad (20)$$

where $\mathcal{K}_b(\boldsymbol{u}) = \frac{1}{b^d} \prod_{j=1}^{d} \kappa\left(\frac{u_j}{b}\right)$, and $\kappa(\cdot)$ is a kernel function satisfying Assumption 4(a).

Estimation of Nuisance Parameters. Step 1(a) estimates the nuisance parameters $\hat{\eta}_{h,k} = (\hat{\mu}_{h,k}, \hat{\pi}_k)$ using data from $\mathcal{N}_k^c$. Estimating the conditional mean function $\mu_h(\boldsymbol{x}, \boldsymbol{w})$ reduces to a standard forecasting problem, for which a wide range of machine learning methods may be employed, including tree-based algorithms, boosting, and neural networks. These methods are particularly suitable given the high dimensionality of controls $\boldsymbol{W}_t$ and their proven forecasting accuracy (Medeiros et al., 2021; Babii et al., 2022; Goulet Coulombe et al., 2022; Hauzenberger et al., 2025). The only requirement is that the convergence rate of the estimator satisfies the conditions in the next section.

For the stabilized weights, we transform the density ratio estimation into a classification problem following Gutmann and Hyvärinen (2012). Let $\{(\tilde{\boldsymbol{X}}_s, \tilde{\boldsymbol{W}}_s)\}_{s=1}^{\tilde{T}}$ be an artificial sample generated by first randomly drawing $\tilde{\boldsymbol{X}}_s$ from the empirical distribution of $\{\boldsymbol{X}_t\}_{t=1}^{T}$, and then independently drawing $\tilde{\boldsymbol{W}}_s$ from the empirical distribution of $\{\boldsymbol{W}_t\}_{t=1}^{T}$. The resulting artificial sample $\{(\tilde{\boldsymbol{X}}_s, \tilde{\boldsymbol{W}}_s)\}_{s=1}^{\tilde{T}}$ thus mimics the joint distribution under independence, i.e., with density $f_X(\boldsymbol{X}_t) f_W(\boldsymbol{W}_t)$.

Define a binary label L_t such that $L_t = 1$ for the actual sample and $L_t = 0$ for the artificial sample. We pool both samples into $\{(\boldsymbol{X}_t, \boldsymbol{W}_t, L_t)\}_{t=1}^{T} \cup \{(\tilde{\boldsymbol{X}}_s, \tilde{\boldsymbol{W}}_s, L_s)\}_{s=1}^{\tilde{T}}$. The

[6] That is, $\mathcal{N}_{k,t}^{c}$ includes all observations in block k except observation t and its neighbors.

conditional probability of the label L_t given $(\boldsymbol{X}_t, \boldsymbol{W}_t)$ can be expressed using Bayes' rule:

$$P(L_t = 0 \mid \boldsymbol{X}_t,\, \boldsymbol{W}_t) = \frac{f_X(\boldsymbol{X}_t) f_W(\boldsymbol{W}_t) P(L_t = 0)}{f_{XW}(\boldsymbol{X}_t,\, \boldsymbol{W}_t) P(L_t = 1) + f_X(\boldsymbol{X}_t) f_W(\boldsymbol{W}_t) P(L_t = 0)}\ , \tag{21}$$

since the actual sample is drawn from the joint distribution $f_{XW}(\boldsymbol{X}_t, \boldsymbol{W}_t)$. Direct calculation shows that the stabilized weights can be expressed as

$$\pi(\boldsymbol{X}_t,\, \boldsymbol{W}_t) = \frac{P(L_t = 0 \mid \boldsymbol{X}_t,\, \boldsymbol{W}_t)}{1 - P(L_t = 0 \mid \boldsymbol{X}_t,\, \boldsymbol{W}_t)} \cdot \frac{P(L_t = 1)}{P(L_t = 0)}\ . \tag{22}$$

Thus, the stabilized weights can be estimated as

$$\hat{\pi}(\boldsymbol{X}_t,\, \boldsymbol{W}_t) = \frac{\hat{P}(L_t = 0 \mid \boldsymbol{X}_t,\, \boldsymbol{W}_t)}{1 - \hat{P}(L_t = 0 \mid \boldsymbol{X}_t, \boldsymbol{W}_t)} \cdot \frac{T}{\tilde{T}}\ . \tag{23}$$

Under mild regularity conditions, the stabilized weights inherit the convergence rate of $\hat{P}(L_t = 0 \mid \boldsymbol{X}_t, \boldsymbol{W}_t)$.[7] This motivates us to estimate (21) with a variety of machine learning methods, such as tree-based classifiers, support vector machines, and neural networks. A related approach is employed in Hu et al. (2024) for a different application.

Our classification-based approach is motivated by recent literature on direct density ratio estimation (e.g., Ai et al., 2021; Matsushita et al., 2023). However, existing methods typically rely on kernel or sieve approximations that suffer from the curse of dimensionality with high-dimensional $\boldsymbol{W}_t$. In contrast, our classification formulation accommodates high-dimensional controls flexibly via modern machine learning methods.

[7] Formally, if $\hat{P}(L_t = 0 \mid \boldsymbol{X}_t, \boldsymbol{W}_t) = P(L_t = 0 \mid \boldsymbol{X}_t, \boldsymbol{W}_t) + O_p(r_T^{-1})$, then

$$\hat{\pi}(\boldsymbol{X}_t,\, \boldsymbol{W}_t) = \pi(\boldsymbol{X}_t,\, \boldsymbol{W}_t) + \frac{T}{\tilde{T}} \cdot (1 - P(L_t = 0 \mid \boldsymbol{X}_t, \boldsymbol{W}_t))^{-2} O_p(r_T^{-1})\ .$$

When $T/\tilde{T} \xrightarrow{p} c \in (0, \infty)$ and $P(L_t = 0 \mid \boldsymbol{X}_t,\, \boldsymbol{W}_t) \neq 1$, both estimators share the same rate. Although a large artificial sample size $\tilde{T}$ reduces $\frac{T}{\tilde{T}}$, it also increases $P(L_t = 0 | \boldsymbol{X}_t,\, \boldsymbol{W}_t)$ due to label imbalances. Therefore, it is recommended to set $\tilde{T} \approx T$ in practice.

Cross-fitting. The cross-fitting procedure aims to control the overfitting bias from the first-step estimation of nuisance functions. Without cross-fitting, first-step estimation errors would be correlated with second-step errors, leading to non-negligible asymptotic bias (Chernozhukov et al., 2018). This is particularly useful when using flexible machine learning methods for $\hat{\mu}_{h,k}$ and $\hat{\pi}_k$. Alternatively, one could impose strong restrictions on the nuisance functions, such as sparsity in the conditional expectation (Belloni et al., 2014; Su et al., 2019), but this may introduce model misspecification bias.

Construction of the Pseudo-Outcome. Step 1(b) implements a sample analogue of the pseudo-outcome (15) where we estimate the marginal expectation $\mathbb{E}_W[\mu_h(\boldsymbol{X}_t, \boldsymbol{W}_t)] = \int_{\mathcal{W}} \mu_h(\boldsymbol{X}_t, \boldsymbol{v}) f_W(\boldsymbol{v}) \mathrm{d}\boldsymbol{v}$ using leave-out estimation. As in Step 1(a), we exploit the β-mixing property of the data, which ensures that observations $\boldsymbol{X}_t$ and $\boldsymbol{W}_s$ separated sufficiently in time are asymptotically independent. This allows us to approximate the marginal expectation via averaging over the empirical distribution of $\boldsymbol{W}_s$ within $\mathcal{N}_{k,t}^{\mathrm{c}}$.

Nonparametric Regression. Finally, Step 2 is a standard local linear regression problem. The average structural function $\phi_h(\boldsymbol{x})$ and its derivative $\psi_h(\boldsymbol{x})$ have the closed-form solution:

$$\begin{bmatrix} \hat{\phi}_h(\boldsymbol{x}) \\ b\hat{\psi}_h(\boldsymbol{x}) \end{bmatrix} = \left[\frac{1}{T} \sum_{t=1}^{T} \mathcal{K}_{t,b,x} \tilde{\boldsymbol{X}}_{t,b,x} \tilde{\boldsymbol{X}}'_{t,b,x} \right]^{-1} \left[\frac{1}{T} \sum_{k=1}^{K} \sum_{t \in \mathcal{N}_k} \mathcal{K}_{t,b,x} \tilde{\boldsymbol{X}}_{t,b,x} Y_{t+h}(\hat{\eta}_{h,k}) \right] \tag{24}$$

where $\mathcal{K}_{t,b,x} = \mathcal{K}_b(\boldsymbol{X}_t - \boldsymbol{x})$ and $\tilde{\boldsymbol{X}}_{t,b,x} = (1, b^{-1}(\boldsymbol{X}_t - \boldsymbol{x})')'$. In principle, alternative nonparametric regression methods such as sieve estimators may also be employed (Chen, 2007).

4 Asymptotic properties

This section develops the asymptotic properties of the high-dimensional local projection estimator introduced in the previous section. Throughout, we assume that the vector of observables $\boldsymbol{Z}_t = (Y_t, \boldsymbol{X}_t', \boldsymbol{W}_t')'$ is defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and that

potential outcomes are generated by the nonseparable model (7) as $Y_{t+h}(\boldsymbol{x}) = g_h(\boldsymbol{x}, \boldsymbol{U}_{t,h})$. We denote the support of $\boldsymbol{Z}_t$ by $\mathcal{Z}$.

Assumption 2 (DGP)**.**

(a) The process $\boldsymbol{Z}_t$ is strictly stationary and β-mixing with mixing coefficients satisfying $\beta(n) \leq c_1 \exp(-c_2 n)$ for some constants $c_1 \geq 0$ and $c_2 > 0$.

(b) For some $q > 2$, the conditional moments $\mathbb{E}\left[(Y_{t+h} - \mu_h^0(\boldsymbol{X}_t, \boldsymbol{W}_t))^q \mid \boldsymbol{X}_t = \boldsymbol{x},\ \boldsymbol{W}_t = \boldsymbol{w}\right]$ are uniformly bounded over $\mathcal{X} \times \mathcal{W}$.

Assumption 2 imposes standard mixing and moment conditions that facilitate the application of independence coupling arguments à la Berbee (1987). This ensures asymptotic independence between estimation blocks $\{\boldsymbol{Z}_t\}_{t \in \mathcal{N}_k}$ and their quasi-complements $\{\boldsymbol{Z}_t\}_{t \in \mathcal{N}_k^{\text{qc}}}$. Similar strategies have been used in high-dimensional and dependent data contexts; see Gao et al. (2022) for time series applications and Semenova et al. (2023) for panel data setup.

Our next assumption imposes regularity conditions on the nuisance functions. Denote the limits of the nuisance function estimators as $\overline{\mu}_0, \ldots, \overline{\mu}_H, \overline{\pi}$.

Assumption 3 (Nuisance functions)**.**

(a) For each h, either $\overline{\mu}_h = \mu_h^0$ or $\overline{\pi} = \pi^0$ or both.

(b) The functions $\overline{\mu}_0, \ldots, \overline{\mu}_H$ and $\overline{\pi}$ are uniformly bounded over $\mathcal{X} \times \mathcal{W}$.

(c) For all $k \in \{1, \ldots, K\}$, $h \in \{0, \ldots, H\}$, and any $\boldsymbol{x} \in \mathcal{X}$,

$$\sup_{\boldsymbol{v}: |v_j - x_j| \leq b} \sqrt{\mathbb{E}_W[\delta_{\mu,k,h}(\boldsymbol{v}, \boldsymbol{W}_t)^2]} \cdot \sqrt{\mathbb{E}_W[\delta_{\pi,k}(\boldsymbol{v}, \boldsymbol{W}_t)^2]} = o_p\left((Tb^d)^{-1/2}\right), \tag{25}$$

where $\delta_{\mu,k,h}(\boldsymbol{x}, \boldsymbol{w}) = \hat{\mu}_{h,k}(\boldsymbol{x}, \boldsymbol{w}) - \mu_h^0(\boldsymbol{x}, \boldsymbol{w})$ and $\delta_{\pi,k}(\boldsymbol{x}, \boldsymbol{w}) = \hat{\pi}_k(\boldsymbol{x}, \boldsymbol{w}) - \pi^0(\boldsymbol{x}, \boldsymbol{w})$.

(d) For all k, h, and x, and for some sequence $r_T = o(b^{d/2})$,

$$\sup_{\boldsymbol{v}: |v_j - x_j| \leq b} \mathbb{E}_W\left[\overline{\delta}_{\mu,k,h}(\boldsymbol{v}, \boldsymbol{W}_t)^q \mid \boldsymbol{X}_t = v\right] = O_p(r_T^q), \tag{26}$$

$$\sup_{\boldsymbol{v}:|v_j - x_j| \leq b} \mathbb{E}_W \left[\overline{\delta}_{\pi,k}(\boldsymbol{v}, \boldsymbol{W}_t)^q \mid \boldsymbol{X}_t = v \right] = O_p(r_T^q), \tag{27}$$

where $\overline{\delta}_{\mu,k,h}(\boldsymbol{x}, \boldsymbol{w}) = \hat{\mu}_{h,k}(\boldsymbol{x}, \boldsymbol{w}) - \overline{\mu}_h(\boldsymbol{x}, \boldsymbol{w})$ and $\overline{\delta}_{\pi,k}(\boldsymbol{x}, \boldsymbol{w}) = \hat{\pi}_k(\boldsymbol{x}, \boldsymbol{w}) - \overline{\pi}(\boldsymbol{x}, \boldsymbol{w})$, and q is defined in Assumption 2(b).

Assumption 3(a) is a standard double robustness condition, common in the recent literature on semiparametric inference with machine learning (Chernozhukov et al., 2018; Semenova et al., 2023). It allows consistent estimation when either the conditional mean or stabilized weights are correctly specified. In practice, this can be achieved using flexible nonparametric or machine learning methods. Assumption 3(b) imposes uniform boundedness of the nuisance function estimators, which, when combined with consistency 3(d), ensures that the first-order bias has negligible effects in the second step estimation. Finally, Assumptions 3(c) and 3(d) specify convergence rates of the nuisance function estimators. The use of product convergence rates arises from the fact that one of the nuisance functions can be misspecified, in which case it requires a faster convergence rate of the consistent nuisance function estimator. When it comes to the convergence rates to population limits ($\overline{\mu}_h$ and $\overline{\pi}$), we impose a slightly stronger condition (for some $q > 2$) than the i.i.d. case (Kennedy et al., 2017; Semenova et al., 2023). This is used to bound the cross-covariance due to dependence in the data.

We now state conditions for the nonparametric estimation of the average structural function.

Assumption 4 (Nonparametric Estimation)**.**

(a) The kernel function $\kappa(\cdot)$ is bounded, Lipschitz continuous, supported on $[-1, 1]$, and satisfies $\kappa_1 = 0$, $\kappa_2, \kappa_3 < \infty$ with $\kappa_j = \int \boldsymbol{u}^j \kappa(\boldsymbol{u}) \mathrm{d}\boldsymbol{u}$.

(b) For all $\boldsymbol{u}, \boldsymbol{v} \in \mathcal{X}$, we have $0 < f_X(\boldsymbol{u}) < \infty$ and $|f_{X,l}(\boldsymbol{u}, \boldsymbol{v})| < \infty$ where $f_{X,l}(\cdot, \cdot)$ is the joint density of $\boldsymbol{X}_0$ and $\boldsymbol{X}_l$.

(c) The marginal density $f_X(\boldsymbol{x})$ is twice continuously differentiable with bounded derivatives over $\mathcal{X}$.

(d) The average structural function $\phi_h^0(\cdot)$ is twice continuously differentiable on $\mathcal{X}$.

(e) For some $\delta > 0$, $\mathbb{E}[|Y_{t+h}^*(\overline{\eta}_h)|^{2+\delta} \mid \boldsymbol{X}_t = \boldsymbol{x}] < \infty$ for all $\boldsymbol{x} \in \mathcal{X}$, and $\mathbb{E}[|Y_{t+h}^*(\overline{\eta}_h)|^4] < \infty$.

(f) The intrinsic error $\sigma_h^2(\boldsymbol{x}) \coloneqq \mathrm{Var}[Y_{t+h}^*(\overline{\eta}_h) \mid \boldsymbol{X}_t = \boldsymbol{x}]$ is continuous and bounded on $\mathcal{X}$.

(g) The bandwidth satisfies $b \to 0$ and $Tb^d \to \infty$ as $T \to \infty$.

Assumption 4 imposes standard regularity conditions on kernel-based nonparametric estimators. Compact support and Lipschitz continuity in part 4(a) are assumed for technical simplicity, and can be relaxed following Hansen (2008). The remaining conditions mirror standard requirements in the nonparametric regression literature, applied here to the infeasible pseudo-outcome $Y_{t+h}^*(\overline{\eta}_h)$. Note that under uniform boundedness of the nuisance function estimators, the pseudo-outcome inherits the same regularity properties as the oracle pseudo-outcome $Y_{t+h}^*(\eta_h^0)$.

We are now ready to state the main asymptotic result:

Theorem 1 (Asymptotic Normality)**.** Suppose Assumptions 1–4 hold. If $b \to 0$ and $Tb^d \to \infty$, then for any $\boldsymbol{x} \in \mathcal{X}$,

$$\sqrt{Tb^d}\left(\hat{\phi}_h(\boldsymbol{x}) - \phi_h^0(\boldsymbol{x}) - B_{h,b}(\boldsymbol{x})\right) \xrightarrow{d} \mathcal{N}(0, V(\boldsymbol{x})), \tag{28}$$

where $B_{h,b}(\boldsymbol{x}) = \frac{b^2}{2}\sum_{j=1}^d \frac{\partial^2 \phi_h^0(\boldsymbol{x})}{\partial x_j^2}\kappa_2 + o(b^2)$ and $V(\boldsymbol{x}) = \frac{\sigma_h^2(\boldsymbol{x})}{f_X(\boldsymbol{x})}\int \kappa(\boldsymbol{u})^2 \, \mathrm{d}\boldsymbol{u}$.

Theorem 1 establishes the pointwise asymptotic normality of the high-dimensional local projection estimator, with convergence rate $\sqrt{Tb^d}$, analogous to classical results for local linear regression under both i.i.d. and weak dependence (Fan and Yao, 2003). The bias term is of order b^2 and can be further corrected following Takatsu and Westling (2024). The

variance depends on the intrinsic error and the distribution of the policy variable, reflecting the use of stabilized weights, as is typical in propensity score weighting methods (Imai and Ratkovic, 2014; Wu et al., 2024).

5 Monte Carlo simulations

This section examines the finite-sample performance of the high-dimensional nonparametric local projections estimator.

5.1 Simulation design

The first simulation design is based on a linear VAR(1) model:

$$\boldsymbol{Y}_t = \boldsymbol{A}\boldsymbol{Y}_{t-1} + \boldsymbol{B}\boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(0, \boldsymbol{I}_n), \tag{29}$$

where $\boldsymbol{Y}_t$ is an $n \times 1$ vector of observables, $\boldsymbol{A}$ is the autoregressive coefficient matrix, and $\boldsymbol{B}$ is a lower-triangular contemporaneous impact matrix. We focus on the impulse responses of $\boldsymbol{Y}_t$ to shocks in the last component, $Y_{n,t}$. Because of the triangular structure of $\boldsymbol{B}$, the policy variable $Y_{n,t}$ contemporaneously responds to the remaining $n-1$ variables. Identification of the structural function thus requires conditioning on both the contemporaneous and the lagged values, i.e., $\boldsymbol{W}_t = (\boldsymbol{Y}'_{1:(n-1),t}, \boldsymbol{Y}'_{t-1})'$. This design mimics the high-dimensional time series setting studied by Adamek et al. (2024), but with dense coefficient matrices to better capture the persistence and cross-sectional dependence typical of macroeconomic data.[8]

[8]Notice that in this triangular VAR design, it is necessary to control for lags $\boldsymbol{W}_t = (\boldsymbol{Y}_{1,t-1}, \ldots, \boldsymbol{Y}_{n,t-1})$ even when we aim to estimate the responses to the first component, $\boldsymbol{Y}_{1,t}$. This is because only when we condition on $\boldsymbol{W}_t$ does the randomness in $Y_{1,t}$ come solely from $\epsilon_{1,t}$, and thus satisfying the unconfoundedness assumption 1(a).

The second design introduces nonlinearity via a VAR(1) with quadratic effects:

$$\begin{aligned} W_{j,t} &= \rho_j^w W_{j,t-1} + \sigma_w \epsilon_{j,t}^w, \quad j = 1, \ldots, n \\ X_t &= \rho_x X_{t-1} + \boldsymbol{\beta}_{xw}' \boldsymbol{W}_t + \sigma_x \epsilon_t^x \qquad , \quad \epsilon_{j,t}^w, \epsilon_t^x, \epsilon_t^y \sim N(0,1). \\ Y_t &= \rho_y Y_{t-1} + \boldsymbol{\beta}_{yw}' \boldsymbol{W}_t + \beta_{yx,1} X_t + \beta_{yx,2} X_t^2 + \sigma_y \epsilon_t^y \end{aligned} \tag{30}$$

Here, the controls $W_{j,t}$ follow heterogeneous AR(1) processes; the policy variable X_t depends linearly on its lag and $\boldsymbol{W}_t$; and the outcome Y_t depends nonlinearly on the policy variable through a quadratic term $\beta_{yx,2} X_t^2$, which generates state-dependent impulse responses, resembling the nonlinear factor structures discussed in Gonçalves et al. (2024a).

For both designs, we simulate data for $T \in \{300, 500\}$ and $n \in \{5, 10\}$. The true average structural functions and impulse responses are illustrated in the supplementary materials.

We apply the proposed HDNLP estimator following the algorithm in Section 3.2. For the first-step nuisance function estimation, we use linear local projections for the conditional expectation function $\mu_h(X_t, \boldsymbol{W}_t)$ and gradient boosting decision trees for the stabilized weights $\pi(X_t, \boldsymbol{W}_t)$ (implemented as a classifier).[9] For the second-step local linear estimator, we use the Gaussian kernel with the rule-of-thumb bandwidth $b^{\text{rot}} = 1.06 \hat{\sigma}_X T^{-\frac{1}{5}}$, where $\hat{\sigma}_X$ is the standard deviation of the univariate policy variable X_t.[10] Finally, we implement cross-fitting with $K = 5$ folds.

We compare the HDNLP estimator against two benchmarks: (i) the plug-in estimator of Gonçalves et al. (2024a), which estimates conditional expectations via local linear regression and applies the plug-in formula (3), and (ii) standard linear local projections (Jordà, 2005). We evaluate the performance with the mean integrated squared error (MISE) for structural functions, as well as absolute bias and mean squared error (MSE) for impulse responses.

[9] Results are qualitatively the same if we use more sophisticated machine learning methods for the conditional expectation function. Doing so would slow down simulations significantly, since the machine learning model needs to be estimated $d_Y \times K \times (H+1)$ times.

[10] Bandwidth selection via cross-validation yields qualitatively similar results.

Specifically, the MISE for each horizon h is computed as:

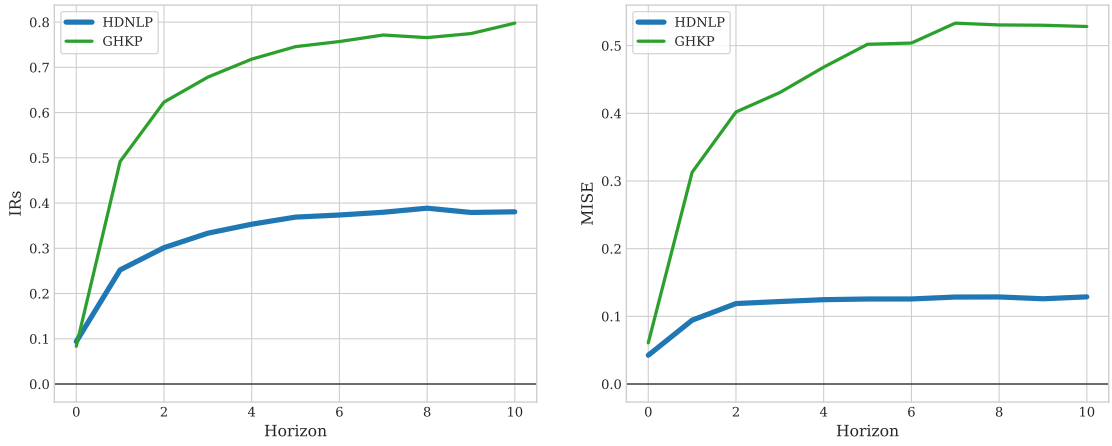
$$\text{MISE}(h) = \int_{u\in[0,1]} \left(\phi_h^0(u) - \hat{\phi}_h(u)\right)^2 du \ . \tag{31}$$

In case of multiple outcomes, we average the metrics across variables in the interest of space.

5.2 Simulation results

Figure 1 displays the MISE defined in Equation 31 for the linear VAR(1) DGP with $n = 5$ and sample sizes $T = 300, 500$. Two key patterns emerge. First, the HDNLP estimator consistently achieves lower MISE than the plug-in estimator across horizons and dimensions with very few exceptions. Second, somewhat surprisingly, the bias of the HDNLP estimator is elevated at short horizons, especially for smaller sample sizes. This reflects two sources of finite-sample biases.

Figure 1: MISE for linear DGP with 5 variables: HDNLP vs. Plug-in estimators



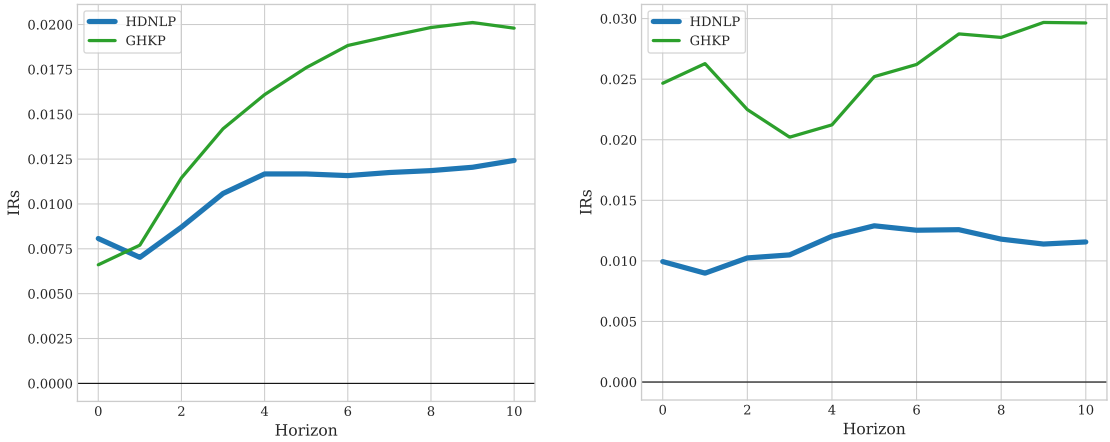
(a) Sample size: $T = 300$ (b) Sample size: $T = 500$

Notes: Mean integrated squared errors. DGP is the VAR(1) with 5 variables shown in Equation 29. Green line is based on the Plug-in estimator, the blue line is based on the HDNLP estimator. See text.

First, cross-fitting with small samples reduces the effective sample size in each fold, increasing first-step estimation error. Second, high persistence in the DGP ($\rho \approx 0.9$) means the quasi-complement and main samples may differ in distribution despite our mixing assump-

tions. These finite-sample issues diminish rapidly as the sample size, T, grows, as supported by the differences in Figure 1(a) and 1(b). For the nonlinear DGP, Figure 2 shows similar patterns: the HDNLP estimator again outperforms the plug-in estimator in terms of MISE, particularly as the sample size grows.

Figure 2: MISE for non-linear DGP with 5 variables: HDNLP vs. Plug-in estimators

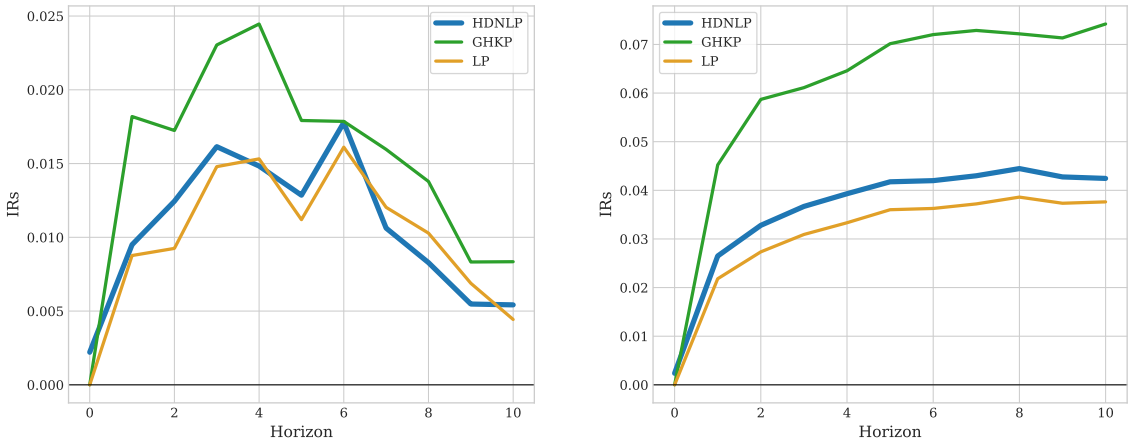


(a) Sample size: $T = 300$ (b) Sample size: $T = 500$

Notes: Mean integrated squared errors. DGP is the non-linear VAR(1) with 5 variables shown in Equation 30. Green line is based on the Plug-in estimator, the blue line is based on the HDNLP estimator. See text.

Next, we evaluate the impulse response functions across two dimensions, bias and mean squared error (MSE). Figure 3 reports the absolute bias and MSE for impulse responses generated by the linear DGP with $n = 5$ variables and sample size $T = 300$. In general, the HDNLP estimator delivers absolute bias comparable to linear local projections at most horizons, though slightly larger at very short horizons due to the small-sample bias issues previously mentioned. In terms of MSE, HDNLP outperforms the plug-in estimator throughout, with particularly large gains at longer horizons. For the nonlinear DGP, Figure 4 shows a similar pattern: HDNLP exhibits larger bias at short horizons for small T, although it achieves sizable MSE improvements at longer horizons where nonlinearities are more pronounced. Experiments reported in the supplementary material confirms that these biases rapidly decrease with larger samples.

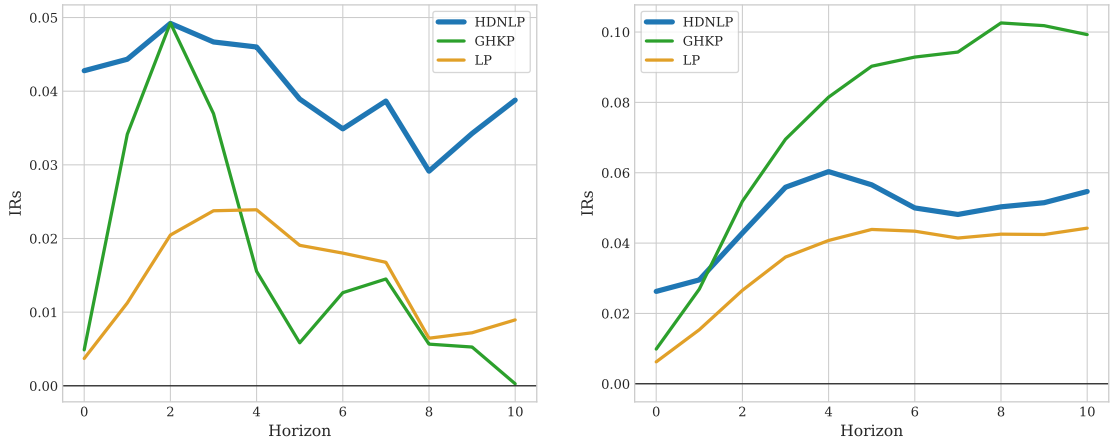
Figure 3: Average impulse responses diagnostics (linear DGP, $n = 5,\ T = 300$)



(a) Absolute Bias (b) Mean Squared Error

Notes: Data generated with linear VAR(1) with 5 variables shown in Equation 29. Green line is based on the Plug-in estimator, the blue line is based on the HDNLP estimator, and the orange line is based on linear local projection estimator. Panel A shows the absolute bias per horizon, whereas panel B shows the Mean Squared Error. See text.

Figure 4: Average impulse responses diagnostics (Nonlinear DGP, $n = 5,\ T = 300$)



(a) Absolute Bias (b) Mean Squared Error

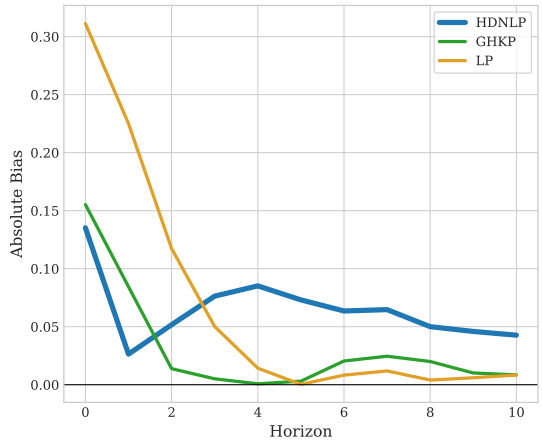
Notes: Data generated with non-linear VAR(1) with 5 variables shown in Equation 30. Green line is based on the Plug-in estimator, the blue line is based on the HDNLP estimator, and the orange line is based on linear local projection estimator. Panel A shows the absolute bias per horizon, whereas panel B shows the Mean Squared Error. See text.

To further explore state-dependent effects, we compute asymmetric impulse responses by comparing responses for negative and positive regions of X_t. Specifically, we consider two

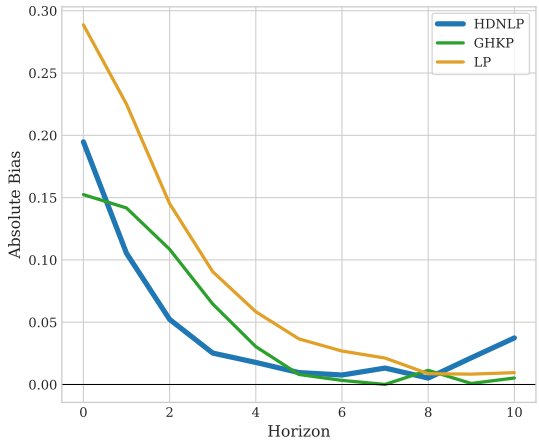
region-specific impulse responses defined as:

$$\hat{\beta}_h^- = \int_{[-0.5,-0.25]} \hat{\phi}_h(x+\delta) - \hat{\phi}_h(x)\mathrm{d}x, \quad \hat{\beta}_h^+ = \int_{[0,0.25]} \hat{\phi}_h(x+\delta) - \hat{\phi}_h(x)\mathrm{d}x, \quad \delta = 0.25 \ .$$

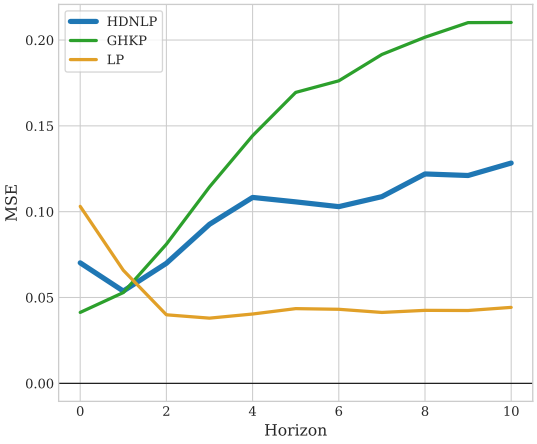
Figure 5: Region-specific Impulse responses for Nonlinear DGP ($n = 5,\ T = 300$)



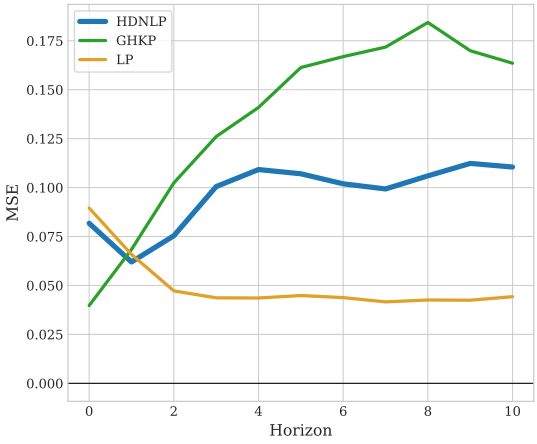
(a) Absolute Bias, $x \in [-0.5, -0.25]$



(b) Absolute Bias, $x \in [0, 0.25]$



(c) Mean Squared Error, $x \in [-0.5, -0.25]$



(d) Mean Squared Error, $x \in [0, 0.25]$

Notes: Data generated with non-linear VAR(1) with 5 variables shown in Equation 30. Green line is based on the Plug-in estimator, the blue line is based on the HDNLP estimator, and the orange line is based on linear local projection estimator. Panel A shows the absolute bias per horizon, whereas panel B shows the Mean Squared Error. See text.

Figure 5 shows that the HDNLP estimator achieves substantially smaller bias than linear local projections for the nonlinear DGP, while maintaining lower MSE than the plug-in estimator. Compared with the linear local projections, the HDNLP estimator is generally

associated with larger variance, since it uses only part of the data. However, it is still more precise than the linear local projections, especially at shorter horizons.

Overall, the simulation results demonstrate that the proposed high-dimensional nonparametric local projections estimator effectively recovers both average structural functions and impulse responses, outperforming existing methods in most cases in terms of bias and MSE under both linear and nonlinear settings though these improvements become more visible with larger sample sizes given the computational demands of our estimator. Consistent patterns hold for different sample sizes as well as dimension of the model, as reported in the supplementary materials. However, practitioners should be cautious with very persistent data and small samples. In such situations, linear local projections appear to be more robust.

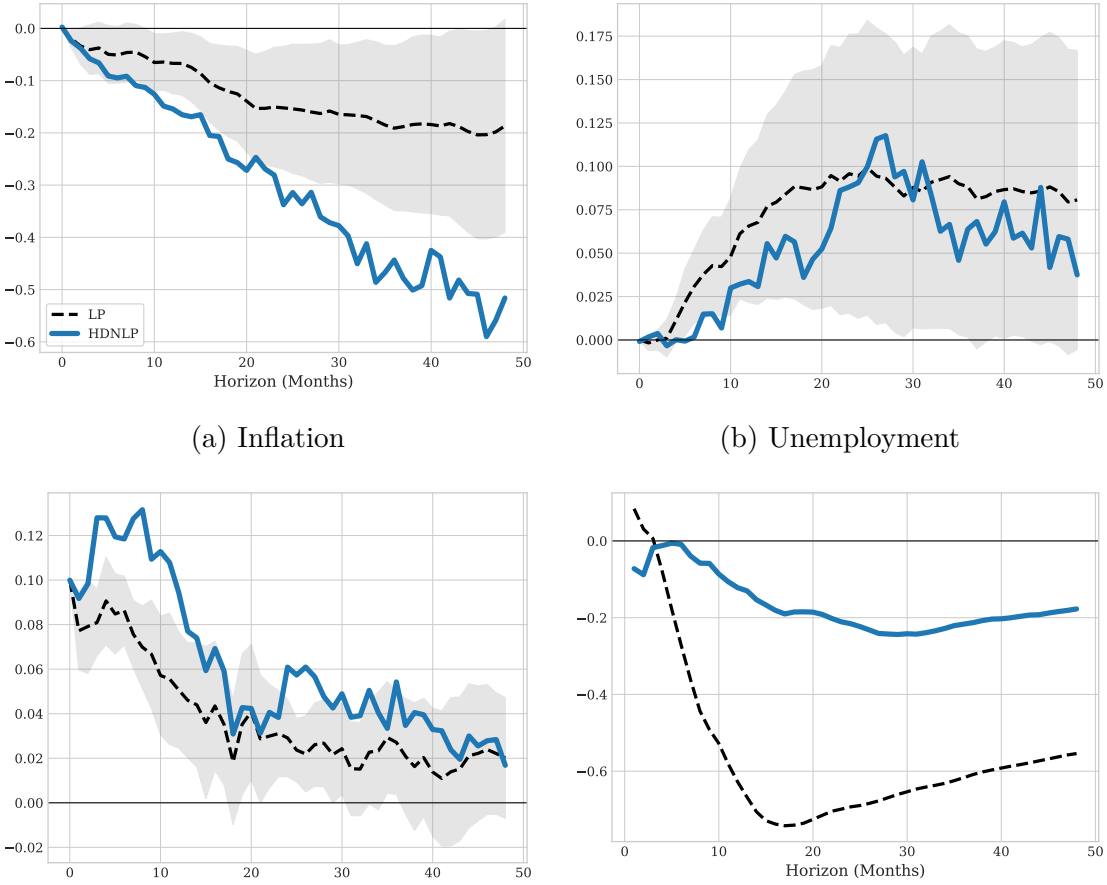
6 The economy's response to financial conditions redux

In a recent study, Forni et al. (2024) reinforce the findings of a growing literature on the nonlinear effects that financial conditions can have on the economy (see, e.g. Brunnermeier et al., 2013; Brunnermeier and Sannikov, 2014, for a survey). Relatedly, Alessandri et al. (2025) show that financial conditions can amplify or attenuate the effects of monetary policy whereas Barnichon et al. (2022) show that, by taking into account the asymmetric effects of credit conditions on the economy, the effects of financial market disruptions on the economy are large.

In this section we revisit these questions by reexamining the evidence in the Forni et al. (2024) study but using the methods introduced in previous sections. In particular, financial conditions are measured by the excess bond premium (EBP) from Gilchrist and Zakrajšek (2012). Hence we analyze the responses of industrial production (INDPRO), unemployment rate (UNRATE), CPI inflation (CPIAUCSL) , the Chicago Fed's National Financial Con-

ditions Index (NFCI), the S&P 500 Composite Stock Price Index (SP500), and the federal funds rate (FFR)[11] to EBP. All variables are transformed to ensure stationarity, with details summarized in the Supplementary materials.

Figure 6: Average Linear Responses for Financial Shocks



(a) Inflation

(b) Unemployment

(c) EBP

(d) Cumulative Response Ratio

Notes: Cumulative response ratio refers to the ratio of cumulative unemployment responses relative to the cumulative inflation responses. The blue line is based on the HDNLP estimator, and the dashed line is based on linear local projection estimator. See text.

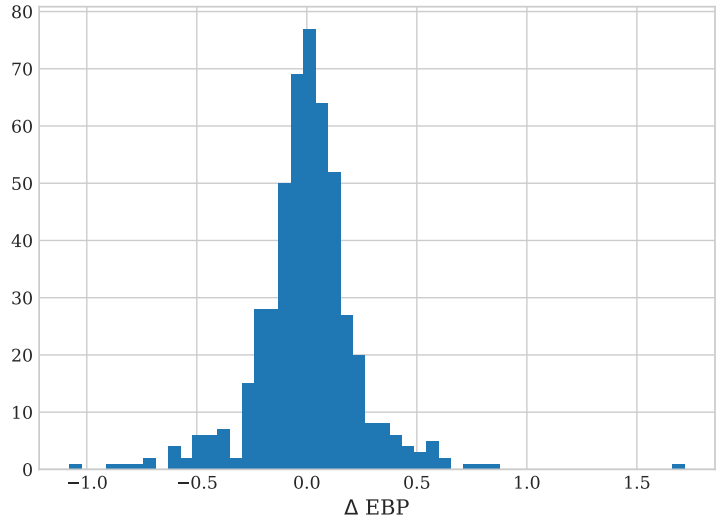
To identify the effects of financial shocks, we include contemporaneous slow-moving macroeconomic variables (INDPRO, CPIAUCSL, and UNRATE) and 12 lags of all variables as controls, following the timing restrictions from Gilchrist and Zakrajšek (2012). The resulting high-dimensional control set contains 87 regressors, rendering standard nonpara-

[11] We replace the federal funds rate by the Wu and Xia (2016) shadow rate when the federal funds rate hits the zero lower bound.

metric approaches computationally infeasible.

Figure 6 is organized into 4 panels, with the top row displaying the responses of inflation and the unemployment rate calculated with traditional local projections (or LP for short, shown as a dashed line) and using our HDNLP estimator (shown as a solid line). The bottom row displays the response of EBP to its own shock as reference in panel (c), and the cumulative response ratio (CRR) of unemployment to inflation, i.e., $\text{CRR}_h = \frac{\sum_{j=1}^{h} \theta_j^{\text{UR}}}{\sum_{j=1}^{h} \theta_j^{\text{CPI}}}$ is shown on panel (d). The CRR measures the average trade-off between unemployment and inflation, that is, how much unemployment is generated for each percentage point decline in the rate of inflation. Thus the CRR normalization adjusts for differences in the impulse response of EBP that could explain the differences in the responses of other variables. Moreover, the CRR provides a smoother and more easily interpretable statistic.[12]

Figure 7: Distribution of EBP Changes



Notes: Distribution of monthly changes in EBP.

The CRR displayed in panel (d) of Figure 6 thus provides a clear distinction between the LP and HDNLP estimators. Following a shock to financial conditions as captured in panel (c) of that figure, for every percentage point decline in the rate of inflation, the unemployment rate increases by about 0.8 percentage point after about 18 months when relying on the

[12]Responses of other variables not reported in Figure 6 are available upon request.

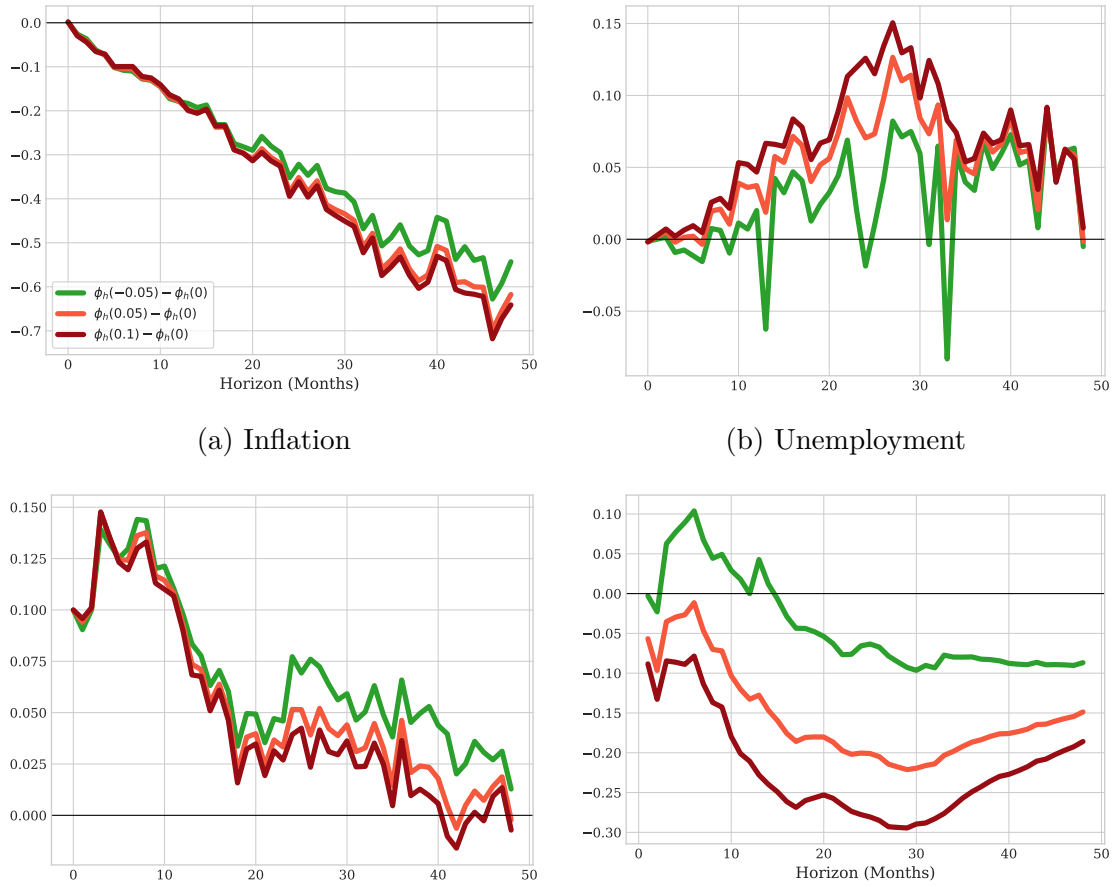
LP estimator, whereas that number is closer to 0.2 percentage point and bottoming out at around 30 months after impact for the HDNLP estimator, a trade-off that is about 4 times smaller and more delayed. The reason is that the response of inflation appears to be stronger with the HDNLP estimator since the differences in the unemployment response are de minimis.

However, Barnichon et al. (2022) and Forni et al. (2024) show that bad financial shocks have larger effects than good shocks. Hence, next we investigate whether the sign and the size of shocks to financial conditions have different effects on economic conditions. In order to explore this type of asymmetry, Figure 7 displays the distribution of EBP shocks in our dataset. Based on this empirical distribution, we consider three scenarios: a modest favorable financial shock (an EBP decline of 0.05), a modest adverse shock (an EBP increase of 0.05), and a large adverse shock (an EBP increase of 0.10). The resulting responses are reported in Figure 8. For ease of comparison, we normalize all responses to have the same EBP response on impact.

Figure 8 is organized into the same 4 panels of Figure 6. We normalize the response to a negative EBP shock by flipping the sign of the response to make the responses under each scenario easier to compare. Because of the normalization of the EBP shock used in panels (a)-(c) of the figure (and the noise of the responses themselves) one would conclude that the data offers little evidence that credit shocks have asymmetric effects on the economy that depend on the magnitude of the EBP shock. However, panel (d) again displays the CRR, which avoids potential normalization issues, as we discussed earlier.

Thus, under the intermediate scenario, we largely replicate the CRR chart shown in panel (d) of Figure 6. That is, the CRR bottoms out at about 30 months with an implied trade-off of 0.2 percentage point in the unemployment rate for every decrease of 1 percentage point in the rate of inflation. Now consider the more favorable scenario (displayed in green color). The trade-off bottoms out at about 0.1 percentage point (rather than 0.2). In contrast, the scenario where the economy experiences a larger contraction in credit suggest a trade-off of

Figure 8: Average Asymmetric Responses for Financial Shocks



(a) Inflation

(b) Unemployment

(c) EBP

(d) Cumulative Response Ratio

Notes: Cumulative response ratio refers to the ratio of cumulative unemployment responses relative to the cumulative inflation responses. Horizons beyond $h = 40$ months are excluded due to numerical instability.

about 0.3. Note that these differences are not driven by the size of the EBP shock considered. This is the reason for focusing on the CRR, which takes care of these normalization issues. Hence, the difference between a favorable credit shock versus a very unfavorable credit shock can be 3 times as large. This is in line with Alessandri et al. (2025); Barnichon et al. (2022) and Forni et al. (2024).

7 Conclusion

This paper develops a two-step nonparametric estimator for impulse response functions in high-dimensional local projections setup, where we combine Neyman-orthogonal moments and cross-fitting to estimate the average structural function. The proposed method allows the structural function to be fully nonlinear, accommodates many high-dimensional controls, and remains computationally feasible in empirically relevant sample sizes.

Applying the method to study the effects of financial shocks in the US, we document average impulse responses that are qualitatively consistent yet quantitatively different from their linear local projection counterparts. This suggests linear models systematically underestimate the real damage from financial stress. We also illustrate how the effects of financial shocks exhibit significant nonlinearities, revealing both a strong sign asymmetry based on the shock's direction and a clear size-dependency based on its magnitude.

Several extensions remain open for future work. First, our pseudo-outcome construction naturally extends to quantiles or distribution functions, potentially avoiding the high-dimensional integration required by existing methods (e.g., Su et al., 2019). This allows us to recover distributional impulse responses—tracking how shocks affect the entire distribution of outcomes. Second, while we establish pointwise asymptotic normality, uniform inference over regions or functional spaces can be particularly useful for policy analysis. Finally, although we mainly focus on unconditional impulse responses, the proposed method can be readily applied to estimate conditional responses, i.e., heterogeneous treatment effects, given observable characteristics $\boldsymbol{W}_t$. We leave these extensions to future work.

References

Adamek, R., Smeekes, S., and Wilms, I. (2024). Local projection inference in high dimensions. *The Econometrics Journal*, page utae012.

Ai, C., Linton, O., Motegi, K., and Zhang, Z. (2021). A unified framework for efficient estimation of general treatment models. *Quantitative Economics*, 12:779–816.

Alessandri, P., Jordà, Ò., and Venditti, F. (2025). Decomposing the monetary policy multiplier. *Journal of Monetary Economics*, page 103783.

Angrist, J. D., Jordà, Ò., and Kuersteiner, G. M. (2018). Semiparametric estimates of monetary policy effects: String theory revisited. *Journal of Business and Economic Statistics*, 36:371–387.

Angrist, J. D. and Kuersteiner, G. M. (2011). Causal effects of monetary shocks: Semiparametric conditional independence tests with a multinomial propensity score. *The Review of Economics and Statistics*, 93:725–747.

Babii, A., Ghysels, E., and Striaukas, J. (2022). Machine learning time series regressions with an application to nowcasting. *Journal of Business & Economic Statistics*, 40:1094–1106.

Ballinari, D. and Wehrli, A. (2024). Semiparametric inference for impulse response functions using double/debiased machine learning.

Barnichon, R. and Matthes, C. (2018). Functional approximation of impulse responses. *Journal of Monetary Economics*, 99:41–55.

Barnichon, R., Matthes, C., and Ziegenbein, A. (2022). Are the effects of financial market disruptions big or small? *Review of Economics and Statistics*, 104:557–570.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, 81:608–650.

Berbee, H. (1987). Convergence rates in the strong law for bounded mixing sequences. *Probability Theory and Related Fields*, 74:255–270.

Brunnermeier, M. K., Eisenbach, T. M., and Sannikov, Y. (2013). Macroeconomics with financial frictions: a survey. In *Advances in economics and econometrics: Volume 2, applied economics: Tenth world congress*, volume 50, page 3. Cambridge University Press.

Brunnermeier, M. K. and Sannikov, Y. (2014). A macroeconomic model with a financial sector. *American Economic Review*, 104:379–421.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. In Heckman, J. J. and Leamer, E., editors, *Handbook of Econometrics*, volume 6, pages 5549–5632. North-Holland, Amsterdam.

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. *The Econometrics Journal*, 21:C1–C68.

Colangelo, K. and Lee, Y.-Y. (2026). Double debiased machine learning nonparametric inference with continuous treatments. *Journal of Business & Economic Statistics*, 44:67–79.

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer, New York Berlin Heidelberg.

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.

Gao, L., Shao, Q.-M., and Shi, J. (2022). Refined cramér-type moderate deviation theorems for general self-normalized sums with applications to dependent random variables and winsorized mean. *The Annals of Statistics*, 50:673–697.

Gertler, M. and Karadi, P. (2015). Monetary policy surprises, credit costs, and economic activity. *American Economic Journal: Macroeconomics*, 7:44–76.

Giannone, D., Lenza, M., and Primiceri, G. E. (2021). Economic predictions with big data: The illusion of sparsity. *Econometrica*, 89:2409–2437.

Gilchrist, S. and Zakrajšek, E. (2012). Credit spreads and business cycle fluctuations. *American Economic Review*, 102:1692–1720.

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

Gonçalves, S., Herrera, A. M., Kilian, L., and Pesavento, E. (2024b). State-dependent local projections. *Journal of Econometrics*, 244:105702.

Goulet Coulombe, P., Leroux, M., Stevanovic, D., and Surprenant, S. (2022). How is machine learning useful for macroeconomic forecasting? *Journal of Applied Econometrics*, 37:920–964.

Gutmann, M. U. and Hyvärinen, A. (2012). Noise-contrastive estimation of unnormalized statistical models, with applications to natural image statistics. *Journal of Machine Learning Research*, 13:307–361.

Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, 24:726–748.

Hastie, T., Tibshirani, R., and Friedman, J. H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, NY, 2nd ed edition.

Hauzenberger, N., Huber, F., Klieber, K., and Marcellino, M. (2025). Bayesian neural networks for macroeconomic analysis. *Journal of Econometrics*, 249:105843.

Hu, R., Sejdinovic, D., and Evans, R. J. (2024). A kernel test for causal association via noise contrastive backdoor adjustment. *Journal of Machine Learning Research*, 25:1–56.

Imai, K. and Ratkovic, M. (2014). Covariate balancing propensity score. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 76:243–263.

Imbens, G. W. and Newey, W. K. (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica*, 77:1481–1512.

Imbens, GW. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika*, 87:706–710.

Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. *American Economic Review*, 95:161–182.

Kennedy, E. H., Ma, Z., McHugh, M. D., and Small, D. S. (2017). Non-parametric methods for doubly robust estimation of continuous treatment effects. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 79:1229–1245.

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.

Matsushita, Y., Otsu, T., and Takahata, K. (2023). Estimating density ratio of marginals to joint: Applications to causal inference. *Journal of Business & Economic Statistics*, 41:467–481.

McCracken, M. W. and and Ng, S. (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34:574–589.

Medeiros, M. C., Vasconcelos, G. F. R., Veiga, Á., and Zilberman, E. (2021). Forecasting inflation in a data-rich environment: The benefits of machine learning methods. *Journal of Business & Economic Statistics*, 39:98–119.

Miranda-Agrippino, S. and Ricco, G. (2021). The transmission of monetary policy shocks. *American Economic Journal: Macroeconomics*, 13:74–107.

Nakamura, E. and Steinsson, J. (2018). High-frequency identification of monetary non-neutrality: The information effect. *Quarterly Journal of Economics*, 133:1283–1330.

Paranhos, L. (2025). How do firms' financial conditions influence the transmission of monetary policy? a non-parametric local projection approach. *Journal of Econometrics*, 249:105886.

Ramey, V. A. (2016). Macroeconomic shocks and their propagation. In Taylor, J. B. and Uhlig, H., editors, *Handbook of Macroeconomics*, volume 2, pages 71–162. Elsevier, Amsterdam.

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

Robins, J. M., Hernán, M. Á., and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology:. *Epidemiology*, 11:550–560.

Romer, C. D. and Romer, D. H. (2004). A new measure of monetary shocks: Derivation and implications. *American Economic Review*, 94:1055–1084.

Semenova, V. and Chernozhukov, V. (2021). Debiased machine learning of conditional average treatment effects and other causal functions. *The Econometrics Journal*, 24:264–289.

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.

Su, L., Ura, T., and Zhang, Y. (2019). Non-separable models with high-dimensional data. *Journal of Econometrics*, 212:646–677.

Takatsu, K. and Westling, T. (2024). Debiased inference for a covariate-adjusted regression function. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, page qkae041.

Wu, J. C. and Xia, F. D. (2016). Measuring the macroeconomic impact of monetary policy at the zero lower bound. *Journal of Money, Credit and Banking*, 48:253–291.

Wu, X., Mealli, F., Kioumourtzoglou, M.-A., Dominici, F., and Braun, D. (2024). Matching on generalized propensity scores with continuous exposures. *Journal of the American Statistical Association*, 119:757–772.