

# Causal State-Dependent Local Projections

Joel M. David, Raffaella Giacomini, Xiyu Jiao,  
and Weining Wang\*

Discussion Paper 26/829  
7<sup>th</sup> May 2026

## **School of Economics**

University of Bristol  
Priory Road Complex  
Bristol  
BS8 1TU  
United Kingdom



# Causal State-Dependent Local Projections

Joel M. David, Raffaella Giacomini, Xiyu Jiao, and Weining Wang\*

May 7, 2026

## Abstract

State-dependent local projections (LPs) are widely used to estimate how responses to exogenous aggregate shocks vary as a function of observable state variables, yet their causal interpretation remains unclear. We show that this interpretation obtains under the sufficient condition that the conditional mean is linear in the aggregate shock at each horizon, and that this condition holds in a broad class of canonical micro–macro environments, including first-order perturbation solutions of heterogeneous-agent models and macro-finance models. Under this condition, LPs recover causal impulse responses without requiring specification of the full data-generating process. We further show that the causal interpretation of state-dependent LPs is robust to the choice of state variable. By contrast, commonly used linear interaction LPs generally fail to recover causal objects. We therefore develop a sieve-based nonparametric LP estimator that restores causal interpretation and delivers valid pointwise and uniform inference in micro–macro panels. Empirically, allowing for nonparametric state dependence materially changes both the pattern of heterogeneous firm investment responses and their aggregate implications for the transmission of monetary policy shocks.

**JEL Classification:** C14, C12, C23, E32

**Keywords:** Impulse Response Functions, Identification, Nonparametric Estimation, Uniform Inference, Micro-Macro Panels, Heterogeneous-Agent Models

---

\*David: Federal Reserve Bank of Chicago (joel.david@chi.frb.org); Giacomini: Department of Economics, University College London (r.giacomini@ucl.ac.uk); Jiao: Department of Economics, University of Gothenburg (xiyu.jiao@economics.gu.se); Wang: Department of Economics, University of Bristol (weining.wang@bristol.ac.uk). Acknowledgements. We would like to thank Marco Brianti and Wei Cui for useful comments and suggestions and William Pennington for excellent research assistance. The views expressed here are those of the authors and not necessarily those of the Federal Reserve Bank of Chicago or the Federal Reserve System.

# 1 Introduction

State-dependent local projections (LPs) are widely used in empirical practice, yet their causal interpretation remains unclear, even under standard shock exogeneity assumptions. Kolesár and Plagborg-Møller (2024) show that scalar LPs (regressing outcomes on the shock) can admit a causal interpretation under shock exogeneity, even when the data-generating process (DGP) is nonlinear. Winkler (2026) extends this perspective to state-dependent specifications, showing that LP coefficients can be expressed as weighted averages of underlying effects. By contrast, Gonçalves et al. (2024b) show that state-dependent LPs can fail to recover causal impulse responses in nonlinear environments, even under exogeneity. As emphasized in the survey by Jordà and Taylor (2025), extending causal interpretations beyond scalar LPs to nonlinear or state-dependent settings remains an open challenge. Taken together, these results reveal a fundamental tension: while LP estimates can always be written as averages of underlying effects, it is unclear whether these averages correspond to economically meaningful causal impulse responses. This question is central for empirical practice, where state-dependent LPs are routinely interpreted as measuring how the effects of aggregate shocks vary across economic conditions.

This paper resolves this tension by providing a complete characterization of when state-dependent LPs admit a meaningful causal interpretation as impulse responses. It makes three main contributions. First, we establish that state-dependent LPs recover causal impulse responses under a sufficient condition on the conditional mean. We show that this condition is naturally satisfied in a broad class of canonical micro–macro environments, but is less plausible in purely aggregate settings, where the concerns emphasized by Gonçalves et al. (2024b) remain relevant. This condition is a property of the underlying DGP, while estimation proceeds via standard LPs under shock exogeneity. Second, we clarify the role of specification choices for causal interpretation when this condition is satisfied. While the choice of state variable does not by itself invalidate causal interpretation, imposing linear state dependence generally leads to objects that do not admit a meaningful causal interpretation, except in the knife-edge case in which the true state dependence is exactly linear. Third, we develop a practical sieve-based nonparametric approach to estimating state-dependent impulse responses with LPs and establish valid inference for these objects. When the sufficient condition holds, our approach recovers the causal impulse response function and delivers both pointwise and uniform confidence bands in micro–macro panels. The inference builds on existing results but adapts them to the nonstandard structure of local projections with macro shocks and micro-level state variables.

The central takeaway for empirical researchers is that state-dependent LPs admit a causal interpretation under restrictions on the underlying DGP that are satisfied in a broad class of canonical micro–macro environments. These restrictions do not require specifying the DGP: with an exogenous shock, state-dependent LPs recover causal impulse responses directly, but doing so generally requires moving beyond commonly used linear interaction specifications and adopting a nonparametric approach to modeling state dependence.

Throughout the paper, we focus on micro–macro settings, as these are the environments in

which identifying conditions are most plausibly satisfied. State-dependent LPs are widely used in such settings to study how aggregate shocks affect heterogeneous households and firms. In practice, researchers implement these specifications by regressing a micro outcome  $Y_{i,t+h}$  on an aggregate shock  $X_t$  interacted with micro-level characteristics  $Z_{i,t-1}$ , such as balance sheet variables, financial constraints, or other predetermined state variables, typically imposing linear state dependence (i.e., including  $Z_{i,t-1}X_t$  as a regressor).

Formally, the sufficient condition that delivers a causal interpretation of state-dependent LPs is that the horizon- $h$  conditional mean of the micro outcome is linear in the aggregate shock with a functional coefficient  $g_h(Z_{i,t-1})$  depending on the state:

$$\mathbb{E}[Y_{i,t+h} \mid \mathcal{F}_{t-1}, X_t] = g_h(Z_{i,t-1})X_t + r_{i,h}(\mathcal{F}_{t-1}), \quad (1)$$

where  $\mathcal{F}_{t-1}$  denotes the information available at time  $t-1$ . Under this condition,  $g_h(z)$  equals the total causal impulse response at horizon  $h$  for units with initial state  $z$ . This condition is a restriction on the underlying DGP and does not require specifying or estimating the conditional mean. The key insight is that linearity in the aggregate shock allows the impulse response to be represented as a state-dependent coefficient in a LP, so that identification follows directly under shock exogeneity. The nuisance component  $r_{i,h}(\mathcal{F}_{t-1})$  of the conditional mean need not be modeled. Importantly, identification does not depend on the precise choice of state  $Z$ : as long as it is predetermined, we show that a misspecified state still yields a valid causal interpretation, but as a weighted average over the underlying heterogeneity that is not captured. If condition (1) fails, however, state-dependent LP coefficients are generally not causal.

We then characterize economically relevant classes of micro-macro DGPs under which condition (1) holds and fails. In environments where the aggregate shock enters linearly and propagation dynamics are time-invariant, state-dependent LPs recover causal impulse responses. This structure arises in a broad class of widely used models. In particular, local log-linear solutions of Heterogeneous-Agents (HA) and Heterogeneous Agents New Keynesian (HANK) models and macro-finance models imply that aggregate shocks generate state-dependent exposure across units while leaving the underlying propagation mechanisms unchanged. This implies that the sufficient condition is satisfied in these models. Concretely, a contractionary monetary shock may reduce consumption more for highly leveraged households, depress investment more for highly leveraged firms, and lower wealth more for holders of long-duration assets, even though the structural laws governing consumption smoothing, capital accumulation, and wealth dynamics remain unchanged. When this structure is motivated by linearized solutions, it requires that the first-order approximation remains accurate over the range of aggregate shocks considered. In these environments, state-dependent LPs can recover causal impulse responses directly, without requiring a full specification of the underlying model.

Departures from this structure can arise in a number of ways. One occurs when the relationship between outcomes and the aggregate shock is nonlinear, as in models with occasionally binding constraints, regime switching, or in global nonlinear solutions of heterogeneous-agent models. Another,

more subtle, violation arises when shocks affect not only exposure but also the propagation dynamics of micro outcomes, so that the total response, combining direct and indirect effects, cannot be summarized by a coefficient that depends only on the state. In these cases, state-dependent LP coefficients do not correspond to causal impulse responses, as illustrated by the aggregate setting in Gonçalves et al. (2024b). The micro–macro model of Almuzara et al. (2025) provides a related example that falls outside the class we characterize. Recovering causal impulse responses then requires allowing for fully nonlinear responses to the shock.

That the condition holds in canonical micro–macro environments does not, however, justify the linear state-dependent LP specifications commonly used in empirical work. Even when the sufficient condition holds, recovering the causal impulse response generally requires nonparametric estimation of  $g_h(\cdot)$ . Unless the true state dependence is exactly linear, we show that the commonly used linear specifications capture projection objects that do not generally have a causal interpretation. We therefore develop an easy-to-implement sieve estimation approach that directly recovers the causal state-dependent impulse response function and delivers valid pointwise and uniform inference.

Monte Carlo evidence illustrates the practical implications of our approach. Allowing for nonparametric state dependence is quantitatively important, especially at short horizons and for larger shocks, where linear specifications can substantially distort the impulse response function. Uniform confidence bands exhibit performance consistent with the nonparametric inference literature with data-driven tuning, including some undercoverage in smaller samples, but coverage improves as the time dimension increases. In addition, intermediate terms – future shock–state interactions that enter the LP error and are typically omitted because they do not affect identification – can be exploited to obtain substantially tighter confidence bands.

Applying our approach to firm-level investment responses to monetary policy shocks, we revisit the widely used LP in Ottonello and Winberry (2020), which imposes linear state dependence in firms’ financial conditions, e.g., distance to default. We show that their estimated positive slope is in large part a misspecification artifact: our approach reveals a significantly nonlinear, hump-shaped response, with the largest effects concentrated among firms near the mean of the distribution rather than at the least risky right tail. This pattern of firm-level responsiveness has important quantitative implications: because most firms are concentrated in regions of the distribution where the nonlinear response exceeds the linear approximation, the latter effectively averages across states and attenuates the estimated effects. As a result, our nonparametric estimates imply substantially larger aggregate effects of monetary policy shocks on firms’ investment.

Our framework connects and clarifies several strands of the literature. We speak directly to the empirical literature that studies heterogeneous responses to aggregate shocks using LPs that interact shocks with micro characteristics (e.g., Cloyne et al., 2023; Ottonello and Winberry, 2020; Jeenas, 2023). Our message is as follows. State-dependent LPs can be given a causal interpretation in micro–macro settings where shocks affect units differently through their characteristics but do not change how the economy propagates shocks over time, as in standard linearized heterogeneous-agent and macro-finance models. In these environments, researchers can continue to use state-dependent

LPs to estimate heterogeneous effects. However, commonly used linear interaction specifications generally do not recover causal impulse responses, even when the causal interpretation is valid. Instead, state dependence should be modeled nonparametrically, which can be implemented with a simple modification of standard LP regressions. By contrast, when shocks affect propagation itself, as is more likely in aggregate time-series applications, a causal interpretation is less credible.

Our results reinterpret and extend the recent econometric literature on LPs. Kolesár and Plagborg-Møller (2024) show that scalar LPs admit a causal interpretation as weighted averages of marginal effects under shock exogeneity, even when the true DGP is nonlinear. Bousquet (2025) specializes this result to settings with nonlinearities in the shock. We show that this robustness does not generally extend to state-dependent LPs once heterogeneity is parametrically restricted. Under additional structure on the DGP, however, state-dependent LPs recover the causal impulse response function itself - rather than a weighted average - provided that the state dependence is estimated nonparametrically.

Gonçalves et al. (2024b) show that state-dependent LPs need not recover causal impulse responses when the state is endogenous. Our analysis complements their insight by showing that the crucial mechanism for such failure is nonlinearity of the shock and/or shock-dependent propagation. While both are plausible in aggregate time series, we show that in micro-macro panels there exists a large and economically relevant class of DGPs in which shocks enter linearly and propagation remains stable. In these environments state-dependent LPs are valid even when the state is endogenous (again, provided that state dependence is estimated nonparametrically).

Gonçalves et al. (2024a) recover impulse responses in nonlinear environments by estimating the full response of the outcome as a function of the shock and then constructing impulse responses as differences in this response function across shock realizations. By contrast, we restrict attention to environments in which the conditional mean is linear in the aggregate shock with a state-dependent coefficient. This structure allows us to recover impulse responses directly as the coefficient function in a LP, without estimating the full response surface or performing a second-step transformation, and under weaker martingale-difference exogeneity of the shock rather than i.i.d. shocks.

From an econometric perspective, our estimator can be viewed as a functional coefficient regression in which the coefficient on the aggregate shock is identified by exogeneity. This structure allows us to recover the causal object directly without specifying or estimating the remaining components of the conditional mean, distinguishing our approach from standard implementations that require explicit modeling of nuisance components.

Relative to work on inference in micro-macro panels with aggregate shocks, our contribution is complementary but distinct. Almuzara et al. (2025) study inference for scalar LPs, but their framework does not extend to the state-dependent impulse responses we consider. We instead provide pointwise and uniform inference for state-dependent LPs, building on recent advances in sieve and high-dimensional time-series methods (e.g., Chernozhukov et al., 2014; Zhang and Wu, 2017; Lu et al., 2020; Chen et al., 2024) and carefully adapting them to the macro-micro LP setting, which raises nonstandard challenges for inference.

Our empirical application is related to Paranhos (2025), who proposes a nonparametric extension of Ottonello and Winberry (2020) using random forests to flexibly estimate heterogeneous impulse responses. While her approach relaxes functional form assumptions on how responses vary across firms, it takes as given that such state-dependent impulse responses admit a causal interpretation. In contrast, our contribution is to characterize the conditions under which these objects are identified in the first place. We further show that, even when these conditions hold, a nonparametric representation of state dependence is generally required for the resulting impulse responses to retain a causal interpretation. Within such settings, her approach can be viewed as an alternative way of estimating the same object. Our sieve-based implementation, however, remains closer to standard LP practice and delivers uniform inference for the full impulse response function.

Overall, our paper clarifies when state-dependent LPs admit a causal interpretation and provides practical tools for estimation and inference of heterogeneous responses to aggregate shocks, with particular relevance for micro–macro settings.

## 2 Conceptual Framework and Illustrative Data-Generating Processes

This illustrative section defines the state-dependent causal impulse response of interest and presents the key condition under which state-dependent LPs admit a causal interpretation. It then uses three stylized DGPs and three different specifications of LPs to clarify when a causal interpretation holds and when it fails.

Consider the following stylized nonlinear DGPs for a scalar micro outcome  $Y_{it}$ , exogenous aggregate shock  $X_t$ , and predetermined micro state  $Z_{i,t-1}$ :

$$\text{DGP 1 (linear shock, constant propagation): } Y_{it} = g(Z_{i,t-1})X_t + \rho Y_{i,t-1} + \varepsilon_{it},$$

$$\text{DGP 2 (shock-dependent propagation): } Y_{it} = g(Z_{i,t-1})X_t + \rho_t Y_{i,t-1} + \varepsilon_{it},$$

$$\text{DGP 3 (nonlinear shock): } Y_{it} = f(Z_{i,t-1}, X_t) + \rho Y_{i,t-1} + \varepsilon_{it},$$

where  $g(\cdot)$  is nonlinear in  $Z$ ,  $f(\cdot, \cdot)$  is nonlinear in  $X$ , and in DGP 2 the propagation coefficient  $\rho_t$  depends on the aggregate shock, so that perturbations to  $X_t$  affect future propagation.

These DGPs isolate three distinct channels through which aggregate shocks affect outcomes: heterogeneous exposure to a shock that enters linearly, but with constant propagation (DGP 1), shock-dependent propagation (DGP 2), and nonlinear dependence on the shock (DGP 3). As we show below, only in the first case state-dependent LPs recover causal impulse responses.

Our object of interest is the finite- $\delta$  impulse response at horizon  $h$  for a given state,

$$\text{IRF}_h(\delta \mid z) = \mathbb{E}[Y_{i,t+h}(\delta) - Y_{i,t+h} \mid Z_{i,t-1} = z] = \mathbb{E}[\mathbb{E}(Y_{i,t+h}(\delta) - Y_{i,t+h} \mid \mathcal{F}_{t-1}, X_t) \mid Z_{i,t-1} = z],$$

where  $Y_{i,t+h}(\delta)$  is the potential outcome when the shock at time  $t$  is perturbed from  $X_t$  to  $X_t + \delta$ .

The central question is whether the total effect of a perturbation to  $X_t$  at horizon  $h$  can be expressed as a function of the initial state multiplying the shock. If so, the impulse response admits a representation

$$\text{IRF}_h(\delta \mid z) = g_h(z) \delta,$$

where  $g_h(z)$  captures both the direct effect of the shock and all indirect effects arising from its propagation over time. This multiplicative structure is the key condition under which LPs recover state-dependent causal impulse responses.

Under DGP 1, a perturbation  $\delta$  changes  $Y_{it}$  by  $g(Z_{i,t-1})\delta$ . Because propagation is linear and time-invariant, the key condition is satisfied:

$$\text{IRF}_h(\delta \mid z) = \rho^h g(z) \delta.$$

This extends to more general lag structures: if the shock enters linearly and propagation dynamics are time-invariant, the impulse response admits the multiplicative form  $\text{IRF}_h(\delta \mid z) = g_h(z)\delta$ . All dynamic feedback effects are absorbed into the single coefficient  $g_h(z)$ .

Under DGP 2, the key condition fails because a perturbation to  $X_t$  affects future propagation. At  $h = 1$ ,

$$\mathbb{E}[Y_{i,t+1}(\delta) - Y_{i,t+1} \mid \mathcal{F}_{t-1}, X_t] = \mathbb{E}[\rho_{t+1}(\delta) Y_{it}(\delta) - \rho_{t+1} Y_{it} \mid \mathcal{F}_{t-1}, X_t],$$

so the response depends on both the perturbed outcome  $Y_{it}(\delta)$  and the perturbed propagation coefficient  $\rho_{t+1}(\delta)$ . Because both objects vary with the shock, the effect involves their interaction and is not, in general, proportional to  $\delta$  with a coefficient depending only on the initial state. This environment is a micro-macro analogue of the aggregate DGP considered by Gonçalves et al. (2024b) where state-dependent LPs fail. Our formulation highlights that this failure arises from shock-dependent propagation, rather than from endogeneity of the state per se.

Under DGP 3, nonlinearity in the shock means that even at  $h = 0$ :

$$Y_{it}(\delta) - Y_{it} = f(Z_{i,t-1}, X_t + \delta) - f(Z_{i,t-1}, X_t),$$

which depends on the level of  $X_t$  and is not proportional to  $\delta$ . No representation of the form  $g_h(Z_{i,t-1})\delta$  for the impulse response generally exists and the key condition fails.

In sum, only in environments like DGP 1 can the impulse response at each horizon be expressed as a state-dependent coefficient on the shock.

We now discuss what different LP specifications recover in these different DGPs. Consider:

$$\text{Scalar LP: } Y_{i,t+h} = \beta_h X_t + u_{i,t+h}, \quad (2a)$$

$$\text{Linear state dependence LP: } Y_{i,t+h} = \alpha_h X_t + \beta_h Z_{i,t-1} X_t + u_{i,t+h}, \quad (2b)$$

$$\text{Nonparametric state dependence LP (our proposal): } Y_{i,t+h} = b_h^\top \phi(Z_{i,t-1}) X_t + u_{i,t+h}. \quad (2c)$$

where  $\phi(\cdot)$  is a vector of basis functions allowing for flexible state-dependence.



Table 1: What LPs recover under three stylized DGPs

|                                                                    | DGP 1                                  | DGP 2                 | DGP 3                 |
|--------------------------------------------------------------------|----------------------------------------|-----------------------|-----------------------|
| <b>Scalar LP</b>                                                   | Average causal effect                  | Average causal effect | Average causal effect |
| <b>Linear state dependence LP (common empirical specification)</b> | Projection (not causal in general)     | Projection            | Projection            |
| <b>Nonparametric state dependence LP (our proposal)</b>            | Causal state-dependent effect $g_h(z)$ | Projection            | Projection            |

Kolesár and Plagborg-Møller (2024) show that scalar LPs as (2a) retain a clean average causal interpretation even when the DGP is nonlinear, and thus the result applies to all three DGPs.

By contrast, an LP with linear state dependence as (2b) forces the regression to approximate  $g_h(z)$  by its best linear projection. Unless  $g_h(\cdot)$  is exactly linear, we show that the interaction slope  $\beta_h$  does not coincide with the impulse response at any particular state or with any well-defined aggregation of state-specific impulse responses.

This reveals a fundamental asymmetry. The robustness of scalar LPs relies on averaging across heterogeneous effects. Once one conditions on the state, this logic no longer protects the estimator from functional-form misspecification. Modeling heterogeneity is therefore more demanding than estimating average effects. Many empirical papers interpret the coefficient on a linear shock–state interaction as the causal effect for high- versus low-state units. Our analysis shows that such an interpretation is generally not justified when state dependence is nonlinear.

Allowing flexible nonparametric dependence on the state  $Z_{i,t-1}$  in our proposed LP specification - such as (2c) - instead permits the regression to recover directly the causal impulse response function  $g_h(z)$  if the data are generated by DGP 1.

In DGP 2 and DGP 3, by contrast, the impulse response does not admit a multiplicative representation in the first place and state-dependent LP coefficients (both linear and nonparametric) are generally projections without a clean causal interpretation.

Table 1 summarizes the central message of the paper.

In the remainder of the paper, we (i) derive a sufficient condition under which state-dependent LPs recover causal impulse responses without specifying the full DGP; (ii) characterize when this condition holds or fails, admitting environments with state-dependent exposure and constant propagation (such as DGP 1) while excluding those with shock-dependent propagation or nonlinear-in-shock effects (such as DGP 2–3); and (iii) show that standard linear state-dependent LPs do not recover this object except in a knife-edge case, motivating a nonparametric approach. We argue that this admissible class includes many canonical micro–macro DGPs and develop a nonparamet-

ric approach for estimation and inference of causal state-dependent impulse responses via LPs in these settings.

### 3 Identification of State-Dependent Impulse Responses

We now discuss identification and the conditions under which it holds or fails.

#### 3.1 Identification and the Limits of Linear State-Dependent Local Projections

This section provides a general characterization of when LPs recover state-dependent causal impulse responses. Rather than specifying a full structural model or DGP, we impose a sufficient condition on the horizon-specific conditional mean that delivers a causal interpretation of state-dependent LPs. We then show that, even under this condition, standard linear state-dependent LPs generally fail to recover the causal object, motivating our nonparametric approach.

We observe a panel  $\{Y_{it}, X_t, Z_{it}, W_{it}\}$  for  $i = 1, \dots, N$  and  $t = 1, \dots, T$ , where  $Y_{it}$  is a micro outcome,  $X_t$  is an aggregate shock,  $Z_{it}$  is a micro state, and  $W_{it}$  collects control variables that may be included in the LP specification. Let  $\mathcal{F}_{t-1}$  denote the information available at time  $t - 1$ ,

$$\mathcal{F}_{t-1} = \sigma(\{Y_{is}, Z_{is}, W_{is}\}_{i,s \leq t-1}, \{X_s\}_{s \leq t-1}).$$

Our object of interest is the state-dependent impulse response of  $Y$  to a one-time perturbation of the realized shock at date  $t$ . Let  $Y_{i,t+h}(\delta)$  denote the potential outcome when the realized shock at date  $t$  is increased from  $X_t$  to  $X_t + \delta$ , holding all other shocks at their realized values. The finite- $\delta$  causal effect conditional on the initial micro state is

$$\text{IRF}_Y(h, \delta \mid Z_{i,t-1} = z) := \mathbb{E}[Y_{i,t+h}(\delta) - Y_{i,t+h} \mid Z_{i,t-1} = z].$$

For each horizon  $h \geq 0$ , we consider the LP

$$Y_{i,t+h} = g_h(Z_{i,t-1}) X_t + W_{i,t-1}^\top \gamma_h + u_{i,t+h}, \quad (3)$$

where  $g_h(\cdot)$  is an unknown function and  $W_{i,t-1}$  are controls. The empirical object of interest is the function  $g_h(\cdot)$ , interpreted as the horizon- $h$  impulse response per unit shock conditional on  $Z_{i,t-1} = z$ .

The following high-level assumptions suffice for identification.

**Assumption 1. (Shock exogeneity).** The aggregate shock is a martingale difference sequence (m.d.s.) with respect to  $\mathcal{F}_{t-1}$ :

$$E[X_t \mid \mathcal{F}_{t-1}] = 0.$$

**Assumption 2. (Predetermined state and controls).**  $(Z_{i,t-1}, W_{i,t-1})$  are  $\mathcal{F}_{t-1}$ -measurable.

**Assumption 3. (Conditional mean condition).** For each horizon  $h \geq 0$ ,

$$\mathbb{E}[Y_{i,t+h} \mid \mathcal{F}_{t-1}, X_t] = g_h(Z_{i,t-1})X_t + r_{i,h}(\mathcal{F}_{t-1}). \quad (4)$$

Assumptions 1–2 are standard: the state and controls are predetermined, and the aggregate shock is exogenous. Assumption 3 is the high-level analogue of the multiplicative structure highlighted in Section 2, and is the key condition for identification. It requires that the total effect of the date- $t$  shock  $X_t$  on the horizon- $h$  conditional mean - including all direct and indirect channels operating through the endogenous evolution of the system - be linear in  $X_t$ , with a slope that depends only on the initial micro state  $Z_{i,t-1}$ . The nuisance component  $r_{i,h}(\mathcal{F}_{t-1})$  is unspecified and unrestricted and need not coincide with the linear projection onto the controls  $W_{i,t-1}$  used in the LP specification (3).<sup>1</sup> In particular, it may contain arbitrary lagged dynamics, fixed effects, and nonlinear dependence on past aggregate and idiosyncratic variables, provided these components are  $\mathcal{F}_{t-1}$ -measurable.

Note that, in many empirical applications, researchers instead use cumulative changes  $Y_{i,t+h} - Y_{i,t}$  as the dependent variable in LPs. Under Assumption 3, the corresponding estimand is  $g_h(z) - g_0(z)$ . Thus, cumulative LPs recover cumulative causal impulse responses under the same conditions.

Under Assumptions 1–3,  $g_h(\cdot)$  is identified:

$$g_h(z) = \frac{\mathbb{E}[X_t Y_{i,t+h} \mid Z_{i,t-1} = z]}{\mathbb{E}[X_t^2 \mid Z_{i,t-1} = z]}.$$

The following result clarifies that a causal interpretation is still available, even when the state variable  $Z_{i,t-1}$  does not coincide with the true state governing heterogeneity.

**Proposition 1 (State choice preserves causal interpretation).** Suppose Assumptions 1–2 hold and that, for some  $s_{i,t-1} \in \mathcal{F}_{t-1}$ ,

$$\mathbb{E}[Y_{i,t+h} \mid \mathcal{F}_{t-1}, X_t] = \tilde{g}_h(s_{i,t-1})X_t + r_{i,h}(\mathcal{F}_{t-1}).$$

Let  $Z_{i,t-1}$  be any  $\mathcal{F}_{t-1}$ -measurable state, and define

$$g_h(z) = \frac{\mathbb{E}[X_t Y_{i,t+h} \mid Z_{i,t-1} = z]}{\mathbb{E}[X_t^2 \mid Z_{i,t-1} = z]}, \quad \mathbb{E}[X_t^2 \mid Z_{i,t-1} = z] > 0.^2$$

<sup>1</sup>The controls  $W_{i,t-1}$  are included in the LP as a projection device, whereas  $r_{i,h}(\mathcal{F}_{t-1})$  denotes the unknown conditional mean component. Note that identification does not require correct specification of  $r_{i,h}(\mathcal{F}_{t-1})$ . Instead, controls serve to partial out predetermined variation, improving efficiency and defining the conditioning set for the impulse response. As a result, different control sets correspond to different conditional impulse responses, each of which is consistently estimated under Assumption 3. Unit fixed effects can be absorbed into  $r_{i,h}(\mathcal{F}_{t-1})$  without affecting identification but shift the variation used to estimate  $g_h(\cdot)$  toward within-unit variation in exposure to the aggregate shock over time.

<sup>2</sup>With predetermined controls, the same result holds after partialling them out.

Then

$$g_h(z) = \mathbb{E}[w_h(s_{i,t-1}; z) \tilde{g}_h(s_{i,t-1}) \mid Z_{i,t-1} = z], \quad w_h(s; z) = \frac{\mathbb{E}[X_t^2 \mid s_{i,t-1} = s, Z_{i,t-1} = z]}{\mathbb{E}[X_t^2 \mid Z_{i,t-1} = z]}.$$

The weights satisfy  $w_h(s; z) \geq 0$  and  $\mathbb{E}[w_h(s_{i,t-1}; z) \mid Z_{i,t-1} = z] = 1$ . Thus,  $g_h(z)$  is a convex (variance-weighted) average of the underlying causal responses among observations with  $Z_{i,t-1} = z$ . For simplicity, we present the analysis for a scalar state variable, but the results extend directly to multivariate states.

*Remark 1* (Interpretation and robustness to state choice). Proposition 1 highlights a robustness property of state-dependent LPs to the choice of state. Even when the researcher does not observe the true source of heterogeneity,  $g_h(z)$  remains causal because  $X_t$  is as-good-as-randomly assigned within any  $\mathcal{F}_{t-1}$ -measurable group.<sup>3</sup> The estimand therefore has a clear interpretation as the weighted average causal response for the subpopulation defined by  $Z_{i,t-1} = z$ , although it may aggregate over underlying heterogeneity in ways that depend on the choice of state. This implies that empirical state-dependent LPs need not recover the true structural source of heterogeneity to remain informative: they identify well-defined subpopulation causal effects, even under coarse or misspecified state variables.

*Remark 2* (Heterogeneity). Writing the response as a common function  $g_h(\cdot)$  does not impose homogeneity across units. If the true response  $\tilde{g}_{h,i}(s_{i,t-1})$  is unit-specific, the same argument implies that  $g_h(z)$  is a convex, variance-weighted average of  $\tilde{g}_{h,i}(s_{i,t-1})$  among observations with  $Z_{i,t-1} = z$ . Thus, heterogeneity is allowed, but the estimand should be interpreted as an average causal response for the subpopulation defined by  $Z_{i,t-1} = z$ .

The causal object of interest is the function  $g_h(\cdot)$  itself. A natural estimator is therefore non-parametric. By contrast, much of the applied literature assumes that  $g_h(\cdot)$  is linear and interprets the slope coefficient as a state-dependent effect. The next result shows that this interpretation fails in general.

**Proposition 2 (Why linear state-dependent LPs generally fail).** Suppose Assumptions 1–3 hold and that  $g_h(\cdot)$  is differentiable almost everywhere. The coefficient on the linear interaction term in a state-dependent LP (e.g., equation (2b) in Section 2) can be written as

$$\beta_h = \int g'_h(z) \omega_h(z) dz,$$

where  $g'_h(z)$  denotes the derivative of the impulse response function, and  $\omega_h(z)$  is a weighting function that depends on the distribution of the data, satisfies  $\int \omega_h(z) dz = 1$ , but is not restricted to be nonnegative. Thus, the linear interaction coefficient does not, in general, correspond to a well-defined causal effect. Appendix A provides a more detailed and formal characterization of this

---

<sup>3</sup>This mirrors the logic of Local Average Treatment Effects (Imbens and Angrist, 1994): conditioning on  $Z_{i,t-1}$  defines a subpopulation, and the LP recovers a positively weighted average of heterogeneous causal responses within that group.

result.

In contrast to the scalar LP case studied by Kolesár and Plagborg-Møller (2024), where the estimand is a positively weighted average of heterogeneous causal effects and admits a clear sub-population interpretation, the weights here need not be positive. As a result, the linear-interaction coefficient does not aggregate causal effects across units but instead combines marginal effects from different regions of the state space with potentially opposing signs. Economically, this breaks the usual interpretation of the coefficient as describing how responses vary with the state: a positive (or negative) estimate need not imply that higher- $z$  units respond more (or less) to the shock.

### 3.2 Economic Interpretation and Admissible Data-Generating Processes

Assumption 3 is a restriction on the horizon- $h$  conditional mean of the outcome, rather than a full characterization of the DGP. The assumption is not testable without further structure, but it is implied by economically meaningful classes of DGPs. This subsection provides an economic interpretation of this condition and highlights classes of canonical micro-macro DGPs in which it is satisfied. Importantly, this characterization is not used for estimation; rather, it clarifies the environments in which state-dependent LPs recover economically meaningful causal impulse responses.

As we show below, a transparent case in which Assumption 3 is satisfied is when the effect of aggregate shocks operates through state-dependent exposure of units to a shock that enters linearly, while propagation dynamics are time-invariant.

#### 3.2.1 Admissible DGPs.

(i) *Log-linearized heterogeneous-agent models.*

Assumption 3 is naturally satisfied by first-order (log-)linear solutions of HA/HANK models. For any household type (grid point)  $i$  with predetermined idiosyncratic state  $s_{i,t-1}$ , the linearization implies that the conditional mean of a micro outcome is affine in the unanticipated aggregate shock, with a type-specific coefficient that may depend flexibly on the micro state. Concretely, perturbation methods used in quantitative HANK, including the Reiter approach formalized by Bayer and Luetticke (2020), write the heterogeneous-agent equilibrium as a nonlinear functional difference equation and approximate it around a stationary equilibrium without aggregate risk. The first-order solution is a linear state-space system: the law of motion for the relevant aggregate and distributional states, and the mapping from these states to individual outcomes, are governed by fixed coefficients that do not depend on the realization of the shock.<sup>4</sup>

Mapping to our notation, let  $Y_{it}$  be any household-level outcome (e.g. consumption or earnings), let  $X_t$  denote the aggregate shock, and let  $Z_{i,t-1}$  be a chosen predetermined micro characteristic.

---

<sup>4</sup>See Bayer and Luetticke (2020) for the perturbation logic and the resulting linear state-space representation: the linearized law of motion for the high-dimensional state vector is governed by a fixed matrix  $O$  and a fixed mapping from states to controls  $G$  (Bayer and Luetticke, 2020, p. 1262).

At impact, the linearized solution implies

$$\mathbb{E}[Y_{i,t} \mid \mathcal{F}_{t-1}, X_t] = g_0(s_{i,t-1})X_t + r_{i,0}(\mathcal{F}_{t-1}).$$

Because all subsequent propagation is governed by the same fixed linear law of motion and policy rules, the effect of the shock at date  $t$  remains linear as the system is iterated forward. Hence, for each horizon  $h$ ,

$$\mathbb{E}[Y_{i,t+h} \mid \mathcal{F}_{t-1}, X_t] = g_h(s_{i,t-1})X_t + r_{i,h}(\mathcal{F}_{t-1}).$$

Thus, Assumption 3 holds exactly within the linearized model. The restriction is not that heterogeneity is linear in the micro state:  $g_h(\cdot)$  may be highly nonlinear because it reflects the nonlinear household problem evaluated at different points of the state space. Rather, the restriction is that the shock affects outcomes through a linearized system whose propagation coefficients are fixed and therefore not themselves altered by the shock.

This structure can be given a microfoundation using the sufficient-statistics representation of Auclert (2019). In his partial-equilibrium micro block, an individual's consumption response to an aggregate shock can be decomposed, to first order, into a substitution effect and a wealth (revaluation) effect given by the product of the individual's marginal propensity to consume (MPC) and a balance-sheet exposure term (Auclert, 2019, pp. 1–2). In our notation, this implies that the response can be written as a loading on the shock that depends on MPCs and balance-sheet positions, which are unrestricted (and potentially nonlinear) functions of the micro state. This decomposition characterizes the cross-sectional structure of the response, while the linearized dynamics that are standard in perturbation-based HANK models ensure that propagation across horizons is governed by time-invariant coefficients, so that the effect at horizon  $h$  can be summarized by  $g_h(s_{i,t-1})$ . This perspective also motivates our empirical choice of a low-dimensional state  $Z_{i,t-1}$ : while the full micro state  $s_{i,t-1}$  is high-dimensional, responses depend primarily on MPCs and balance-sheet exposures, which are low-dimensional functions of predetermined characteristics such as liquid wealth. Conditioning on  $Z_{i,t-1}$  therefore targets the key sources of heterogeneity emphasized by the model.

The preceding discussion is exact for the linearized equilibrium and approximate for the underlying nonlinear economy. In perturbation-based HANK solutions, the equilibrium is approximated around a stationary steady state, typically without aggregate risk. Our object of interest is a *shock perturbation*: the realized aggregate shock is changed from  $X_t$  to  $X_t + \delta$ , and the finite- $\delta$  impulse response compares outcomes under these two shock realizations. Within the linearized system, this comparison is exact for any  $\delta$ , because the approximating equilibrium mapping is linear in the shock by construction. Thus, replacing  $X_t$  by  $X_t + \delta$  changes outcomes proportionally to  $\delta$ , so Assumption 3 holds exactly within the linearized model.

When the linearized system is used as an approximation to the underlying nonlinear economy, the relevant requirement is that the first-order approximation remain accurate over the range of shock realizations being compared, including the paths induced by  $X_t$  and by  $X_t + \delta$ . This is a local approximation requirement, not a restriction on the size of  $\delta$  per se. Moderate perturbations may

be well approximated when the induced path remains in a region where the equilibrium mapping is smooth and close to linear. By contrast, the approximation can deteriorate when the perturbation moves the economy into regions where nonlinearities are quantitatively important, for example because constraints become binding for many agents, aggregate constraints such as the zero lower bound bind, or the shock induces large reallocations across agents with different exposures. In such cases, the counterfactual path associated with  $X_t + \delta$  is no longer well summarized by the linearized dynamics around the steady state.<sup>5</sup>

(ii) *Macro and macro-finance models.*

Consider a panel VAR for the joint process  $(Y_{it}, W_{it})^\top$  with state-dependent exposure to the aggregate shock and time-invariant propagation:

$$\begin{pmatrix} Y_{it} \\ W_{it} \end{pmatrix} = \alpha_i + \lambda_t + A_i(L) \begin{pmatrix} Y_{i,t-1} \\ W_{i,t-1} \end{pmatrix} + \begin{pmatrix} \lambda_Y(Z_{i,t-1}) \\ \lambda_W(Z_{i,t-1}) \end{pmatrix} X_t + \begin{pmatrix} \eta_{it} \\ \nu_{it} \end{pmatrix}, \quad (5)$$

where  $A_i(L)$  is a lag polynomial with possibly heterogeneous but time-invariant coefficients. The functions  $\lambda_Y(\cdot)$  and  $\lambda_W(\cdot)$  govern exposure of all variables to the aggregate shock and may be constant or arbitrary measurable functions of the predetermined state  $Z_{i,t-1}$ .

Because the system is linear in  $X_t$  and the propagation coefficients are fixed, iterating (5) forward implies that, for each horizon  $h$ ,

$$\mathbb{E}[Y_{i,t+h} \mid \mathcal{F}_{t-1}, X_t] = r_{i,h}(\mathcal{F}_{t-1}) + g_h(Z_{i,t-1}) X_t,$$

for some measurable function  $g_h(\cdot)$ . The function  $g_h(Z_{i,t-1})$  summarizes both the direct effect of  $X_t$  on  $Y_{it}$  (through  $\lambda_Y$ ) and all indirect effects operating through the endogenous post- $t$  path of  $W$  and  $Y$  (through  $\lambda_W$  and the propagation  $A_i(L)$ ).

This structure allows rich state dependence in exposure while ruling out changes in the dynamics governing the evolution of  $(Y, W)$  following the shock. Under this exposure–propagation separation, Assumption 3 holds and state-dependent LPs recover causal impulse responses.

This structure is consistent with leading heterogeneous-firm monetary models. In Ottonello and Winberry (2020), monetary policy shocks shift the intertemporal price system and firms’ financing conditions. In this setting,  $Y_{it}$  is firm investment or capital growth,  $X_t$  is the monetary shock, and  $Z_{i,t-1}$  is a predetermined financial state (e.g., net worth, leverage, or distance-to-default). The model generates strong nonlinearity in exposure because firms’ marginal conditions depend on default risk and exhibit threshold effects (see their Figure 2), so the function  $\lambda(\cdot)$  need not be linear. Importantly, while the financial state evolves endogenously after the shock, the structural forces governing propagation (e.g., depreciation and adjustment costs) remain time-invariant. As a result, the shock shifts the level of  $Y_{i,t+h}$  without altering its propagation dynamics.

---

<sup>5</sup>If  $\mathbb{E}[Y_{i,t+h} \mid \mathcal{F}_{t-1}, X_t]$  is twice differentiable in  $X_t$ , a second-order Taylor expansion implies that the finite-shock response to a perturbation  $\delta$  equals its linear approximation plus a remainder of order  $O(\delta^2)$ , governed by the curvature of the conditional mean over the relevant range.

A closely related implication emerges in Cui et al. (2024). Their mechanism generates nonlinear exposure: the effect of an interest-rate shock on risk-taking depends on a predetermined leverage constraint, and the threshold for adopting risky projects can be hump-shaped in the interest rate. In our notation,  $Y_{it}$  denotes firm-level risk-taking or investment,  $X_t$  the interest-rate shock, and  $Z_{i,t-1}$  the predetermined financial constraint. The nonlinearity arises from equilibrium choice conditions, rather than from changes in the propagation dynamics of  $Y$  following the shock.

### 3.2.2 Inadmissible DGPs.

(i) *Nonlinear dependence on the shock.* Assumption 3 excludes environments in which the conditional mean is nonlinear in the shock, e.g.,

$$\mathbb{E}[Y_{i,t+h} \mid \mathcal{F}_{t-1}, X_t] = f_h(Z_{i,t-1}, X_t) + r_{i,h}(\mathcal{F}_{t-1}).$$

Examples include regime-switching policy, zero-lower-bound nonlinearities, or large discrete interventions. In these cases, the effect of a perturbation depends on the magnitude or sign of  $X_t$ .

(ii) *State-dependent propagation.* Assumption 3 also fails when shocks affect the propagation of the outcome itself. In such environments, shocks alter not only contemporaneous outcomes but also the law of motion governing their future dynamics through their effect on endogenous states, so that the horizon- $h$  response cannot be summarized as a function of the initial state alone. Such mechanisms arise, for example, in sovereign default models such as Almuzara et al. (2025), where aggregate shocks affect borrowing conditions and default risk, which in turn shape future dynamics. In these environments, shocks propagate through endogenous states, so that the law of motion itself depends on the shock.

(iii) *Fully nonlinear HA/HANK models.* The previous mechanisms arise naturally in fully nonlinear heterogeneous-agent environments beyond first-order (log-linear) approximations. In such models, aggregate shocks can move the economy across regions of the state space where the local dynamics differ, for example due to occasionally binding constraints, kinked adjustment costs, or endogenous regime shifts. As a result, shocks affect not only the level of  $Y_{it}$  but also its propagation. Formally, if

$$Y_{i,t+1} = f(Y_{it}, Z_{it}, X_t),$$

and  $\partial f / \partial Y$  depends on states that are themselves affected by the shock, then the propagation of  $Y$  becomes shock-dependent. In that case, the horizon- $h$  response depends on the entire nonlinear state path induced by  $X_t$ , and the conditional mean generally cannot be written in the multiplicative form of Assumption 3.

Such violations of Assumption 3 are economically plausible in aggregate time-series settings, where shocks can alter persistence. The critique of Gonçalves et al. (2024b) that state-dependent LPs do not recover finite- $\delta$  impulse responses is therefore valid in these settings. Our condition targets a different empirical setting: micro-macro panels in which cross-sectional heterogeneity



operates through state-dependent exposure to common aggregate shocks that enter linearly, while the propagation of individual outcomes is governed by time-invariant structural primitives (e.g., depreciation, adjustment costs, or idiosyncratic risk processes). In that environment, Assumption 3 isolates state dependence in exposure from propagation and restores a transparent causal interpretation of state-dependent LPs (provided the state dependence is estimated nonparametrically, which we discuss next).

## 4 Estimation and Inference

This section develops a sieve estimator of the state-dependent impulse response function  $g_h(\cdot)$  and associated inference procedures. The parameter of interest is the coefficient function  $g_h(z)$  defined by the conditional mean restriction established in Section 3. Section 3 shows that this function is the causal object of interest and that, in general, recovering it requires nonparametric estimation.

The resulting LP estimation problem has a structure in which the aggregate shock enters with a functional coefficient  $g_h(Z_{i,t-1})$ , while additional regressors (such as predetermined controls) enter as finite-dimensional components.

Our setting builds on standard econometric tools - sieve approximation, HAC covariance estimation, and multiplier bootstrap methods for uniform inference - but applies them in a LP framework with features that are not covered by off-the-shelf results.

On the one hand, the LP structure with an exogenous aggregate shock is particularly convenient: because  $X_t$  is a martingale difference, the nuisance component of the conditional mean is orthogonal to the regressor of interest. As a result,  $g_h(\cdot)$  can be identified without specifying the rest of the conditional mean, unlike in standard regression settings.

On the other hand, the LP environment introduces several nonstandard features that require adapting existing sieve methods. First, the object of interest is a *functional coefficient* multiplying an aggregate shock, rather than a generic conditional mean. Second, overlapping outcomes across horizons induce time-series dependence in the regression errors. Third, the aggregate shock generates cross-sectional dependence across units. Finally, inference targets the entire function  $g_h(\cdot)$ , rather than a finite-dimensional parameter, making uniform inference essential.

Taken together, these features imply that while our approach relies on standard building blocks, their combination in a macro-micro LP setting is nonstandard and requires careful adaptation and verification of the underlying asymptotic arguments.

Relative to the general approach to nonlinear impulse responses considered by Gonçalves et al. (2024b), we restrict attention to the class of environments characterized in Section 3, in which the shock enters the conditional mean linearly and heterogeneity operates through a state-dependent functional coefficient. Their approach recovers impulse responses by first estimating the response of the outcome as a function of the shock, allowing for general nonlinear dependence, and then constructing impulse responses as differences in this response function evaluated at different shock realizations and averaged over the distribution of the shock. While differencing removes nuisance

components in population, the procedure involves estimating a nonparametric response function and then transforming it, so that estimation error from the first step carries over into the second-step construction of the impulse response. By contrast, we exploit the linear-in-shock structure implied by our identifying restriction to eliminate the nuisance component at the level of the identifying moment and estimate the coefficient function directly, without requiring a second-step transformation. The two approaches also differ in their assumptions on the shock process: their construction relies on i.i.d. shocks to justify averaging over the shock distribution, whereas our approach only requires martingale-difference exogeneity.

The remainder of this section formalizes the estimator and inference procedures. Full technical details and proofs are collected in Appendix B and Appendix C.

#### 4.1 Sieve Estimator

For each horizon  $h \in \{0, 1, \dots, H\}$ , we estimate  $g_h(\cdot)$  using a sieve approximation. Let  $\{\phi_{j,J}(z)\}_{j=1}^J$  denote a basis of cubic B-splines defined on the support of  $Z_{i,t-1}$ , where interior knots are located at equally spaced empirical quantiles and with a constant term included. For a given sieve dimension  $J$ , approximate

$$g_h(z) \approx \phi_J(z)^\top b_h, \quad \phi_J(z) = (\phi_{1,J}(z), \dots, \phi_{J,J}(z))^\top. \quad (6)$$

Substituting (6) into the LP yields the estimation equation

$$Y_{i,t+h} = \sum_{j=1}^J b_{h,j} [\phi_{j,J}(Z_{i,t-1}) X_t] + W_{i,t-1}^\top \gamma_h + u_{i,t+h}. \quad (7)$$

Let  $D_{it}(J)$  denote the vector of regressors in (7). The OLS estimator is

$$\hat{\theta}_h(J) = \left( \sum_{i=1}^N \sum_{t=1}^{T-h} D_{it}(J) D_{it}(J)^\top \right)^{-1} \left( \sum_{i=1}^N \sum_{t=1}^{T-h} D_{it}(J) Y_{i,t+h} \right), \quad (8)$$

with  $\hat{b}_h(J)$  denoting the subvector corresponding to the sieve coefficients. The associated estimator of the impulse response function is

$$\hat{g}_{h,J}(z) = \phi_J(z)^\top \hat{b}_h(J). \quad (9)$$

We select the sieve dimension  $J$  using the Akaike information criterion (additional selection criteria are examined in the simulation section and in the Appendix and perform similarly):

$$\text{AIC}_h(J) = N(T-h) \log \left( \frac{1}{N(T-h)} \sum_{i=1}^N \sum_{t=1}^{T-h} \hat{u}_{i,t+h}(J)^2 \right) + 2K_J, \quad (10)$$

where  $\hat{u}_{i,t+h}(J)$  denote the OLS residuals from (7) and  $K_J = J + \dim(W_{i,t-1})$  is the total number

of estimated parameters. The selected number of sieves is

$$\hat{J}_h = \arg \min_{J \in \mathcal{J}} \text{AIC}_h(J). \quad (11)$$

In practice, we take the candidate set  $\mathcal{J} = \{4, \dots, 20\}$ , with the upper bound serving as a finite-sample safeguard against an overly rich basis. Here  $J = 4$  gives a global cubic polynomial, while larger values of  $J$  add interior knots to the cubic B-spline basis and allow for more flexible nonlinearities. The final estimator is

$$\hat{g}_h(z) = \phi_{\hat{J}_h}(z)^\top \hat{b}_h(\hat{J}_h). \quad (12)$$

The following algorithm summarizes the steps involved in estimating the impulse response function.

**Algorithm 1 (Sieve LP estimator of the impulse response function).** For each horizon  $h = 0, \dots, H$ :

1. Construct cubic B-spline basis functions for  $Z_{i,t-1}$  with a constant term included and interior knots equally placed at empirical quantiles.
2. For each  $J \in \mathcal{J}$ , form the  $J$  interaction regressors  $\phi_{j,J}(Z_{i,t-1}) X_t$ ,  $j = 1, \dots, J$ , and stack them with  $W_{i,t-1}$  to form  $D_{it}(J)$ .
3. Estimate (7) by OLS and compute residuals  $\hat{u}_{i,t+h}(J)$ .
4. Compute  $\text{AIC}_h(J)$  for each  $J \in \mathcal{J}$  and select  $\hat{J}_h$  via (11).
5. Set  $\hat{g}_h(z) = \phi_{\hat{J}_h}(z)^\top \hat{b}_h(\hat{J}_h)$ .

## 4.2 Inference

We consider both pointwise and uniform inference for the impulse response function  $g_h(\cdot)$ .

Uniform inference is important because the empirical object is often the entire state-dependent response function, rather than the response at a single preselected state. Economic conclusions frequently depend on features of the whole curve - such as whether responses are monotone, hump-shaped, concentrated in the tails, or significantly different across regions of the state distribution. Pointwise intervals can give a misleading impression when researchers inspect many values of the state simultaneously. By contrast, uniform confidence bands allow for valid statements about the shape of the response function over the relevant support, while controlling coverage jointly across states. In the types of applications considered here, this is particularly important for distinguishing a genuinely nonlinear transmission pattern from sampling variation around a linear approximation.

Let  $D_{it}$ ,  $\hat{b}_h$ ,  $\hat{u}_{i,t+h}$  denote, respectively, the vector of regressors, the sieve coefficient estimator, OLS residual in (7) corresponding to the AIC-selected number of basis functions  $\hat{J}_h$ . Under

regularity conditions, the estimator admits the linear representation

$$\sqrt{NT}(\hat{b}_h - b_h) = \frac{1}{\sqrt{NT}} \sum_{t=1}^{T-h} \sum_{i=1}^N \psi_{h,i,t} + o_p(1), \quad (13)$$

for a suitable influence function  $\psi_{h,i,t}$  that reflects the projection of the sieve regressors on the control variables. The precise form of  $\psi_{h,i,t}$  and the regularity conditions are given in Appendices B and C.

Define the sample second moment matrix

$$\frac{1}{N(T-h)} \sum_{i=1}^N \sum_{t=1}^{T-h} D_{it} D_{it}^\top = \begin{pmatrix} \hat{A}_{11,h} & \hat{A}_{12,h} \\ \hat{A}_{21,h} & \hat{A}_{22,h} \end{pmatrix},$$

where the first block corresponds to  $X_t \phi_{\hat{J}_h}(Z_{i,t-1})$  and the second to  $W_{i,t-1}$ . Form the partialled-out sieve regressor

$$\tilde{P}_{h,i,t} = X_t \phi_{\hat{J}_h}(Z_{i,t-1}) - \hat{A}_{12,h} \hat{A}_{22,h}^{-1} W_{i,t-1}, \quad (14)$$

and the score process

$$\hat{s}_{h,t} = \sum_{i=1}^N \tilde{P}_{h,i,t} \hat{u}_{i,t+h}. \quad (15)$$

We estimate the long-run covariance matrix of the score process by

$$\hat{\Omega}_h = \hat{\Gamma}_h(0) + \sum_{k=1}^L w_k \left\{ \hat{\Gamma}_h(k) + \hat{\Gamma}_h(k)^\top \right\}, \quad \hat{\Gamma}_h(k) = \frac{1}{N(T-h)} \sum_{t=k+1}^{T-h} \hat{s}_{h,t} \hat{s}_{h,t-k}^\top, \quad (16)$$

where  $w_k = 1 - k/(L+1)$  and  $L = \lfloor 4[N(T-h)/100]^{2/9} \rfloor$ . The covariance matrix of  $\hat{b}_h$  is estimated by

$$\hat{V}_h = \frac{1}{N(T-h)} \hat{A}_{11,h}^{-1} \hat{\Omega}_h \hat{A}_{11,h}^{-1\top}, \quad (17)$$

where

$$\hat{\hat{A}}_{11,h} = \hat{A}_{11,h} - \hat{A}_{12,h} \hat{A}_{22,h}^{-1} \hat{A}_{21,h}.$$

The pointwise variance of  $\hat{g}_h(z)$  is

$$\hat{\sigma}_h^2(z) = \phi_{\hat{J}_h}(z)^\top \hat{V}_h \phi_{\hat{J}_h}(z). \quad (18)$$

Thus, for fixed  $z$  on the empirical support of  $Z_{i,t-1}$ , the  $(1 - \alpha)$  pointwise confidence interval is

$$\mathcal{I}_h(z) = [\hat{g}_h(z) \pm z_{1-\alpha/2} \hat{\sigma}_h(z)], \quad (19)$$

where  $z_{1-\alpha/2}$  is the standard normal critical value.

For uniform inference, let  $\mathcal{Z}_G$  be a fine grid covering the empirical support of  $Z_{i,t-1}$ . For each

bootstrap draw  $b = 1, \dots, B$ , generate

$$\xi_h^{(b)} \sim \mathcal{N}(0, I_{\hat{J}_h}),$$

and form the Gaussian process

$$G_h^{(b)}(z_m) = \phi_{\hat{J}_h}(z_m)^\top \hat{V}_h^{1/2} \xi_h^{(b)}, \quad z_m \in \mathcal{Z}_G. \quad (20)$$

Compute the supremum statistic

$$T_h^{(b)} = \max_{z_m \in \mathcal{Z}_G} |G_h^{(b)}(z_m)|, \quad (21)$$

and let  $c_{h,1-\alpha}$  be the empirical  $(1-\alpha)$  quantile of  $\{T_h^{(b)}\}_{b=1}^B$ . The  $(1-\alpha)$  uniform confidence band is

$$\mathcal{C}_h(z_m) = [\hat{g}_h(z_m) \pm c_{h,1-\alpha}], \quad z_m \in \mathcal{Z}_G. \quad (22)$$

The distinction is that  $\mathcal{I}_h(z)$  provides coverage at a fixed state  $z$ , while  $\mathcal{C}_h(\cdot)$  provides simultaneous coverage over all grid points in  $\mathcal{Z}_G$ .

The following algorithms summarize the construction of pointwise confidence intervals and uniform confidence bands.<sup>6</sup>

**Algorithm 2 (Pointwise confidence intervals).** For each horizon  $h = 0, \dots, H$ :

1. Estimate  $\hat{g}_h(z)$  in (12) and obtain residuals  $\hat{u}_{i,t+h}$ .
2. Form  $\tilde{P}_{h,i,t}$  in (14), use them to compute the score process  $\{\hat{s}_{h,t}\}_{t=1}^{T-h}$  in (15), and then compute the long-run covariance estimator  $\hat{\Omega}_h$  in (16).
3. Compute the covariance estimator  $\hat{V}_h$  in (17).
4. For each  $z_m \in \mathcal{Z}_G$ , compute  $\hat{\sigma}_h^2(z_m)$  in (18) and construct  $\mathcal{I}_h(z_m)$  in (19).

**Algorithm 3 (Uniform confidence bands).** For each horizon  $h = 0, \dots, H$ :

1. Starting from Step 3 of Algorithm 2, take  $\hat{g}_h(z)$  and  $\hat{V}_h$  as given.
2. For  $b = 1, \dots, B$ , draw  $\xi_h^{(b)} \sim \mathcal{N}(0, I_{\hat{J}_h})$ .
3. For each  $z_m \in \mathcal{Z}_G$ , compute  $G_h^{(b)}(z_m)$  in (20).
4. Compute  $T_h^{(b)}$  in (21) and find the empirical  $(1-\alpha)$  quantile of  $\{T_h^{(b)}\}_{b=1}^B$ .
5. Construct  $\mathcal{C}_h(z_m)$  in (22) for  $z_m \in \mathcal{Z}_G$ .

---

<sup>6</sup>For cumulative specifications where the outcome is  $Y_{i,t+h} - Y_{i,t}$ , the corresponding estimand is  $g_h(z) - g_0(z)$ . Inference proceeds analogously by estimating the LP with the transformed outcome.

### 4.3 Asymptotic Properties

We summarize the main theoretical result; formal conditions and proofs are provided in Appendix C.

**Theorem 1.** *Under regularity conditions, the estimator  $\hat{g}_h(\cdot)$  is uniformly consistent for  $g_h(\cdot)$  over compact subsets of the support of  $Z_{i,t-1}$ . For each fixed  $z$ ,  $\hat{g}_h(z)$  is asymptotically normal. Moreover, Algorithm 3 delivers asymptotically valid uniform confidence bands for  $g_h(\cdot)$ .*

## 5 Simulation Study

All simulations are based on a DGP satisfying Assumption 3. We run 300 Monte Carlo replications of data generated as

$$Y_{it} = g(Z_{i,t-1})X_t + 0.8Y_{i,t-1} + \varepsilon_{it}, \quad i = 1, \dots, N; \quad t = 1, \dots, T, \quad (23)$$

where  $X_t \sim \mathcal{N}(0, 1)$  and  $\varepsilon_{it} \sim \mathcal{N}(0, 1)$ . The state variable evolves as

$$Z_{it} = \mu_i + \xi_{it}, \quad \mu_i \sim \mathcal{N}(0, 3), \quad \xi_{it} = 0.8\xi_{i,t-1} + \sqrt{1 - 0.8^2}v_{it}, \quad v_{it} \sim \mathcal{N}(0, 1).$$

We discard the first 500 observations and retain  $T$ . Results are robust to including fixed effects (not reported).

We focus on the cubic specification

$$g(z) = 0.5z + 0.3z^2 - 0.25z^3. \quad (24)$$

The corresponding horizon- $h$  causal impulse response is

$$\text{IRF}_Y^{\text{true}}(h, \delta \mid Z_{i,t-1} = z) = 0.8^h g(z) \delta. \quad (25)$$

We estimate the nonparametric state-dependent LP

$$Y_{i,t+h} = g_h(Z_{i,t-1})X_t + \gamma_h Y_{i,t-1} + u_{i,t+h}, \quad h = 0, \dots, H,$$

using the spline sieve estimator described in Section 4. The estimated impulse response is

$$\text{IRF}_Y^{\text{sieve}}(h, \delta \mid Z_{i,t-1} = z) = \hat{g}_h(z) \delta. \quad (26)$$

We select the sieve dimension  $J$  using AIC, GCV, and LASSO, alongside an oracle choice ( $J = 4$ ). For AIC and GCV, we search over  $\mathcal{J} = \{4, \dots, 20\}$ , with GCV given by

$$\text{GCV}_h(J) = \frac{\sum_{i,t} \hat{u}_{i,t+h}(J)^2}{N(T-h) \left(1 - \frac{K_L}{N(T-h)}\right)^2}, \quad \hat{J}_h = \arg \min_{J \in \mathcal{J}} \text{GCV}_h(J).$$

For LASSO, we set a maximal dimension  $J_{\text{lasso}} = 50$  and estimate

$$\hat{\theta}_h = \arg \min_{\theta_h} \left\{ \sum_{i,t} (Y_{i,t+h} - D_{it}(J_{\text{lasso}})^\top \theta_h)^2 + \lambda_h \|\theta_h\|_1 \right\},$$

where  $\lambda_h$  is chosen by cross-validation. The selected dimension  $\hat{J}_h$  is the number of nonzero sieve coefficients.

### 5.1 Sieve Estimation: Gains over Linear LPs

We evaluate the performance of the sieve estimator for the state-dependent impulse response and its gains relative to standard linear state-dependent LPs. Under the DGP, which satisfies Assumption 3, the causal impulse response is well defined as  $g_h(z)$ , so the simulation isolates the role of functional-form restrictions. We set  $N = 500$ ,  $T = 200$ , and select the number of basis functions using AIC.

For comparison, we estimate the linear state-dependent LP

$$Y_{i,t+h} = (\alpha_h + \beta_h Z_{i,t-1})X_t + \gamma_h Y_{i,t-1} + u_{i,t+h},$$

which implies

$$\text{IRF}_Y^{\text{linear}}(h, \delta \mid Z_{i,t-1} = z) = (\hat{\alpha}_h + \hat{\beta}_h z)\delta. \quad (27)$$

Figure 1 reports Monte Carlo averages and 10% to 90% percentile bands for the sieve and linear IRFs, together with the true response, for a unit shock ( $\delta = 1$ ) over a grid of 500 points in  $[-4.65, 4.65]$ .

Figure 1 highlights the main gain from sieve estimation relative to the linear state-dependent LP. When the true response varies nonlinearly with the state, the linear specification yields a distorted profile with substantial bias, particularly in the tails. By contrast, the spline sieve closely tracks the true IRF across the entire range of  $z$ , with tight Monte Carlo dispersion. The two estimators can also imply opposite signs in some regions, leading to qualitatively different economic conclusions.

As the horizon increases and the response becomes more linear, the discrepancy between the two approaches attenuates, although the sieve estimator continues to recover the shape more accurately. This reflects both improved approximation and the fact that the sieve targets the causal state-dependent IRF, whereas the linear specification recovers a projection that generally differs from the causal response.

These results are not driven by the choice of basis. Appendix D shows that the same conclusions hold when the true  $g(z)$  is generated by a Fourier series but estimated using splines (Figure 7).

We next quantify the gains of nonparametric estimation using the root integrated mean squared error (RIMSE). Table 2 compares the sieve IRF to the linear IRF for different horizons  $h$  and impulse sizes  $\delta \in \{0.5\sigma_X, \sigma_X, 2\sigma_X\}$ , where  $\sigma_X$  is the standard deviation of the shock  $X_t$  (here equal to 1).

The results reinforce the visual evidence. The sieve estimator delivers substantially lower

Cubic IRFs across horizons with sieve spline AIC and linear interaction LPs ( $N = 500$ ,  $T = 200$ ,  $\delta = 1$ )

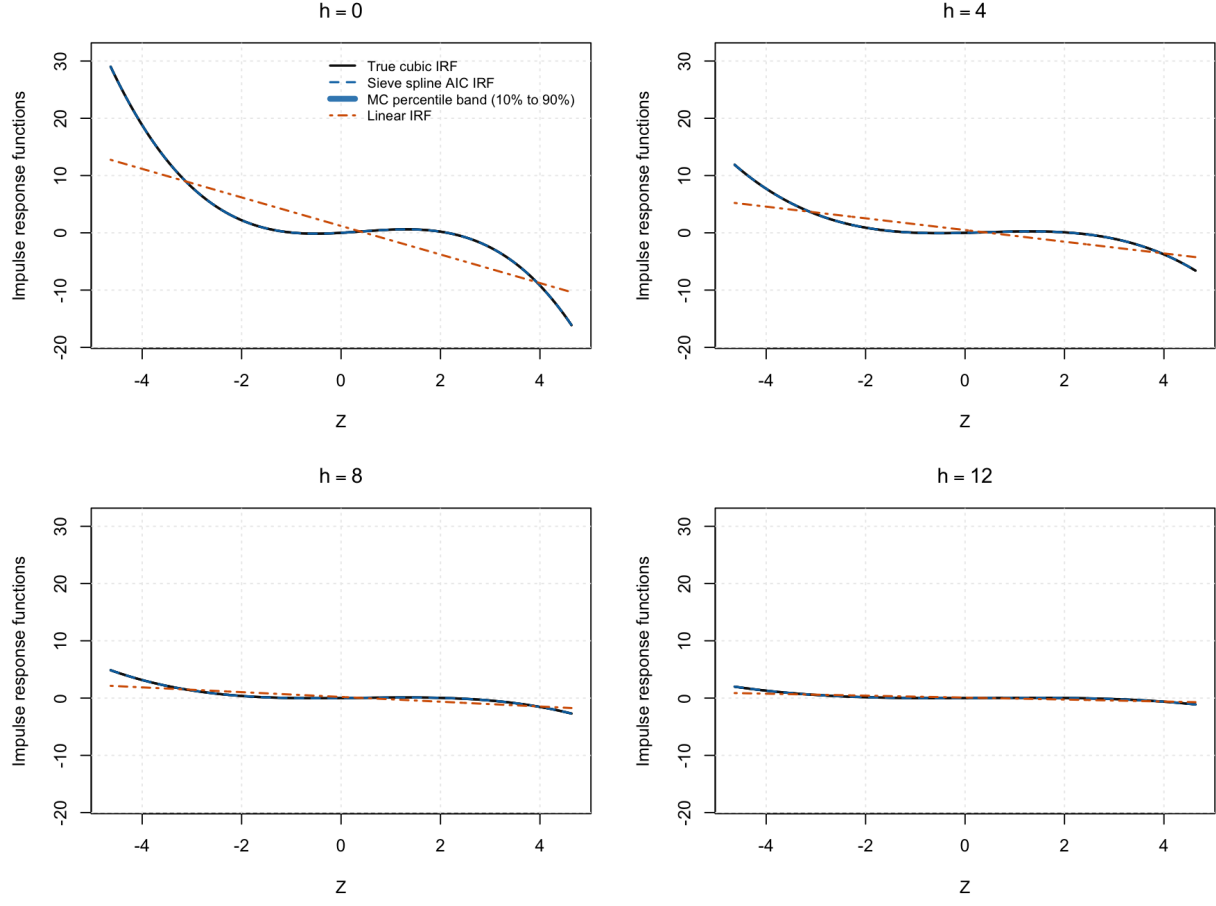


Figure 1: Monte Carlo comparison of sieve and linear IRFs under the cubic DGP

Table 2: RIMSE comparison of sieve and linear IRFs

| $\delta$ | $0.5\sigma_X$ |        | $\sigma_X$ |        | $2\sigma_X$ |        |
|----------|---------------|--------|------------|--------|-------------|--------|
| $h$      | Sieve         | Linear | Sieve      | Linear | Sieve       | Linear |
| 0        | 0.03          | 20.04  | 0.05       | 40.08  | 0.10        | 80.15  |
| 4        | 0.04          | 8.21   | 0.08       | 16.42  | 0.17        | 32.83  |
| 8        | 0.05          | 3.36   | 0.09       | 6.73   | 0.18        | 13.45  |
| 12       | 0.04          | 1.38   | 0.09       | 2.75   | 0.17        | 5.51   |



RIMSE than the linear specification. The gains are largest at short horizons, where nonlinearities are most pronounced, and remain meaningful even at longer horizons despite the impulse response becoming almost linear. Importantly, the gains also widen with  $\delta$ , indicating that the cost of misspecifying nonlinear state dependence becomes more pronounced for larger shocks. Overall, the findings confirm that accounting for nonlinearities is important for accurately recovering the impulse response function.

## 5.2 Sieve Inference: Uniform-band Performance and Controlling for Intermediate Shocks

We evaluate the finite-sample performance of sieve-based uniform confidence bands across sample sizes  $(N, T) \in \{100, 500\} \times \{40, 120, 200\}$  and horizons  $h \in \{0, 4, 8, 12\}$ .

For the spline sieve, we consider an oracle specification ( $J = 4$ ) and data-driven choices based on AIC, GCV, and LASSO. The main text focuses on AIC; results for all selectors are reported in the appendix.

Given  $J$ , we construct uniform confidence bands for  $g_h(z)\delta$  (with  $\delta = 1$ ) using Algorithm 3 with  $B = 2000$  bootstrap draws and nominal coverage  $1 - \alpha = 0.95$ . Performance is evaluated over a grid of 500 points on  $[\min\{Z_{it}\}, \max\{Z_{it}\}]$  using coverage (fraction of replications in which the true IRF lies within the band over the grid) and average band width.

Before turning to overall band performance, we examine the role of including intermediate shocks in the LP. For  $h \geq 1$ ,  $Y_{i,t+h}$  depends on future interaction terms  $g_j(Z_{i,t+j-1})X_{t+j}$  for  $j = 1, \dots, h$ .

We compare specifications that include or omit these terms, approximating each  $g_j(\cdot)$  using the same sieve dimension. Under exogeneity, the intermediate terms are orthogonal to  $X_t$ , so their omission does not affect identification, but including them improves efficiency.<sup>7</sup>

Table 3: Uniform-band performance under specifications with and without intermediate shocks for the sieve with AIC selection at nominal coverage  $1 - \alpha = 0.95$

| $h$ | Intermediate Shocks |       | No Intermediate Shocks |       |
|-----|---------------------|-------|------------------------|-------|
|     | Coverage            | Width | Coverage               | Width |
| 0   | 0.84                | 0.47  | 0.84                   | 0.47  |
| 4   | 0.83                | 0.78  | 0.81                   | 95.04 |
| 8   | 0.81                | 0.77  | 0.84                   | 79.77 |
| 12  | 0.80                | 0.79  | 0.85                   | 76.21 |

Table 3 shows that the two specifications coincide at  $h = 0$ , where no intermediate shocks are present. For  $h \geq 4$ , coverage is similar, but including intermediate shocks yields substantially tighter bands, indicating a sizable efficiency gain. We therefore include intermediate shocks in all subsequent simulations and in the empirical application.

<sup>7</sup>Orthogonality follows from the multiplicative structure and the exogeneity of future shocks:  $X_{t+j}$  is mean independent of past information, including  $X_t$ , while  $g_j(Z_{i,t+j-1})$  is measurable with respect to that information.

Table 4 reports results for  $N \in \{100, 500\}$  and  $T \in \{40, 120, 200\}$  at horizons  $h \in \{0, 4\}$  under AIC selection; additional results are reported in Appendix D. Performance improves with  $T$ , as coverage increases and bands narrow, while differences across  $N$  are limited. At longer horizons, coverage declines modestly and bands widen.

Table 4: Uniform-band performance of the sieve with AIC selection at nominal coverage  $1 - \alpha = 0.95$

|                                            | $N = 100$ |       | $N = 500$ |       |
|--------------------------------------------|-----------|-------|-----------|-------|
|                                            | Coverage  | Width | Coverage  | Width |
| <i>Panel A: Horizon <math>h = 0</math></i> |           |       |           |       |
| $T = 40$                                   | 0.73      | 1.02  | 0.76      | 0.67  |
| $T = 120$                                  | 0.84      | 0.84  | 0.83      | 0.53  |
| $T = 200$                                  | 0.87      | 0.69  | 0.84      | 0.47  |
| <i>Panel B: Horizon <math>h = 4</math></i> |           |       |           |       |
| $T = 40$                                   | 0.74      | 1.70  | 0.72      | 1.12  |
| $T = 120$                                  | 0.82      | 1.34  | 0.79      | 0.83  |
| $T = 200$                                  | 0.83      | 1.16  | 0.83      | 0.78  |

Overall, the AIC-based bands exhibit some undercoverage relative to the nominal level. Appendix D shows that coverage is close to nominal for the oracle selector, implying that the gap is driven by smoothing-parameter selection rather than by the band construction itself. This is consistent with the nonparametric literature, where data-driven tuning introduces additional variability that can lead to undercoverage in finite samples, especially for uniform inference.

### 5.3 Sieve Robustness: Sensitivity to Selector Choice

We assess the sensitivity of sieve estimation to the choice of selector for the number of basis functions at horizon  $h = 4$ . Estimation results are qualitatively similar across selectors and are therefore not reported. Table 5 reports the corresponding inference results. Oracle, AIC, and GCV yield similar coverage with relatively tight bands, while LASSO achieves slightly higher coverage at the cost of substantially wider bands. Overall, sieve estimation is robust to selector choice, although different selectors imply a coverage–width trade-off.

Table 5: Uniform-band performance across selector choices at horizon  $h = 4$  under nominal coverage  $1 - \alpha = 0.95$

| Selector | Coverage | Width |
|----------|----------|-------|
| Oracle   | 0.92     | 0.55  |
| AIC      | 0.83     | 0.78  |
| GCV      | 0.83     | 0.78  |
| LASSO    | 0.85     | 2.70  |

## 6 Empirical Application

To demonstrate the empirical relevance of our nonparametric state-dependent LP, we apply our methodology to study the effect of monetary policy shocks on firm investment in the spirit of Ottonello and Winberry (2020), Cloyne et al. (2023) and Jeenas (2023), among others. Specifically, we study how the investment response to these macroeconomic shocks varies across heterogeneous firms, with firm-level financial conditions serving as the state variable. This setting fits our framework to the extent that monetary policy shocks operate through heterogeneous balance-sheet exposure, while propagation dynamics are well approximated as stable over the relevant horizon.

Following Ottonello and Winberry (2020) (OW), we use distance to default (D2D) – an estimate of the probability of default using leverage ratios and equity price volatility – as our primary measure of firm financial conditions.<sup>8</sup> Firms with high D2D are at lower risk of default and hence safer, i.e., more financially sound. Firms with low D2D are the opposite. For ease of comparability, we use the same monetary policy shocks as OW, identified from high-frequency changes in Federal Funds rate futures in a narrow window around FOMC announcements. As in OW, we measure firm investment using Compustat quarterly data as the change in the log book value of the firm’s capital stock, denoted  $k_{i,t} \equiv \log K_{i,t}$ , and compute D2D following the approach in Gilchrist and Zakrajšek (2012). Throughout, we standardize D2D so that the units are in standard deviations relative to the sample mean, as do OW. We follow the same sample selection criteria as in OW.

We estimate the following LP specification:

$$k_{i,t+h} - k_{i,t-1} = \alpha_{sth} + g_h(Z_{i,t-1})X_t + \rho\Delta k_{i,t-1} + \Gamma_h W_{i,t-1} + u_{i,t+h}. \quad (28)$$

Here, the dependent variable measures the firm’s cumulative (net) investment from quarter  $t$  through  $t+h$ ;  $\alpha_{sth}$  denotes sector-time-horizon fixed-effects;  $g_h(Z_{i,t-1})$  is our nonlinear function of the firm’s financial condition (D2D);  $X_t$  is the monetary policy shock;  $\Delta k_{i,t-1}$  is the firm’s lagged one-period investment rate;  $W_{i,t-1}$  is a set of firm-level controls; and  $u_{i,t+h}$  is an error term.<sup>9</sup>

We estimate two versions of (28) that differ only in the specification of the  $g(\cdot)$  function. First, we impose a linear function such that  $g_h(Z_{i,t-1}) = \beta_h Z_{i,t-1}$ . Second, we implement the flexible nonlinear approach developed in Section 4. We display our results in Figure 2, which plots the

---

<sup>8</sup>D2D is equal to the (log) ratio of firm value to debt (an inverse measure of leverage), appropriately adjusted for the volatility of firm value. It can be interpreted as the number of standard deviations the (log) ratio of value to debt must fall below its mean for the firm to default on its debt. Although OW also study leverage, we focus on D2D for two main reasons: first, there is a large empirical literature in corporate finance that documents the performance of D2D as an indicator of the likelihood that a firm will declare bankruptcy and/or default on a bond and the key economic importance of D2D in determining default intensities (Duffie et al., 2007; Schaefer and Strebulaev, 2008; Duffie, 2011; Atkeson et al., 2017); second, OW find weaker results using leverage, perhaps because it is a less precise indicator of firms’ financial soundness.

<sup>9</sup>As in OW, we include as controls  $Z_{i,t-1}$ , total assets, sales growth, current assets as a share of total assets, a fiscal quarter indicator, and the interaction of  $Z_{i,t-1}$  with lagged GDP growth. Our specification differs from OW in three ways: first, we control for the firm’s lagged investment rate; second, they use the firm’s demeaned financial position relative to its average level; third, they include firm fixed effects. None of these differences qualitatively changes our conclusions. Finally, we include time fixed effects only to facilitate comparison with OW, while noting that their inclusion may raise additional identification issues that are beyond the scope of this paper. .

impulse response function – the investment response to the monetary policy shock – on impact and at horizons of four, eight and 12 quarters following the shock. Each panel in the figure displays the investment response across the entire distribution of firms as a function of their pre-shock financial condition, i.e.,  $g_h(Z_{i,t-1})$ , accompanied by uniform confidence bands around the nonlinear estimate constructed using Algorithms 1 and 3. The shock is normalized as a 25 basis point surprise interest rate cut.

## 6.1 Nonlinear response heterogeneity

Panel A of Figure 2 sharply illustrates our main findings. Imposing a linear specification yields a positive impact coefficient, which implies a monotonic upward-sloping relationship between a firm’s investment response and financial condition – less risky, more financially sound firms with higher D2D are more responsive to monetary policy shocks. The result corroborates the influential findings in OW, who also estimate a positive coefficient. In contrast, the nonlinear IRF shows quite a different pattern: a non-monotonic hump-shaped response that peaks at approximately the mean of the distribution, rather than at the least risky right tail. Thus, although responsiveness is increasing for firms that are close to default, as found in previous work, the relationship turns negative around the mean, implying that there is a large portion of the distribution where riskier, lower distance to default firms are more responsive to monetary policy shocks, not less.<sup>10</sup>

Although our econometric model does not reveal the precise economic forces driving these patterns, one interpretation is that firms close to default are investment constrained and respond little, firms near the middle of the distribution are most sensitive to interest rates, and the safest firms less so. Thus, our method uncovers a richer, more nuanced relationship between shock sensitivity and financial conditions that is masked by the linear specification, underscoring the importance of allowing for more general state-dependence when analyzing microeconomic responses to macroeconomic shocks.<sup>11</sup>

Panels B–D of Figure 2 report IRFs at four-quarter intervals up to 12 quarters after the shock. The non-monotonicity becomes more pronounced over time: firms near the center of the D2D distribution exhibit increasingly larger cumulative investment responses than those at either tail.<sup>12</sup> In contrast, the linear specification misses these dynamics and suggests that the response of high-D2D firms continues to grow relative to that of firms closer to default.

To illustrate the economic implications of response heterogeneity, Figure 3 plots the differential in cumulative impulse responses across selected points of the D2D distribution under the linear and nonlinear specifications, which directly captures how the transmission of shocks varies across firms over the post-shock horizon. We report pointwise confidence intervals using Algorithm 2,

<sup>10</sup>Due to right-skewness in the distribution, the median D2D is slightly below the mean.

<sup>11</sup>Our results are qualitatively unchanged under various modifications to our main specification; for example, including firm fixed-effects somewhat dampens responsiveness across the distribution, but the hump-shaped pattern we uncover remains.

<sup>12</sup>Although not shown, the cumulative IRF changes little beyond 12 quarters, indicating that the effects of the shock have largely dissipated.

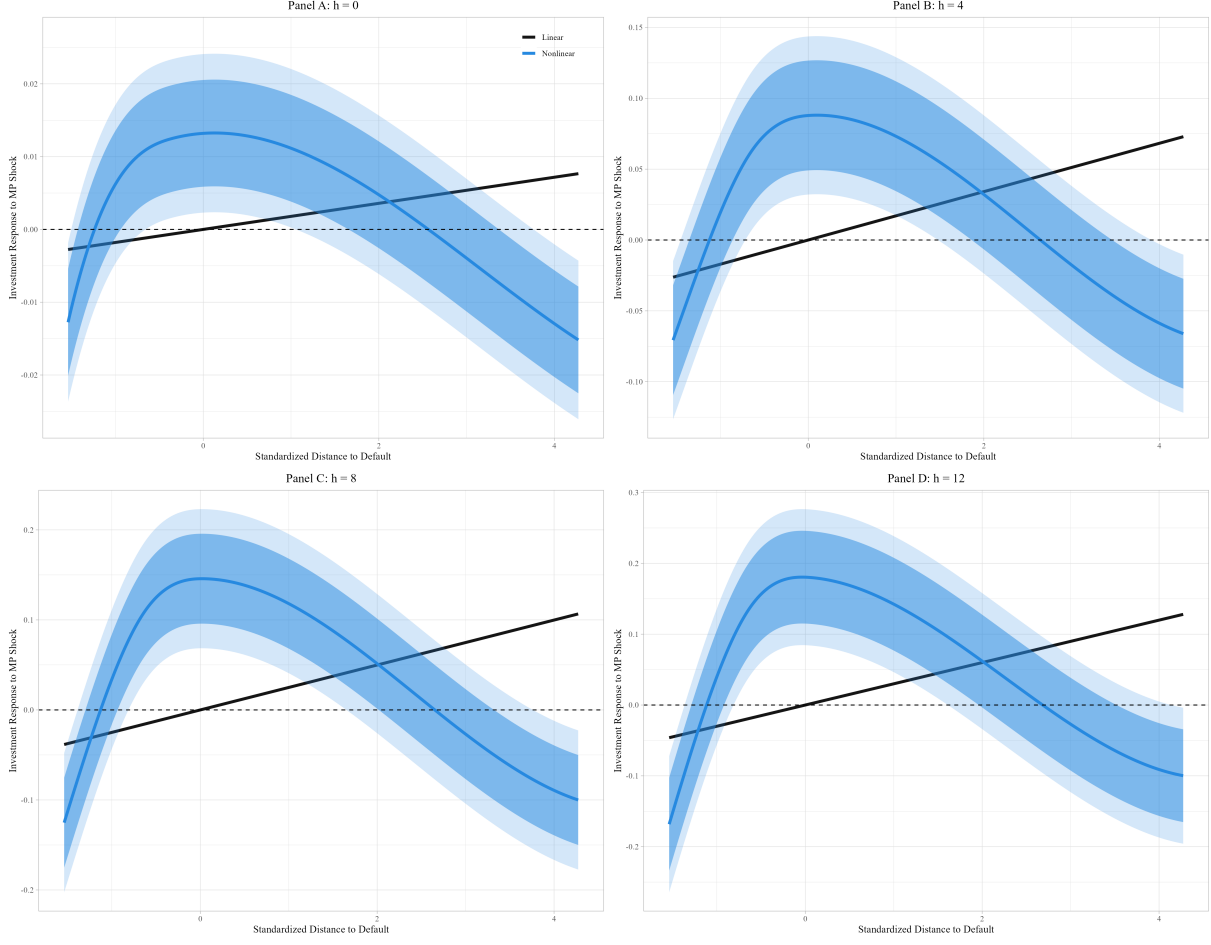


Figure 2: Impulse Response of Firm-Level Investment to a Monetary Policy Shock

*Notes:* This figure displays local projection impulse response functions to a 25 b.p. monetary policy shock across firms with different levels of distance to default. The units of distance to default are standard deviations relative to the sample mean. The linear specification imposes a linear interaction between the shock and distance to default. The nonlinear specification implements the nonlinear function of distance to default. Dark (light) shaded regions report 68% (90%) uniform confidence bands.

since the objects of interest are differences in impulse responses at specific states.<sup>13</sup> In Panel A, we compare a firm at the 50th percentile of the D2D distribution to one at the 5th percentile. Although both sets of estimates predict that the firm at the 50th percentile will invest more than the firm at the 5th percentile in response to the monetary policy shock, the linear estimates significantly underpredict the difference; for example, after 16 quarters, the linear estimates yield a cumulative investment differential of just under 5%, whereas the nonlinear estimates suggest a cumulative differential exceeding 20%. In Panel B we similarly compare a firm at the 95th percentile of the state distribution to the 50th; in this case, the two sets of estimates yield virtually opposite

<sup>13</sup>Specifically, we consider  $\hat{g}_h(Z_a) - \hat{g}_h(Z_b)$ , where  $Z_a$  and  $Z_b$  denote two fixed points of the state distribution, such as the 50th and 5th percentiles, or the 95th and 50th percentiles, of  $Z_{i,t-1}$ . Accordingly, instead of using the pointwise variance estimator in (18) directly, we construct pointwise confidence intervals using the variance  $(\phi_{\hat{J}_h}(Z_a) - \phi_{\hat{J}_h}(Z_b))^\top \hat{V}_h(\phi_{\hat{J}_h}(Z_a) - \phi_{\hat{J}_h}(Z_b))$  in Algorithm 2.

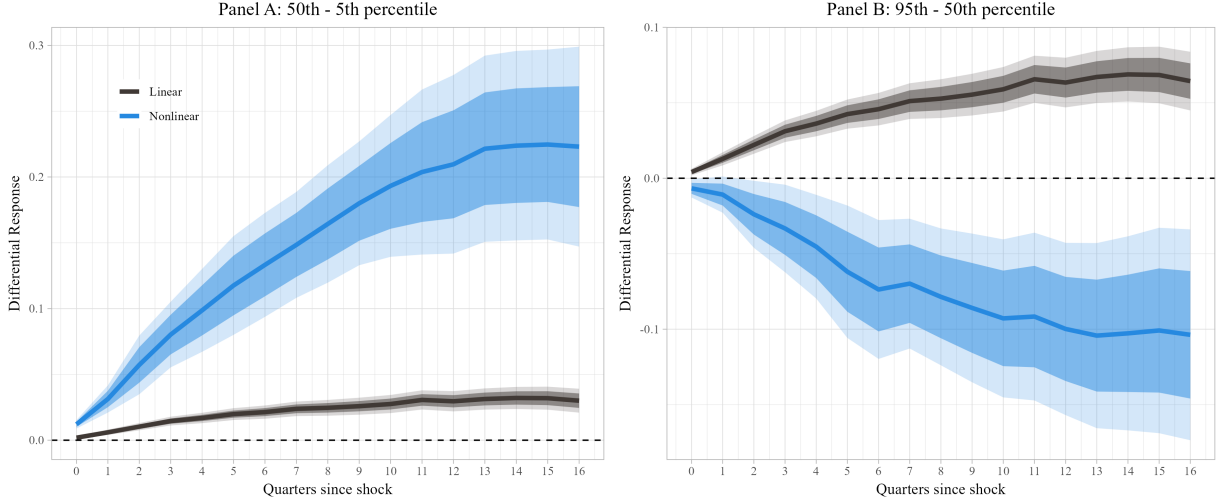


Figure 3: Differential Investment Responses to a Monetary Policy Shock

*Notes:* This figure displays the differential in impulse response functions (IRF) to a 25 b.p. monetary policy shock for firms with different levels of distance to default. The linear specification imposes a linear interaction between the shock and distance to default. The nonlinear specification implements the nonlinear function of distance to default. Panel A plots the difference in IRFs for a firm at the 50th percentile of the distance to default distribution relative to a firm at the 5th percentile; Panel B plots the difference in IRFs for a firm at the 95th percentile relative to a firm at the 50th percentile. Dark (light) shaded regions report 68% (90%) pointwise confidence intervals.

results: the linear estimates predict a positive differential of almost 10% at 16 quarters following the shock, whereas the nonlinear estimates show that this difference is in fact negative and close to -10%. Thus, the linear specification underestimates the strength of the positive relationship between responsiveness and D2D in the left tail of the distribution, and misses entirely on the negative relationship in the right tail. Similar results hold at the shorter horizons as well. In short, imposing a linear specification can lead to misleading conclusions regarding the investment response of firms across the financial distribution.

## 6.2 Aggregate implications of nonlinear responsiveness

Our results have important quantitative implications for the role of financial heterogeneity in the transmission of monetary policy: because most firms are concentrated in regions of the distribution where the nonlinear response exceeds the linear approximation, the latter effectively averages across states and attenuates the estimated impact of monetary policy shocks.

The response of aggregate investment to the monetary policy shock operating through financial heterogeneity is equal to the average of the firm-level responses weighted by their shares of the aggregate capital stock at the time of the shock, i.e.,

$$\Delta k_{t+h} = \underbrace{\sum_i \frac{K_{i,t-1}}{\sum_i K_{i,t-1}} g_h(Z_{i,t-1})}_{\text{aggregate responsiveness}} X_t ,$$

where  $\Delta k_{t+h} = \log K_{t+h} - \log K_{t-1}$  and  $K_t = \sum_i K_{it}$  denotes the aggregate capital stock.<sup>14</sup> As the expression highlights, aggregate responsiveness depends not only on individual responses, but the joint distribution of those responses with firm shares of aggregate capital.<sup>15</sup>

We calculate aggregate responsiveness upon shock impact under both the linear and nonlinear specifications of  $g_h(\cdot)$  using the estimates from Panel A of Figure 2 and the distribution of firm capital on a quarter-by-quarter basis. We plot the results in Figure 4. There are two main takeaways: first, the linear specification significantly underestimates the sensitivity of aggregate investment to monetary policy. On average, the linear specification implies that financial heterogeneity adds about 0.09 percentage points (p.p.) to the responsiveness of aggregate investment. In contrast, the nonlinear specification implies that financial heterogeneity adds about 1.0 p.p., roughly an order of magnitude larger.<sup>16</sup> The finding implies that not only does the linear specification miss on important aspects of the distribution of micro-level responses, but also that those micro-level biases affect “bottom-up” calculations of the implications of micro-level responses for the macro-level response. The result stems largely from the fact that most firms are concentrated in regions of the distribution where the nonlinear response exceeds the linear approximation and hence the linear specification attenuates the estimated aggregate impact of the macroeconomic shock.

The second key takeaway from Figure 4 is that the time-series properties of aggregate responsiveness are quite different across the two specifications; indeed, they have a negative correlation of about -0.7. In other words, the linear specification implies that monetary policy is generally less effective exactly when the nonlinear specification says it will be more so. As a particularly stark example, take the Great Financial Crisis (GFC) and its immediate aftermath from late 2007 through the end of 2009. When firms’ financial conditions deteriorated and more firms moved closer to default, the positive coefficient from the linear LP predicts that monetary policy would be less effective in stimulating investment; in contrast, because the pre-recession economy was above the typical level of financial health (the mean D2D was high), worsening financial conditions pushed firms closer to (and then below) the long-run mean, which is the region where responsiveness is highest, according to our estimates. Thus, the two approaches yield very different implications regarding the states of micro-level financial conditions when monetary policy is most effective: the linear estimates imply monetary policy is most effective at times when firms tend to be financially the strongest (responsiveness from the linear estimates also tends to be high when dispersion in D2D is high), whereas the nonlinear estimates imply monetary policy is most effective when firms tend to be closer to the long-run average level of D2D (responsiveness from the nonlinear estimates also tends to be high when dispersion in D2D is low). In sum, allowing for a flexible approach to estimating the impact of macro shocks on micro variables can be critical for properly assessing the

<sup>14</sup>This calculation captures only the incremental impact of the shock operating via heterogeneity in firms’ financial conditions; the impact of the shock that is common across firms is picked up by the sector-time fixed-effects in equation (28).

<sup>15</sup>A similar result is used in David and Gourio (2023) to decompose the impact of monetary policy on aggregate investment across tangible and intangible capital.

<sup>16</sup>The result is consistent with previous work finding that aggregate non-residential private fixed investment tends to be highly sensitive to monetary policy relative to other categories of GDP (David and Gourio, 2023).

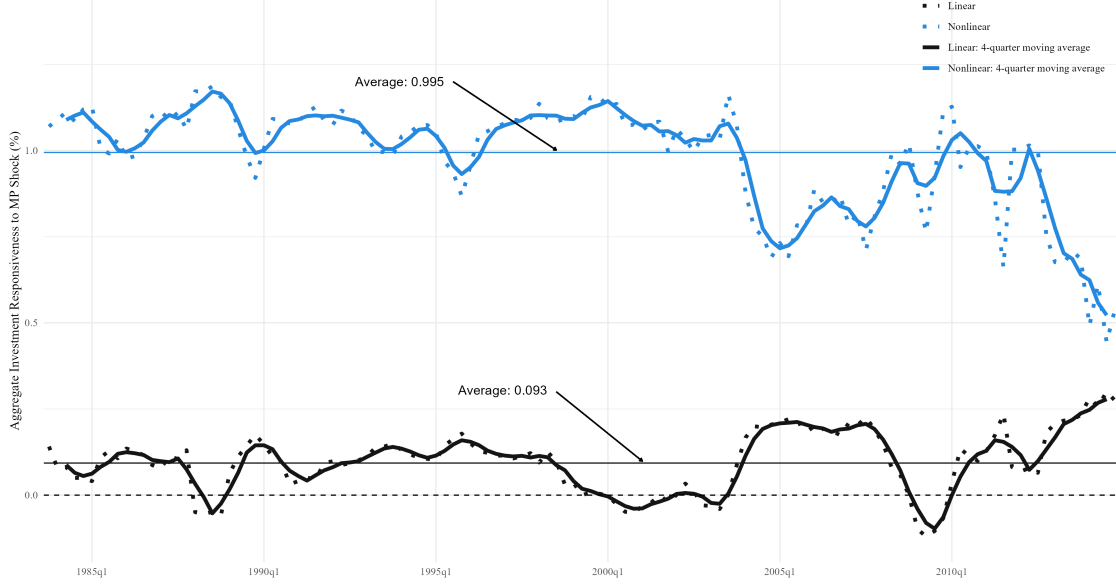


Figure 4: Aggregate Investment Response to a Monetary Policy Shock

*Notes:* This figure displays the responsiveness of aggregate investment to a 25 b.p. monetary policy shock operating through financial heterogeneity across firms as implied by linear and nonlinear local projections. The linear specification imposes a linear interaction between the shock and distance to default in the local projection. The nonlinear specification implements the nonlinear function of distance to default. Each point in the figure displays the impact response of aggregate investment to the monetary policy shock had such a shock occurred at each point in time. Dashed lines show calculated responsiveness; solid lines show a 4-quarter centered moving average.

macro-level implications of micro-level heterogeneity.

## 7 Conclusion

State-dependent local projections are widely used to study heterogeneous responses to aggregate shocks, yet their causal interpretation in nonlinear environments has remained unclear. This paper provides a unified resolution. It shows that state-dependent LPs recover causal impulse responses under a sufficient condition on the conditional mean that is satisfied in environments with state-dependent exposure to a linear shock and constant propagation dynamics. This structure is naturally satisfied in linearized (first-order) solutions of heterogeneous-agent macroeconomic models and in a large class of macro-finance models.

The analysis also reveals that common empirical practice is generally not innocuous. Even in environments where a causal interpretation is valid, standard linear interactions that are misspecified fail to recover the impulse response function and instead identify non-causal projection objects. Recovering the causal estimand therefore generally requires estimating state dependence nonparametrically. We provide a nonparametric sieve approach that achieves this in LPs and delivers asymptotically valid inference both pointwise and for the entire response function.

These results reframe how LPs should be used in practice. In nonlinear settings where shocks affect both exposure and propagation, state-dependent LPs generally lack a causal interpretation.



In contrast, in micro–macro environments where heterogeneity operates through exposure to an aggregate shock that enters linearly and propagation is governed by stable structural primitives, state-dependent LPs can recover economically meaningful causal effects when implemented non-parametrically. This distinction reconciles the use of state-dependent LPs in empirical work with the structure of modern heterogeneous-agent and macro–finance models.

More broadly, the paper highlights a general lesson: identifying heterogeneous causal effects is fundamentally more demanding than estimating average effects, and requires moving beyond parametric approximations. By clarifying both the scope and the limitations of state-dependent LPs, the framework developed here provides a foundation for credible empirical analysis of heterogeneous transmission mechanisms in macroeconomics and macro-finance.

## A Appendix A: Causal Interpretation of Linear State-Dependent Local Projections

To characterise the consequences of misspecifying  $g_h(z)$  as linear, we adopt a simplified setup relative to Section 4. Suppose the user estimates:

$$Y_{i,t+h} = \gamma_h + \beta_h Z_{i,t-1} X_t + u_{i,t+h}, \quad (29)$$

while the true state-dependence is nonlinear:

$$Y_{i,t+h} = \gamma_h + g_h(Z_{i,t-1}) X_t + u_{i,t+h}, \quad (30)$$

We analyse the resulting misspecification in Lemma 1, which generalises Proposition 1 in Kolesár and Plagborg-Møller (2024).

### Assumption 4.

- (a)  $\{(Z_{i,t-1}, X_t)\}$  satisfies:  $\{Z_{i,t-1}\}$  is independent across  $i$  and stationary in  $t$ ;  $\{X_t\}$  is stationary;  $(Z_{i,t-1}, X_t)$  has joint distribution  $F_{Z,X}$ . Cross-sectional dependence via  $X_t$  is allowed.
- (b)  $Z_{i,t-1}$  has a continuous distribution with support  $\mathcal{Z} \subseteq \mathbb{R}$ .
- (c)  $g_h(z)$  is continuously differentiable on  $\mathcal{Z}$ .
- (d)  $\text{Cov}(Z_{i,t-1} X_t, u_{i,t+h}) = 0$ .

### Assumption 5.

- (a)  $Z_{i,t-1}$  and  $X_t$  have finite fourth moments, implying  $\mathbb{E}[(Z_{i,t-1} X_t)^2] < \infty$ , and  $\text{Var}(Z_{i,t-1} X_t) > 0$ .
- (b)  $\mathbb{E}[|g_h(Z_{i,t-1}) X_t| (1 + |Z_{i,t-1} X_t|)] < \infty$ , ensuring applicability of Fubini's theorem.
- (c)  $\int_{\mathcal{Z}} |\omega_h(z) g'_h(z)| dz < \infty$ .

Assumptions 4–5 are weaker than Assumptions 1–3 and Assumption 6: Assumption 4(d) replaces the conditional mean restriction with  $\text{Cov}(Z_{i,t-1} X_t, u_{i,t+h}) = 0$ , sufficient for the OLS probability limit but not for nonparametric identification of  $g_h$ ; Assumption 4(c) requires only continuous differentiability, whereas the main framework assumes  $g_h \in \mathcal{H}^\kappa(\mathcal{Z})$  with  $\kappa > \frac{1}{2}$ . Hence, any DGP satisfying the main assumptions also satisfies those here, and Lemma 1 applies as a special case.

**Lemma 1** (Non-causal estimand under linear misspecification). *Suppose Assumptions 4 and 5 hold, and the DGP is (30). Then the OLS estimand  $\beta_h$  from (29) satisfies*

$$\beta_h = \int_{\mathcal{Z}} \omega_h(z) g'_h(z) dz,$$

where

$$\omega_h(z) := \frac{\text{Cov}(X_t \mathbf{1}\{Z_{i,t-1} \geq z\}, Z_{i,t-1} X_t)}{\text{Var}(Z_{i,t-1} X_t)},$$

and  $\int_{\mathcal{Z}} \omega_h(z) dz = 1$ .

The weight  $\omega_h(z)$  need not be non-negative. Hence  $\beta_h$  is neither  $g_h(z)$  at any state nor a positively weighted average of  $g'_h(z)$ . Unlike the scalar LP case of Kolesár and Plagborg-Møller (2024), where the estimand is a convex combination of causal effects, sign variation in  $\omega_h(z)$  allows offsetting contributions. As a result,  $\beta_h$  may be near zero despite strong nonlinearity, or have the opposite sign to  $g'_h(z)$  over much of  $\mathcal{Z}$ . Thus, linear state-dependent LPs do not recover a causally interpretable object when  $g_h(\cdot)$  is nonlinear. Related issues are discussed in Kitagawa et al. (2025).

### A.1 An Illustrative Example

In practice,  $g_h(z)$  is rarely linear over  $\mathcal{Z}$ , nor is  $\omega_h(z)$  strictly positive. When these conditions fail, the linear coefficient  $\beta_h$  can be misleading. To illustrate the consequences of imposing  $g_h(z) = \beta_h z$  under nonlinearity, we provide the following example.

Let  $Z_{i,t-1} \sim \mathcal{U}[1, 3]$  and consider the non-monotonic function

$$g_h(z) = -\frac{1}{5}z^5 + 2z^4 - \frac{91}{12}z^3 + \frac{27}{2}z^2 - \frac{45}{4}z + \frac{107}{30}, \quad z \in [1, 3],$$

with derivative

$$g'_h(z) = -z^4 + 8z^3 - \frac{91}{4}z^2 + 27z - \frac{45}{4} = (z-1)(z-1.5)(z-2.5)(3-z), \quad z \in [1, 3].$$

From  $g'_h(z)$ , it follows that  $g_h(z)$  is strictly increasing on  $[1, 1.5]$ , decreasing on  $[1.5, 2.5]$ , and increasing again on  $[2.5, 3]$ . The function  $g_h(z)$  and its derivative  $g'_h(z)$  are plotted in Figure 5.

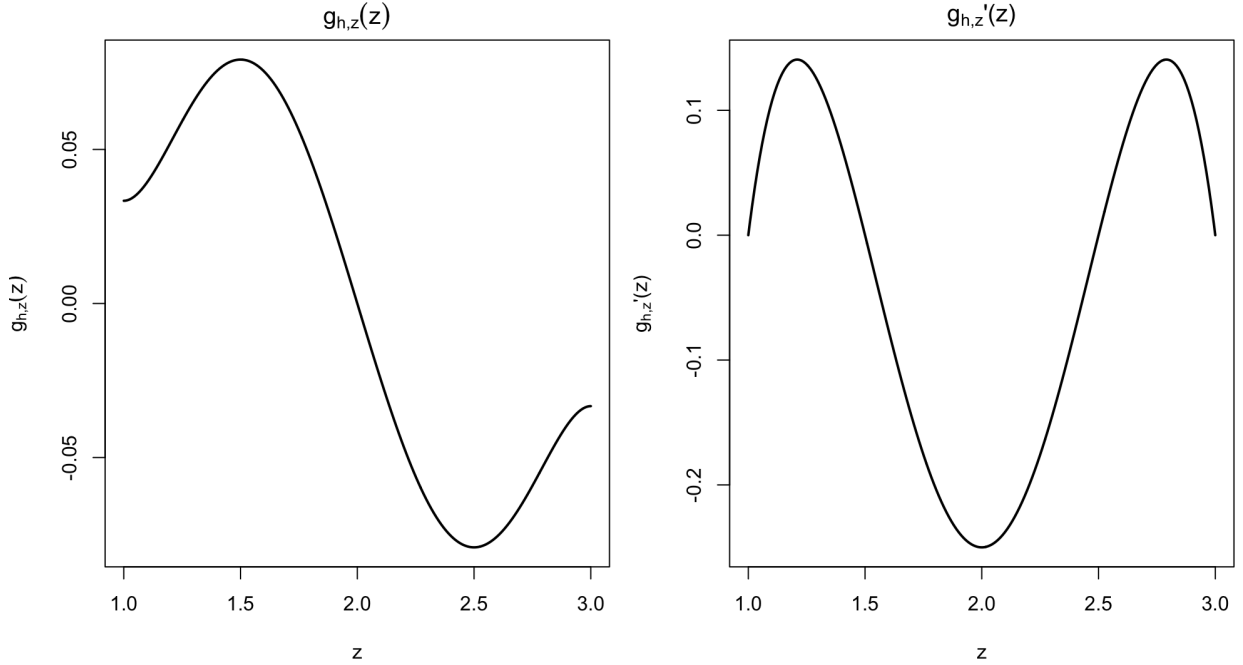


Figure 5: True nonlinear function  $g_h(z)$  and its derivative  $g'_h(z)$

For illustration, set  $X_t = 4 - Z_{i,t-1}$ , a deterministic linear function of  $Z_{i,t-1}$ . This choice is made solely for tractability and to obtain a closed-form expression for  $\omega_h(z)$ ; it is not intended to satisfy the exogeneity condition of Assumption 6. The qualitative conclusion - that  $\omega_h(z)$  need not be non-negative - does not rely on this specification. Then the weight function  $\omega_h(z)$  is

$$\begin{aligned}\omega_h(z) &= \frac{\text{Cov}(X_t \mathbf{1}\{Z_{i,t-1} \geq z\}, Z_{i,t-1} X_t)}{\text{Var}(Z_{i,t-1} X_t)} \\ &= -\frac{15}{32}(3z^4 - 32z^3 + 118z^2 - 176z + 87) \\ &= \frac{15}{32}(z-1)(3-z)(3z^2 - 20z + 29), \quad z \in [1, 3].\end{aligned}$$

One can verify that  $\int_1^3 \omega_h(z) dz = 1$ . Since  $3z^2 - 20z + 29 = 0$  has roots  $(10 - \sqrt{13})/3 \approx 2.13$  and  $(10 + \sqrt{13})/3 \approx 4.54$ , the weight function  $\omega_h(z)$  is positive on  $[1, 2.13]$  and negative on  $(2.13, 3]$ . The weight function is plotted in Figure 6.

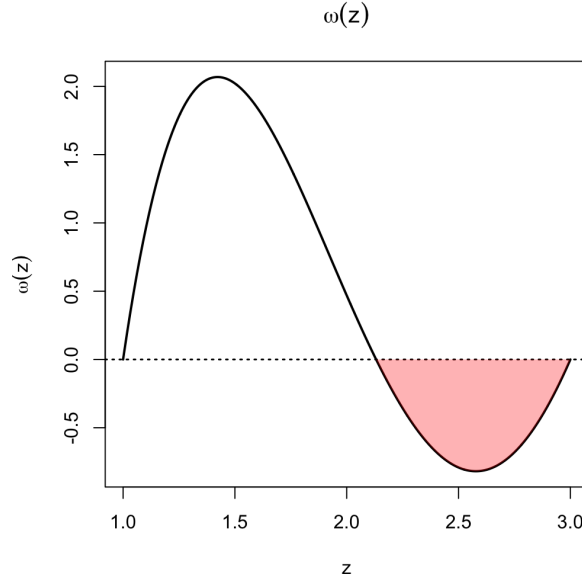


Figure 6: Weight function  $\omega_h(z)$

Under the true model (30) with  $g_h(z)$  and  $Z_{i,t-1}$ ,  $X_t$  as specified above, Lemma 1 gives

$$\beta_h = \int_1^3 \omega_h(z) g'_h(z) dz = -\frac{1}{28}.$$

The issue is that  $\omega_h(z)$  is negative over part of its support (Figure 6), so  $\beta_h$  cannot be interpreted as a weighted average slope or treatment effect. Although  $\beta_h = -1/28$  is the correct weighted average derivative, sign variation in  $\omega_h(z)$  implies that positive slopes of  $g_h(z)$  where  $\omega_h(z) < 0$  are subtracted, leading to cancellation of opposing marginal effects. Consequently,  $\beta_h$  does not reflect the shape of  $g_h(z)$  and cannot be interpreted as a representative marginal effect at any  $z \in \mathcal{Z}$ .

## B Appendix B: Technical Details and Asymptotic Theory for the Sieve Method

This appendix provides the formal setup, technical details, and asymptotic theory underlying the sieve estimator and inference procedures described in Section 4. It serves as a bridge between the main-text exposition and the proofs of asymptotic theory in Appendix C. In particular, the appendix introduces the stacked notation, block representations, residual definitions, and regularity conditions needed to establish consistency, asymptotic normality, and the validity of pointwise confidence intervals and uniform confidence bands.

### B.1 Sieve Estimation and Inference

For each horizon  $h \in \{0, 1, \dots, H\}$ , recall our main LP specification:

$$Y_{i,t+h} = g_h(Z_{i,t-1})X_t + W_{i,t-1}^\top \gamma_h + u_{i,t+h}.$$

Our goal is to estimate and conduct inference on the potentially nonlinear function  $g_h(\cdot)$ . We approximate  $g_h(\cdot)$  using a sieve expansion. Let  $\{\phi_{1,J}(z), \dots, \phi_{J,J}(z)\}$  denote a collection of sieve basis functions. In the main text, we use cubic B-splines, but the notation below also covers other bases, such as Fourier series or wavelets. To simplify notation in this appendix and in the proofs in Appendix C, we suppress the dependence on the sieve dimension  $J$  whenever  $J$  is fixed or clear from context, and write

$$\phi(z) = (\phi_1(z), \dots, \phi_J(z))^\top.$$

The sieve approximation is

$$g_h(z) = \sum_{j=1}^J b_{h,j} \phi_j(z) + R_h(z) = \phi(z)^\top b_h + R_h(z),$$

where  $b_h = (b_{h,1}, \dots, b_{h,J})^\top$  collects the sieve coefficients,  $R_h(z)$  is the sieve approximation error, and  $J$  controls the dimension of the sieve space. In the main text, we select  $J$  by AIC over a finite candidate set. In the simulations, we also examine alternative selection rules, including generalized cross-validation (GCV) and LASSO-based selection, and compare their finite-sample performance.

Substituting the sieve approximation yields estimation equation

$$Y_{i,t+h} = X_t \phi(Z_{i,t-1})^\top b_h + W_{i,t-1}^\top \gamma_h + R_h(Z_{i,t-1})X_t + u_{i,t+h}. \quad (31)$$

For notation stacked across individuals  $i$ , define

$$\begin{aligned} Y_{t+h} &= (Y_{1,t+h}, \dots, Y_{N,t+h})^\top, & Z_{t-1} &= (Z_{1,t-1}, \dots, Z_{N,t-1})^\top, \\ W_{t-1} &= \begin{pmatrix} W_{1,t-1}^\top \\ \vdots \\ W_{N,t-1}^\top \end{pmatrix} \in \mathbb{R}^{N \times Q}, & u_{h,t+h} &= (u_{1,t+h}, \dots, u_{N,t+h})^\top. \end{aligned}$$

Let

$$\Phi(Z_{t-1}) = \begin{pmatrix} \phi(Z_{1,t-1})^\top \\ \vdots \\ \phi(Z_{N,t-1})^\top \end{pmatrix} \in \mathbb{R}^{N \times J}, \quad R_h(Z_{t-1}) = (R_h(Z_{1,t-1}), \dots, R_h(Z_{N,t-1}))^\top.$$

Then the stacked estimation equation is

$$Y_{t+h} = X_t \Phi(Z_{t-1}) b_h + W_{t-1} \gamma_h + R_h(Z_{t-1}) X_t + u_{h,t+h}. \quad (32)$$

Define the composite sieve-regression error

$$v_{h,i,t+h} = R_h(Z_{i,t-1}) X_t + u_{i,t+h}, \quad v_{h,t+h} = R_h(Z_{t-1}) X_t + u_{h,t+h}.$$

This composite error includes both the original regression error and the sieve approximation error. In the main text, the same object is denoted more simply by  $u_{i,t+h}$  in the finite-dimensional sieve regression; throughout this appendix, we use  $v_{h,i,t+h}$  to distinguish the composite sieve-regression error from the structural LP error  $u_{i,t+h}$  in the original LP.

Also define  $D_{it}^\top = (X_t \phi(Z_{i,t-1})^\top, W_{i,t-1}^\top)$ ,  $\theta_h^\top = (b_h^\top, \gamma_h^\top)$ , and

$$D_t = \begin{pmatrix} D_{1t}^\top \\ \vdots \\ D_{Nt}^\top \end{pmatrix} = (X_t \Phi(Z_{t-1}), W_{t-1}) \in \mathbb{R}^{N \times (J+Q)}.$$

Then (31) and (32) can be written as

$$Y_{i,t+h} = D_{it}^\top \theta_h + v_{h,i,t+h}, \quad Y_{t+h} = D_t \theta_h + v_{h,t+h}. \quad (33)$$

Estimating (33) by OLS yields

$$\begin{aligned} \hat{\theta}_h &= \left( \sum_{i=1}^N \sum_{t=1}^{T-h} D_{it} D_{it}^\top \right)^{-1} \left( \sum_{i=1}^N \sum_{t=1}^{T-h} D_{it} Y_{i,t+h} \right) \\ &= \left( \sum_{t=1}^{T-h} D_t^\top D_t \right)^{-1} \left( \sum_{t=1}^{T-h} D_t^\top Y_{t+h} \right). \end{aligned}$$

Writing  $\widehat{\theta}_h^\top = (\widehat{b}_h^\top, \widehat{\gamma}_h^\top)$ , the estimated impulse response function is

$$\widehat{g}_h(z) = \phi(z)^\top \widehat{b}_h.$$

The OLS residual from the sieve regression is

$$\widehat{v}_{h,i,t+h} = Y_{i,t+h} - D_{it}^\top \widehat{\theta}_h, \quad \widehat{v}_{h,t+h} = Y_{t+h} - D_t^\top \widehat{\theta}_h.$$

In the main text, the same residual is denoted more simply by  $\widehat{u}_{i,t+h}$ . When a data-driven sieve dimension  $\widehat{J}_h$  is used, all quantities in this appendix are evaluated at  $J = \widehat{J}_h$ .

Since our main interest is on  $g_h(\cdot)$ , and hence on  $b_h$ , it is useful to separate the sieve coefficients from the coefficients on the controls. Let  $n_h = N(T - h)$  and define

$$\widehat{A}_h = \frac{1}{n_h} \sum_{i=1}^N \sum_{t=1}^{T-h} D_{it} D_{it}^\top = \begin{pmatrix} \widehat{A}_{11,h} & \widehat{A}_{12,h} \\ \widehat{A}_{21,h} & \widehat{A}_{22,h} \end{pmatrix}, \quad \widehat{B}_h = \frac{1}{n_h} \sum_{i=1}^N \sum_{t=1}^{T-h} D_{it} Y_{i,t+h} = \begin{pmatrix} \widehat{B}_{1,h} \\ \widehat{B}_{2,h} \end{pmatrix}.$$

Equivalently, in the stacked notation,

$$\widehat{A}_h = \frac{1}{n_h} \sum_{t=1}^{T-h} \begin{pmatrix} X_t^2 \Phi(Z_{t-1})^\top \Phi(Z_{t-1}) & X_t \Phi(Z_{t-1})^\top W_{t-1} \\ X_t W_{t-1}^\top \Phi(Z_{t-1}) & W_{t-1}^\top W_{t-1} \end{pmatrix},$$

$$\widehat{B}_h = \frac{1}{n_h} \sum_{t=1}^{T-h} \begin{pmatrix} X_t \Phi(Z_{t-1})^\top Y_{t+h} \\ W_{t-1}^\top Y_{t+h} \end{pmatrix}.$$

Here  $\widehat{A}_{11,h} \in \mathbb{R}^{J \times J}$ ,  $\widehat{A}_{22,h} \in \mathbb{R}^{Q \times Q}$ ,  $\widehat{A}_{12,h} = \widehat{A}_{21,h}^\top \in \mathbb{R}^{J \times Q}$ ,  $\widehat{B}_{1,h} \in \mathbb{R}^J$ , and  $\widehat{B}_{2,h} \in \mathbb{R}^Q$ .

As in the main text, let

$$\widehat{\widehat{A}}_{11,h} = \widehat{A}_{11,h} - \widehat{A}_{12,h} \widehat{A}_{22,h}^{-1} \widehat{A}_{21,h}$$

denote the Schur complement. The block matrix inversion formula gives

$$\widehat{b}_h = \widehat{\widehat{A}}_{11,h}^{-1} \left( \widehat{B}_{1,h} - \widehat{A}_{12,h} \widehat{A}_{22,h}^{-1} \widehat{B}_{2,h} \right).$$

This representation is useful both computationally and conceptually:  $\widehat{b}_h$  can be viewed as the coefficient from regressing the residualized sieve interaction  $X_t \phi(Z_{i,t-1})$  on the residualized outcome, after partialling out  $W_{i,t-1}$ . Analogously, the estimation-error representation follows as

$$\widehat{b}_h - b_h = \widehat{\widehat{A}}_{11,h}^{-1} \left[ \frac{1}{n_h} \sum_{t=1}^{T-h} \left( X_t \Phi(Z_{t-1})^\top - \widehat{A}_{12,h} \widehat{A}_{22,h}^{-1} W_{t-1}^\top \right) v_{h,t+h} \right]. \quad (34)$$

For feasible inference, it is convenient to collect the partialled-out sieve regressors at each date  $t$  into the  $J \times N$  matrix

$$\widetilde{P}_{h,t} = X_t \Phi(Z_{t-1})^\top - \widehat{A}_{12,h} \widehat{A}_{22,h}^{-1} W_{t-1}^\top \in \mathbb{R}^{J \times N}.$$

Equivalently,

$$\tilde{P}_{h,t} = (\tilde{P}_{h,1,t}, \dots, \tilde{P}_{h,N,t}),$$

where the individual-level partialled-out sieve regressor defined in the main text is

$$\tilde{P}_{h,i,t} = X_t \phi(Z_{i,t-1}) - \hat{A}_{12,h} \hat{A}_{22,h}^{-1} W_{i,t-1} \in \mathbb{R}^J.$$

The score process used for inference can then be written either in stacked form or as a cross-sectional average:

$$\hat{s}_{h,t} = \tilde{P}_{h,t} \hat{v}_{h,t+h} = \sum_{i=1}^N \tilde{P}_{h,i,t} \hat{v}_{h,i,t+h}.$$

As in the main text, we estimate the long-run covariance matrix of  $\{\hat{s}_{h,t}\}_{t=1}^{T-h}$  by  $\hat{\Omega}_h$  and the covariance matrix of  $\hat{b}_h$  by  $\hat{V}_h$ . Pointwise and uniform inference for  $g_h(\cdot)$  then proceeds using Algorithms 2 and 3 in Section 4.2.

## B.2 Asymptotic Theory

**Assumption 6** (Regularity Conditions). Maintain Assumptions 1–3. The following additional technical conditions hold.

1. **(Cross-Sectional and Time-Series Structure)** The array  $\{(Z_{i,t-1}, W_{i,t-1})\}$  is i.i.d. across  $i$  and strictly stationary over  $t$ . The variable  $Z_{i,t-1}$  has a continuous distribution with support  $\mathcal{Z}$ . The true function  $g_h(\cdot)$  belongs to the bounded Hölder ball

$$\mathcal{G}_M = \left\{ g \in \mathcal{H}^\kappa(\mathcal{Z}) \mid \|g\|_{\mathcal{H}^\kappa} \leq M \right\},$$

where  $\kappa > \frac{1}{2}$ ,  $M < \infty$ , and

$$\|g\|_{\mathcal{H}^\kappa} := \sum_{s \leq \lfloor \kappa \rfloor} \sup_z |D^s g(z)| + \sum_{s = \lfloor \kappa \rfloor} \sup_{z \neq z'} \frac{|D^s g(z) - D^s g(z')|}{|z - z'|^{\kappa - \lfloor \kappa \rfloor}}.$$

The class  $\mathcal{G}_M$  is of VC-type: there exist constants  $A, V < \infty$  such that for all  $\varepsilon > 0$ ,

$$\mathcal{N}(\varepsilon, \mathcal{G}_M, \|\cdot\|_\infty) \leq \left( \frac{A}{\varepsilon} \right)^V.$$

2. **(Error Process)** Let  $\gamma_h := A_{22,h}^{-1} \mathbb{E}[W_{i,t-1}(Y_{i,t+h} - g_h(Z_{i,t-1})X_t)]$  be the population projection coefficient of the nuisance onto the controls. Define the regression residual

$$u_{i,t+h} := Y_{i,t+h} - g_h(Z_{i,t-1})X_t - W_{i,t-1}^\top \gamma_h,$$



which admits a moving-average representation

$$u_{h,t+h} = \sum_{l=0}^{\infty} a_{h,l} \eta_{t+h-l},$$

where  $\{\eta_t\}$  is i.i.d. with mean zero and finite fourth moments,  $\eta_t$  is independent of  $\mathcal{F}_{t-1}$ , and

$$\sum_{l=0}^{\infty} \|a_{h,l}\| < \infty, \quad \sum_{l=0}^{\infty} l^{1+\delta} \|a_{h,l}\| < \infty \quad \text{for some } \delta > 0.$$

Moreover,  $\mathbb{E}[u_{i,t+h}^4] < \infty$ . The long-run covariance matrix of the effective score satisfies

$$\lambda_{\min}(\Omega_h^*) \geq c > 0.$$

3. **(Moment Conditions)**  $Z_{i,t-1}$ ,  $W_{i,t-1}$ , and  $X_t$  have finite fourth moments.

4. **(Matrix Regularity)**

$$\lambda_{\min}(A_{11,h}) \geq c_1 > 0, \quad \lambda_{\max}(A_{11,h}) \leq C_1 < \infty, \quad s_{\min}(A_{22,h}) > 0.$$

5. **(Sieve Basis and Approximation)** Let  $\phi(z) = (\phi_1(z), \dots, \phi_J(z))^\top$ .

$$\sup_{z \in \mathcal{Z}} \|\phi(z)\| \leq C\sqrt{J}, c_2 \leq s_{\min}\left(\mathbb{E}[\phi(Z_{i,t-1})\phi(Z_{i,t-1})^\top]\right) \leq s_{\max}\left(\mathbb{E}[\phi(Z_{i,t-1})\phi(Z_{i,t-1})^\top]\right) \leq C_2.$$

6. **(Growth Conditions)**

$$J \rightarrow \infty, \quad \frac{J^2}{NT} \rightarrow 0.$$

Assumptions 1–3 in the main text collect the structural identification conditions. Assumption 1 is the key exogeneity requirement, ensuring the instrument  $X_t$  is orthogonal to all past information. Assumption 2 ensures the state variable and controls are predetermined, so they enter the conditioning set  $\mathcal{F}_{t-1}$  before the shock is realised. Assumption 3 imposes the state-dependent linear structure on the conditional mean, which defines the impulse response function  $g_h$  to be estimated.

Assumption 6(1) adds the cross-sectional and time-series regularity needed for the asymptotic theory. Cross-sectional independence and stationarity over  $t$  are standard in large- $N$ , large- $T$  panels. The bounded Hölder ball  $\mathcal{G}_M$  with  $\kappa > \frac{1}{2}$  is the minimal smoothness for the sieve bias to vanish faster than the estimation variance, and the VC-type entropy condition on  $\mathcal{G}_M$  supports the uniform law of large numbers used in the proof. Under Assumption 6(1), if the sieve basis  $\{\phi_j\}_{j=1}^J$  is a standard basis (B-splines or wavelets), then classical approximation theory (e.g., Schumaker, 1981; DeVore and Lorentz, 1993) implies

$$\sup_{z \in \mathcal{Z}} |R_h(z)| = O(J^{-\kappa}).$$

This rate controls the approximation bias in the proof of Theorem 2, and together with the growth condition  $J^2/(NT) \rightarrow 0$  in Assumption 6(6), yields the overall convergence rate

$$\|\widehat{b}_h - b_h\| = O_p(J^{-\kappa}) + O_p\left(\sqrt{\frac{J}{NT}}\right).$$

Assumption 6(2) defines the regression residual by projecting the nuisance  $r_{i,h}(\mathcal{F}_{t-1})$  onto the controls  $W_{i,t-1}$ , absorbing any approximation error into  $u_{i,t+h}$ . The MA representation permits serial correlation of arbitrary form up to any finite horizon, which is the defining feature of direct local projections relative to VAR-based methods. The non-degeneracy condition  $\lambda_{\min}(\Omega_h^*) \geq c > 0$  ensures the pointwise CLT has a non-degenerate Gaussian limit. Assumption 6(3) imposes finite fourth moments to control higher-order terms in the variance estimation. Assumption 6(4) ensures the sieve second-moment matrix and the control partialling-out step are both well-conditioned uniformly in  $J$ . Assumption 6(5) collects standard regularity conditions on the basis, including the Riesz-type two-sided bound on  $\mathbb{E}[\phi\phi^\top]$  needed for identification of  $b_h$ . Finally, Assumption 6(6) requires  $J^2/(NT) \rightarrow 0$ , balancing bias and variance; the optimal choice  $J \asymp (NT)^{1/(2\kappa+1)}$  achieves the minimax rate over  $\mathcal{H}^\kappa$ .

**Theorem 2.** *Assume Assumption 6. Using the sieve approximation  $g_h(z) = \phi(z)^\top b_h + R_h(z)$  with  $\sup_{z \in \mathcal{Z}} |R_h(z)| = O(J^{-\kappa})$ , the OLS sieve estimator  $\widehat{g}_h(z) = \phi(z)^\top \widehat{b}_h$  satisfies the coefficient rate*

$$\|\widehat{b}_h - b_h\|_2 = O_p\left(\sqrt{\frac{J}{NT}}\right) + O_p(J^{1/2-\kappa}).$$

*If, in addition, the undersmoothing condition*

$$\sqrt{NT} J^{1/2-\kappa} \rightarrow 0$$

*holds, then for each fixed  $z \in \mathcal{Z}$ ,*

$$\frac{\sqrt{NT}(\widehat{g}_h(z) - g_h(z))}{\sigma_h^*(z)} \Rightarrow \mathcal{N}(0, 1),$$

*where*

$$\begin{aligned} \sigma_h^{*2}(z) &:= \phi(z)^\top A_{11,h}^{-1} \Omega_h^* A_{11,h}^{-1} \phi(z), \\ \Omega_h^* &= \frac{1}{N} \sum_{k \in \mathbb{Z}} \mathbb{E}[s_{h,t}^* s_{h,t-k}^{*\top}], \quad s_{h,t}^* := X_t \Phi(Z_{t-1})^\top u_{h,t+h}. \end{aligned}$$

**Theorem 3** (Consistency of the HAC variance-covariance estimator). *Fix  $h \in \{0, 1, \dots, H\}$ . Suppose Assumption 6 holds. Let the residuals be*

$$\widehat{v}_{h,t+h} := Y_{t+h} - D_t \widehat{\theta}_h.$$

Define

$$\hat{q}_{h,t} := X_t \Phi(Z_{t-1})^\top W_{t-1} \in \mathbb{R}^{J \times Q}, \quad \hat{A}_{22,h} := \frac{1}{NT} \sum_{t=1}^{T-h} W_{t-1}^\top W_{t-1} \in \mathbb{R}^{Q \times Q},$$

$$\hat{\mu}_{W\hat{v}} := \frac{1}{NT} \sum_{t=1}^{T-h} W_{t-1}^\top \hat{v}_{h,t+h} \in \mathbb{R}^Q.$$

Let the effective score process be

$$\hat{s}_{h,t}^* := X_t \Phi(Z_{t-1})^\top \hat{v}_{h,t+h} - \hat{q}_{h,t} \hat{A}_{22,h}^{-1} \hat{\mu}_{W\hat{v}} \in \mathbb{R}^J.$$

Let  $\hat{\Omega}_h^*$  be the Bartlett HAC estimator

$$\begin{aligned} \hat{\Omega}_h^* &:= \frac{1}{NT} \sum_{t=1}^{T-h} \hat{s}_{h,t}^* \hat{s}_{h,t}^{*\top} \\ &\quad + \sum_{k=1}^L w_k \left( \frac{1}{NT} \sum_{t=k+1}^{T-h} \left( \hat{s}_{h,t}^* \hat{s}_{h,t-k}^{*\top} + \hat{s}_{h,t-k}^* \hat{s}_{h,t}^{*\top} \right) \right), \end{aligned}$$

with Bartlett weights  $w_k = 1 - \frac{k}{L+1}$ , where the truncation lag  $L = L_{NT}$  satisfies

$$L \rightarrow \infty, \quad \frac{L}{T} \rightarrow 0, \quad \frac{J^2}{T} + \frac{J^2 L}{NT} \rightarrow 0, \quad J^{1-2\kappa} \rightarrow 0.$$

Then, as  $T \rightarrow \infty$  (and either  $N$  fixed or  $N \rightarrow \infty$  subject to Assumption 6(6)),

$$\|\hat{\Omega}_h^* - \Omega_h^*\| \xrightarrow{p} 0, \quad \|\hat{V}_h^* - V_h^*\| \xrightarrow{p} 0,$$

where

$$\hat{V}_h^* := \hat{A}_{11,h}^{-1} \hat{\Omega}_h^* \hat{A}_{11,h}^{-1}, \quad V_h^* := A_{11,h}^{-1} \Omega_h^* A_{11,h}^{-1},$$

and  $\|\cdot\|$  denotes the operator norm. In particular, for any fixed  $z \in \mathcal{Z}$ ,

$$\phi(z)^\top \hat{V}_h^* \phi(z) - \phi(z)^\top V_h^* \phi(z) = o_p(J).$$

**Lemma 2** (Rate of convergence for  $\hat{A}_{11,h}$ ). Define

$$\mathbf{S}_t := \frac{1}{N} \sum_{i=1}^N \phi(Z_{i,t-1}) \phi(Z_{i,t-1})^\top, \quad \hat{A}_{11,h} := \frac{1}{T-h} \sum_{t=1}^{T-h} X_t^2 \mathbf{S}_t,$$

and the population analogue  $A_{11,h} = \mathbb{E}[X_t^2 \phi(Z_{t-1}) \phi(Z_{t-1})^\top]$ . Suppose Assumption 6 holds, and either (i)  $\mathbb{E}[X_t^2 \mid \mathcal{F}_{t-1}] = \sigma_X^2$  a.s. (conditional homoskedasticity), or (ii)  $\mathbb{E}[X_t^4] < \infty$  (Assumption 6(3) strengthened to fourth moments). Then:

**(a) Operator-norm consistency:**

$$\|\hat{A}_{11,h} - A_{11,h}\| = O_p(T^{-1/2} + \sqrt{J/(NT)} + J/(NT)).$$

**(b) Invertibility and bounded inverse:** If  $\lambda_{\min}(A_{11,h}) \geq c_1 > 0$  (Assumption 6(4)), then with probability approaching one,  $A_{11,h}$  is invertible and  $\|\hat{A}_{11,h}^{-1}\| = O_p(1)$ .

**(c) Neumann series expansion:** Under (a) and (b),

$$\hat{A}_{11,h}^{-1} = A_{11,h}^{-1} + R_{A,n}, \quad \|R_{A,n}\| = O_p(T^{-1/2} + \sqrt{J/(NT)} + J/(NT)).$$

**Theorem 4** (Uniform convergence and Gaussian approximation). Fix  $h \in \{0, 1, \dots, H\}$ . Suppose:

- (i) the assumptions of Lemma 2 hold,
- (ii) the sieve basis  $\phi(\cdot)$  is  $O(J^\rho)$ -Lipschitz on  $\mathcal{Z}$  (i.e.,  $\|\phi(z) - \phi(z')\|_2 \leq C J|z - z'|$ ), and the Riesz  $\ell^1$ -localization

$$\sup_{z \in \mathcal{Z}} \|A_{11,h}^{-1} \phi(z)\|_1 = O(J^\rho)$$

holds for some  $\rho \geq 0$  (with  $\rho = 0$  for B-splines and  $\rho = 1/2$  for non-localized orthonormal bases);

- (iii) for some  $q > 4$ ,  $\mathbb{E}|X_t|^{2q} + \mathbb{E}|u_{i,t+h}|^{2q} + \mathbb{E}\|\eta_t\|^{2q} < \infty$ , strengthening the fourth-moment conditions in Assumption 6(2)–(3);
- (iv) the sieve dimension satisfies the undersmoothing and Gaussian-approximation growth conditions

$$\sqrt{NT} J^{1/2-\kappa} \rightarrow 0 \quad \text{and} \quad J^\rho (\log J)^{7/6} = o(T^{1/9}),$$

together with  $J = o(T)$ .

Then the following hold:

**(a) Sup-norm rate.**

$$\sup_{z \in \mathcal{Z}} |\hat{g}_h(z) - g_h(z)| = O_p\left(\frac{J}{\sqrt{NT}}\right) + O_p\left(J^{1/2-\kappa}\right).$$

**(b) Gaussian approximation.**

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}\left(\sqrt{NT} \sup_{z \in \mathcal{Z}} |\hat{g}_h(z) - g_h(z)| \leq t\right) - \mathbb{P}\left(\sup_{z \in \mathcal{Z}} |Z_h(z)| \leq t\right) \right| \rightarrow 0,$$

where

$$Z_h(z) := \phi(z)^\top A_{11,h}^{-1} (\Omega_h^*)^{1/2} \mathcal{N}_J, \quad \mathcal{N}_J \sim \mathcal{N}(0, I_J),$$

is a centered Gaussian process with covariance kernel  $\mathbb{E}[Z_h(z)Z_h(z')] = \sigma_h^{*2}(z, z')$  matching the asymptotic variance of the pointwise CLT in Theorem 2.

The following uniform confidence-band corollary is implied.

**Corollary 1** (Uniform confidence band). *Under the hypotheses of Theorem 4, define the critical value*

$$c_{h,1-\alpha}^* := \inf \left\{ c > 0 : \mathbb{P} \left( \sup_{z \in \mathcal{Z}} |Z_h(z)| \leq c \right) \geq 1 - \alpha \right\}.$$

*Then the band  $\widehat{g}_h(z) \pm c_{h,1-\alpha}^* / \sqrt{NT}$  has asymptotic coverage:*

$$\mathbb{P} \left( \sup_{z \in \mathcal{Z}} |\widehat{g}_h(z) - g_h(z)| \leq c_{h,1-\alpha}^* / \sqrt{NT} \right) \longrightarrow 1 - \alpha.$$

*Remark 3* (Admissible rates). Under hypotheses (i)–(iv) of Theorem 4, the joint rate of  $(N, T, J)$  is constrained by: (a) undersmoothing  $\sqrt{NT} J^{1/2-\kappa} \rightarrow 0$ ; (b) Assumption 6(6),  $J^2 = o(NT)$ ; (c) GA growth  $J^\rho (\log J)^{7/6} = o(T^{1/9})$ ; (d) sieve dimension constraint  $J = o(T)$  (needed for the linear representation in Step 2).

*B-spline bases* ( $\rho = 0$ ). For  $N \asymp T$  and smoothness  $\kappa > 3/2$ , constraints (a) and (b) reduce to

$$\max\left(\frac{1}{2}, \frac{2}{2\kappa-1}\right) < \frac{\log J}{\log T} < 1,$$

which is non-vacuous for  $\kappa > 3/2$ . For example,  $\kappa = 2$  admits  $T^{2/3} \ll J \ll T$ . Constraint (c) is essentially trivial ( $\log J$  appears only logarithmically). Constraint (d) is automatic from (b) when  $N \leq CT$ .

*Orthonormal polynomial bases* ( $\rho = 1/2$ ). The GA growth constraint (c) becomes  $J^{1/2} (\log J)^{7/6}$  which is  $o(T^{1/9})$ , i.e.,  $J \ll T^{2/9}$  up to logarithmic factors. This is incompatible with the sup-norm undersmoothing condition (a) unless  $\kappa > 11/4$ , since (a) requires  $J \gg (NT)^{1/(2\kappa-1)}$ , and for  $N \asymp T$  this lower bound exceeds  $T^{2/9}$  unless  $\kappa > 11/4$ .

*Large- $N$  regimes.* For  $N \gg T$  (e.g.,  $N \asymp T^a$  with  $a > 1$ ), constraint (d)  $J = o(T)$  is no longer implied by (b) and must be imposed separately. Constraint (a) then gives  $J \gg (NT)^{1/(2\kappa-1)} = T^{(1+a)/(2\kappa-1)}$ , which combined with  $J = o(T)$  requires  $(1+a)/(2\kappa-1) < 1$ , i.e.,  $\kappa > (1+a)/2$ . For example,  $N \asymp T^2$  requires  $\kappa > 3/2$  (same as the panel case);  $N \asymp T^3$  requires  $\kappa > 2$ .

## C Appendix C: Proofs

*Proof of Proposition 1.* For any  $z$  with  $\mathbb{E}[X_t^2 \mid Z_{i,t-1} = z] > 0$ ,

$$g_h(z) = \frac{\mathbb{E}[X_t Y_{i,t+h} \mid Z_{i,t-1} = z]}{\mathbb{E}[X_t^2 \mid Z_{i,t-1} = z]}.$$

Using the conditional mean restriction and iterated expectations,

$$\mathbb{E}[X_t Y_{i,t+h} \mid Z_{i,t-1} = z] = \mathbb{E}[\tilde{g}_h(s_{i,t-1}) X_t^2 \mid Z_{i,t-1} = z] + \mathbb{E}[X_t r_{i,h}(\mathcal{F}_{t-1}) \mid Z_{i,t-1} = z].$$

The second term is zero since  $r_{i,h}(\mathcal{F}_{t-1})$  is  $\mathcal{F}_{t-1}$ -measurable and  $\mathbb{E}[X_t \mid \mathcal{F}_{t-1}] = 0$ . Hence

$$g_h(z) = \frac{\mathbb{E}[\tilde{g}_h(s_{i,t-1})X_t^2 \mid Z_{i,t-1} = z]}{\mathbb{E}[X_t^2 \mid Z_{i,t-1} = z]}.$$

Applying iterated expectations over  $s_{i,t-1}$  yields the result with weights proportional to  $\mathbb{E}[X_t^2 \mid s_{i,t-1}, Z_{i,t-1} = z]$ , which are nonnegative and integrate to one.  $\square$

*Proof of Lemma 1.* The misspecified model (29), together with Assumption 4(d), gives

$$\text{Cov}(Z_{i,t-1}X_t, Y_{i,t+h} - \gamma_h - \beta_h Z_{i,t-1}X_t) = 0 \iff \beta_h = \frac{\text{Cov}(Y_{i,t+h}, Z_{i,t-1}X_t)}{\text{Var}(Z_{i,t-1}X_t)}.$$

According to the ground truth model (30)

$$\beta_h = \frac{\text{Cov}(\gamma_h + g_h(Z_{i,t-1})X_t + u_{i,t+h}, Z_{i,t-1}X_t)}{\text{Var}(Z_{i,t-1}X_t)} = \frac{\text{Cov}(g_h(Z_{i,t-1})X_t, Z_{i,t-1}X_t)}{\text{Var}(Z_{i,t-1}X_t)},$$

where the second equality uses Assumption 4(d) and the fact that  $\text{Cov}(\gamma_h, \cdot) = 0$ .

By Assumption 4(a),  $\{Z_{i,t-1}\}$  is independent across  $i$  and  $\{(Z_{i,t-1}, X_t)\}$  is strictly stationary over  $t$  with joint distribution function  $F_{Z,X}(z, x)$ . Without loss of generality, assume  $\mathcal{Z} = \mathbb{R}$  and  $g_h(-\infty) = 0$ , so that  $g_h(z) = \int_{-\infty}^z g'_h(u) du$  by the fundamental theorem of calculus and Assumption 4(c). Then, by Assumption 5(b) and Fubini's theorem (justified by the integrability condition therein),

$$\begin{aligned} \mathbb{E}[g_h(Z_{i,t-1})X_t \cdot Z_{i,t-1}X_t] &= \int_{z,x} g_h(z) z x^2 dF_{Z,X}(z, x) \\ &= \int_{z,x} \left( \int_{-\infty}^z g'_h(u) du \right) z x^2 dF_{Z,X}(z, x) \\ &= \int_{\mathcal{Z}} \left( \int_{\{(z,x): z \geq u\}} z x^2 dF_{Z,X}(z, x) \right) g'_h(u) du \\ &= \int_{\mathcal{Z}} \mathbb{E}[Z_{i,t-1} X_t^2 \mathbf{1}\{Z_{i,t-1} \geq u\}] g'_h(u) du. \end{aligned}$$

An analogous argument gives

$$\mathbb{E}[g_h(Z_{i,t-1})X_t] = \int_{\mathcal{Z}} \mathbb{E}[X_t \mathbf{1}\{Z_{i,t-1} \geq u\}] g'_h(u) du,$$

and therefore

$$\mathbb{E}[g_h(Z_{i,t-1})X_t] \mathbb{E}[Z_{i,t-1}X_t] = \int_{\mathcal{Z}} \mathbb{E}[X_t \mathbf{1}\{Z_{i,t-1} \geq u\}] \mathbb{E}[Z_{i,t-1}X_t] g'_h(u) du.$$

Combining the two displays,

$$\begin{aligned}
& \text{Cov}(g_h(Z_{i,t-1})X_t, Z_{i,t-1}X_t) \\
&= \mathbb{E}[g_h(Z_{i,t-1})X_t \cdot Z_{i,t-1}X_t] - \mathbb{E}[g_h(Z_{i,t-1})X_t] \mathbb{E}[Z_{i,t-1}X_t] \\
&= \int_{\mathcal{Z}} \left[ \mathbb{E}[Z_{i,t-1}X_t^2 \mathbf{1}\{Z_{i,t-1} \geq u\}] - \mathbb{E}[X_t \mathbf{1}\{Z_{i,t-1} \geq u\}] \mathbb{E}[Z_{i,t-1}X_t] \right] g'_h(u) du \\
&= \int_{\mathcal{Z}} \text{Cov}(X_t \mathbf{1}\{Z_{i,t-1} \geq u\}, Z_{i,t-1}X_t) g'_h(u) du.
\end{aligned}$$

Substituting back and renaming the dummy variable  $u \mapsto z$  yields

$$\begin{aligned}
\beta_h &= \frac{\text{Cov}(g_h(Z_{i,t-1})X_t, Z_{i,t-1}X_t)}{\text{Var}(Z_{i,t-1}X_t)} = \int_{\mathcal{Z}} \frac{\text{Cov}(X_t \mathbf{1}\{Z_{i,t-1} \geq z\}, Z_{i,t-1}X_t)}{\text{Var}(Z_{i,t-1}X_t)} g'_h(z) dz \\
&= \int_{\mathcal{Z}} \omega_h(z) g'_h(z) dz.
\end{aligned}$$

It remains to verify the normalisation  $\int_{\mathcal{Z}} \omega_h(z) dz = 1$ . By Assumption 5(b) and Fubini's theorem,

$$\begin{aligned}
\int_{\mathcal{Z}} \omega_h(z) dz &= \frac{1}{\text{Var}(Z_{i,t-1}X_t)} \int_{\mathcal{Z}} \text{Cov}(X_t \mathbf{1}\{Z_{i,t-1} \geq z\}, Z_{i,t-1}X_t) dz \\
&= \frac{1}{\text{Var}(Z_{i,t-1}X_t)} \text{Cov}\left(\int_{\mathcal{Z}} X_t \mathbf{1}\{Z_{i,t-1} \geq z\} dz, Z_{i,t-1}X_t\right).
\end{aligned}$$

Since

$$\int_{\mathcal{Z}} X_t \mathbf{1}\{Z_{i,t-1} \geq z\} dz = X_t \int_{\mathcal{Z}} \mathbf{1}\{Z_{i,t-1} \geq z\} dz = X_t Z_{i,t-1},$$

it follows that

$$\int_{\mathcal{Z}} \omega_h(z) dz = \frac{\text{Cov}(Z_{i,t-1}X_t, Z_{i,t-1}X_t)}{\text{Var}(Z_{i,t-1}X_t)} = \frac{\text{Var}(Z_{i,t-1}X_t)}{\text{Var}(Z_{i,t-1}X_t)} = 1. \quad \square$$

*Proof of Theorem 2.* Using (34), we obtain the linearization

$$\widehat{b}_h - b_h = \widehat{A}_{11,h}^{-1} \left[ \frac{1}{NT} \sum_{t=1}^{T-h} \left( X_t \Phi(Z_{t-1})^\top - \widehat{A}_{12,h} \widehat{A}_{22,h}^{-1} W_{t-1}^\top \right) v_{h,t+h} \right],$$

where

$$v_{h,t+h} = R_h(Z_{t-1})X_t + u_{h,t+h}, \quad R_h(Z_{t-1}) := (R_h(Z_{i,t-1}))_{i=1}^N,$$

with  $R_h(z) = g_h(z) - \phi(z)^\top b_h$  the sieve approximation residual.

By the sieve approximation property in Assumption 6(1),  $\sup_{z \in \mathcal{Z}} |R_h(z)| = O(J^{-\kappa})$ . Using the per-coordinate bound  $\sup_{j,z} |\phi_j(z)| = O(1)$  for B-spline bases, the bracketed contribution from the bias component  $R_h(Z_{t-1})X_t$  is, in  $\ell^2$  norm,  $O_p(J^{1/2-\kappa})$  (with per-coordinate sup-norm rate

$O_p(J^{-\kappa})$ ). Multiplying by  $\widehat{A}_{11,h}^{-1}$  (of operator norm  $O_p(1)$ ) preserves the rate in  $\ell^2$ . Therefore,

$$\widehat{b}_h - b_h = \widehat{A}_{11,h}^{-1} \left[ \frac{1}{NT} \sum_{t=1}^{T-h} \left( X_t \Phi(Z_{t-1})^\top - \widehat{A}_{12,h} \widehat{A}_{22,h}^{-1} W_{t-1}^\top \right) u_{h,t+h} \right] + O_p(J^{1/2-\kappa}), \quad (35)$$

where the  $O_p(J^{1/2-\kappa})$  is in  $\ell^2$  norm.

Decompose the leading term into

$$\begin{aligned} \widehat{b}_h - b_h &= \underbrace{\widehat{A}_{11,h}^{-1} \left[ \frac{1}{NT} \sum_{t=1}^{T-h} X_t \Phi(Z_{t-1})^\top u_{h,t+h} \right]}_{\mathbf{I}_n} - \underbrace{\widehat{A}_{11,h}^{-1} \widehat{A}_{12,h} \widehat{A}_{22,h}^{-1} \left[ \frac{1}{NT} \sum_{t=1}^{T-h} W_{t-1}^\top u_{h,t+h} \right]}_{\mathbf{S}_n} \\ &\quad + O_p(J^{1/2-\kappa}). \end{aligned}$$

By Assumption 6(1),  $\mathbb{E}[X_t | \mathcal{F}_{t-1}] = 0$ , hence  $A_{12,h} = \mathbb{E}[X_t \phi(Z_{i,t-1}) W_{i,t-1}^\top] = 0$ . The deviation  $\widehat{A}_{12,h} - A_{12,h} = \widehat{A}_{12,h}$  is the rate of empirical estimation, with  $\|\widehat{A}_{12,h}\| = O_p(\sqrt{J/(NT)})$  by a standard second-moment argument (analogous to the variance computation in Step 1 of the proof of Theorem 4). Furthermore, by the population projection definition of  $\gamma_h$  in Assumption 6(2),

$$\mu_{Wu} := \mathbb{E}[W_{t-1}^\top u_{h,t+h}] = 0,$$

so  $\frac{1}{NT} \sum_t W_{t-1}^\top u_{h,t+h} = O_p(1/\sqrt{NT})$  alone (with no leading-order constant). Hence

$$\begin{aligned} \|\mathbf{S}_n\|_2 &\leq \|\widehat{A}_{11,h}^{-1}\| \cdot \|\widehat{A}_{12,h}\| \cdot \|\widehat{A}_{22,h}^{-1}\| \cdot \left\| \frac{1}{NT} \sum_t W_{t-1}^\top u_{h,t+h} \right\| \\ &= O_p(1) \cdot O_p(\sqrt{J/(NT)}) \cdot O_p(1) \cdot O_p(1/\sqrt{NT}) = O_p(\frac{\sqrt{J}}{NT}). \end{aligned}$$

This rate is asymptotically dominated by  $\|\mathbf{I}_n\|$  below, so the Frisch–Waugh correction  $\mathbf{S}_n$  is negligible under  $NT \rightarrow \infty$ .

Since  $A_{12,h} = 0$  and  $\|\widehat{A}_{12,h}\| = O_p(\sqrt{J/(NT)})$ ,  $\widehat{A}_{11,h} = \widehat{A}_{11,h} + O_p(J/(NT))$  in operator norm. Throughout the remainder of the proof we may therefore replace  $\widehat{A}_{11,h}^{-1}$  with  $\widehat{A}_{11,h}^{-1}$  without affecting the rates.

**Step 1: Rate for  $\widehat{b}_h - b_h$ .** We show that  $\mathbf{I}_n = O_p(\sqrt{J/(NT)})$  and that  $\mathbf{S}_n$  is asymptotically dominated by  $\mathbf{I}_n$  under the population projection definition of  $\gamma_h$ .

*Rate of  $\mathbf{I}_n$ .* We have

$$\frac{1}{NT} \sum_{t=1}^{T-h} X_t \Phi(Z_{t-1})^\top u_{h,t+h} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^{T-h} X_t \phi(Z_{i,t-1}) u_{i,t+h} \in \mathbb{R}^J.$$

For each coordinate  $j$ , expanding the variance and using i.i.d. across  $i$ , m.d.s. exogeneity of  $X_t$ ,



and absolute summability of the autocovariances of  $\{u_{i,t+h}\}_t$  in Assumption 6(2),

$$\begin{aligned}\mathbf{Var}\left(\frac{1}{NT} \sum_{i,t} X_t \phi_j(Z_{i,t-1}) u_{i,t+h}\right) &= \frac{1}{(NT)^2} \sum_{i,t,t'} \mathbb{E}[X_t X_{t'} \phi_j(Z_{i,t-1}) \phi_j(Z_{i,t'-1}) u_{i,t+h} u_{i,t'+h}] \\ &= O\left(\frac{1}{NT}\right),\end{aligned}$$

where cross-section terms with  $i \neq i'$  vanish by independence, and the cross-time sum (i.e., terms with  $t \neq t'$ ) is bounded by  $\sum_{k \in \mathbb{Z}} \|\mathbb{E}[\cdot \text{cov}(t, t-k) \cdot]\| < \infty$  under the absolutely summable autocovariance condition. Summing over the  $J$  coordinates and applying Markov's inequality,

$$\left\| \frac{1}{NT} \sum_{i,t} X_t \phi(Z_{i,t-1}) u_{i,t+h} \right\|_2 = O_p\left(\sqrt{J/(NT)}\right).$$

By Lemma 2,  $\|\widehat{A}_{11,h}^{-1}\| = \|\widehat{A}_{11,h}^{-1}\| + o_p(1) = O_p(1)$ , so

$$\|\mathbf{I}_n\|_2 = O_p\left(\sqrt{J/(NT)}\right).$$

By Assumption 6 and matrix regularity (Lemma 2 together with  $s_{\min}(A_{22,h}) > 0$ ), we have  $\|\widehat{A}_{22,h}^{-1}\| = O_p(1)$  and  $\|\widehat{A}_{11,h}^{-1}\| = O_p(1)$ .

Under m.d.s. exogeneity (Assumption 6(1)),  $A_{12,h} = 0$ , and the empirical version satisfies

$$\|\widehat{A}_{12,h}\| = O_p\left(\sqrt{J/(NT)}\right)$$

by the same second-moment argument as above (applied to the zero-mean term  $X_t \phi(Z_{i,t-1}) W_{i,t-1}^\top$ ).

By the population projection definition of  $\gamma_h$  in Assumption 6(2),

$$\mu_{Wu} := \mathbb{E}[W_{t-1}^\top u_{h,t+h}] = 0,$$

so the cross-section/time average obeys

$$\left\| \frac{1}{NT} \sum_{t=1}^{T-h} W_{t-1}^\top u_{h,t+h} \right\| = O_p(1/\sqrt{NT}),$$

with no leading-order constant. Combining,

$$\begin{aligned}\|\mathbf{S}_n\|_2 &\leq \|\widehat{A}_{11,h}^{-1}\| \cdot \|\widehat{A}_{12,h}\| \cdot \|\widehat{A}_{22,h}^{-1}\| \cdot \left\| \frac{1}{NT} \sum_t W_{t-1}^\top u_{h,t+h} \right\| \\ &= O_p(1) \cdot O_p\left(\sqrt{J/(NT)}\right) \cdot O_p(1) \cdot O_p(1/\sqrt{NT}) = O_p\left(\frac{\sqrt{J}}{NT}\right).\end{aligned}$$

Hence  $\|\mathbf{S}_n\|_2$  is asymptotically dominated by  $\|\mathbf{I}_n\|_2$  by a factor of  $1/\sqrt{NT}$ .

Adding the bias contribution  $O_p(J^{1/2-\kappa})$  in  $\ell^2$  from (35),

$$\begin{aligned}\|\hat{b}_h - b_h\|_2 &\leq \|\mathbf{I}_n\|_2 + \|\mathbf{S}_n\|_2 + O_p(J^{1/2-\kappa}) \\ &= O_p\left(\sqrt{J/(NT)}\right) + O_p(J^{1/2-\kappa}),\end{aligned}$$

where we used that  $\|\mathbf{S}_n\|_2$  is dominated. This establishes the stated coefficient rate, in  $\ell^2$  norm.

**Step 2: Pointwise CLT for  $\hat{g}_h(z)$ .** Fix  $z \in \mathcal{Z}$ . Multiply (35) by  $\sqrt{NT} \phi(z)^\top$ :

$$\begin{aligned}\sqrt{NT} \phi(z)^\top (\hat{b}_h - b_h) &= \phi(z)^\top \hat{A}_{h,11}^{-1} \left[ \frac{1}{\sqrt{NT}} \sum_{t=1}^{T-h} \left( X_t \Phi(Z_{t-1})^\top - \hat{A}_{12,h} \hat{A}_{22,h}^{-1} W_{t-1}^\top \right) u_{h,t+h} \right] \\ &\quad + \sqrt{NT} O_p(J^{1/2-\kappa}).\end{aligned}$$

Under the undersmoothing condition  $\sqrt{NT} J^{1/2-\kappa} \rightarrow 0$  (consistent with hypothesis (iv) of Theorem 4), the last term divided by  $\sigma_h^*(z) \asymp \sqrt{J}$  is  $o_p(1)$ .

By Step 1,  $\|\hat{A}_{12,h}\| = O_p(\sqrt{J/(NT)})$  and  $\mu_{Wu} = \mathbb{E}[W_{t-1}^\top u_{h,t+h}] = 0$  by the population projection definition of  $\gamma_h$  in Assumption 6(2). Hence

$$\begin{aligned}\left\| \hat{A}_{12,h} \hat{A}_{22,h}^{-1} \cdot \frac{1}{\sqrt{NT}} \sum_t W_{t-1}^\top u_{h,t+h} \right\| &\leq \|\hat{A}_{12,h}\| \cdot \|\hat{A}_{22,h}^{-1}\| \cdot \left\| \frac{1}{\sqrt{NT}} \sum_t W_{t-1}^\top u_{h,t+h} \right\| \\ &= O_p(\sqrt{J/(NT)}) \cdot O_p(1) \cdot O_p(1) = o_p(1)\end{aligned}$$

(using  $\mu_{Wu} = 0$  to bound the partial-sum factor by  $O_p(1)$  rather than  $O_p(\sqrt{NT} \mu_{Wu} + 1)$ ). Thus the Frisch–Waugh correction term is asymptotically negligible, and the leading stochastic term reduces to

$$\mathcal{S}_{1,n} := \frac{1}{\sqrt{NT}} \sum_{t=1}^{T-h} X_t \Phi(Z_{t-1})^\top u_{h,t+h}$$

plus an  $o_p(1)$  remainder.

Define the time-stationary process

$$\tilde{S}_{t,j}(z) := \phi(z)^\top A_{11,h}^{-1} \cdot \frac{1}{\sqrt{N}} \sum_{i=1}^N X_t \phi_j(Z_{i,t-1}) u_{i,t+h} = \tilde{\phi}(z)^\top \cdot \frac{1}{\sqrt{N}} \sum_{i=1}^N X_t \phi_j(Z_{i,t-1}) u_{i,t+h},$$

where  $\tilde{\phi}(z) := A_{11,h}^{-1} \phi(z)$ . By Assumption 6(1)–(3),  $\{\tilde{S}_{t,\cdot}(z)\}_t$  is strictly stationary in  $t$  with mean zero (by m.d.s. exogeneity) and finite second moments. The absolute summability of the MA coefficients  $\sum_l \|a_{h,l}\| < \infty$  in Assumption 6(2) implies absolute summability of the autocovariances of  $\{\tilde{S}_{t,\cdot}(z)\}_t$ , so the long-run variance

$$\Sigma^*(z) := \sum_{k \in \mathbb{Z}} \mathbb{E}[\tilde{S}_{t,\cdot}(z) \tilde{S}_{t-k,\cdot}(z)^\top] = \phi(z)^\top A_{11,h}^{-1} \Omega_h^* A_{11,h}^{-1} \phi(z) = \sigma_h^{*2}(z)$$

is well-defined and bounded above and below by  $\lambda_{\min/\max}(\Omega_h^*) \cdot \|\tilde{\phi}(z)\|^2$ ; in particular,  $\sigma_h^*(z) \asymp \sqrt{J}$ .

By the central limit theorem for stationary processes with absolutely summable autocovariances (Ibragimov and Linnik 1971, Theorem 18.6.5),

$$\frac{1}{\sigma_h^*(z)} \cdot \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \tilde{S}_{t,\cdot}(z) \Rightarrow \mathcal{N}(0, 1).$$

By Lemma 2,  $\hat{A}_{11,h}^{-1} \rightarrow_p A_{11,h}^{-1}$ . Combined with the CLT above and the  $o_p(1)$  remainder from the Frisch–Waugh reduction,

$$\frac{\sqrt{NT} \phi(z)^\top (\hat{b}_h - b_h)}{\sigma_h^*(z)} = \frac{1}{\sigma_h^*(z)} \cdot \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \tilde{S}_{t,\cdot}(z) + o_p(1) \Rightarrow \mathcal{N}(0, 1).$$

The estimation error decomposes as

$$\hat{g}_h(z) - g_h(z) = \phi(z)^\top (\hat{b}_h - b_h) - R_h(z), \quad \sup_z |R_h(z)| = O(J^{-\kappa}).$$

Hence

$$\frac{\sqrt{NT}(\hat{g}_h(z) - g_h(z))}{\sigma_h^*(z)} = \frac{\sqrt{NT} \phi(z)^\top (\hat{b}_h - b_h)}{\sigma_h^*(z)} - \frac{\sqrt{NT} R_h(z)}{\sigma_h^*(z)}.$$

The bias term satisfies  $|\sqrt{NT} R_h(z)/\sigma_h^*(z)| \leq \sqrt{NT} \cdot O(J^{-\kappa})/\sqrt{J} = O(\sqrt{NT} J^{-(\kappa+1/2)})$ , which is  $o(1)$  under the undersmoothing condition  $\sqrt{NT} J^{1/2-\kappa} \rightarrow 0$  (since  $J^{1/2-\kappa} \geq J^{-(\kappa+1/2)}$  for  $J \geq 1$ ). Therefore

$$\frac{\sqrt{NT}(\hat{g}_h(z) - g_h(z))}{\sigma_h^*(z)} \Rightarrow \mathcal{N}(0, 1),$$

which establishes the pointwise CLT.  $\square$

*Proof of Theorem 3.* We split the argument into three steps.

**Step 1: Preliminary limits and matrix regularity.** By the martingale difference restriction  $\mathbb{E}[X_t | \mathcal{F}_{t-1}] = 0$  (Assumption 6(1)) and the  $\mathcal{F}_{t-1}$ -measurability of  $\phi(Z_{i,t-1})$  and  $W_{i,t-1}$ ,

$$A_{12,h} = \mathbb{E}[X_t \phi(Z_{i,t-1}) W_{i,t-1}^\top] = \mathbb{E}[\mathbb{E}(X_t | \mathcal{F}_{t-1}) \phi(Z_{i,t-1}) W_{i,t-1}^\top] = 0.$$

Hence, by definition,  $\tilde{A}_{11,h} = A_{11,h} - A_{12,h} A_{22,h}^{-1} A_{21,h} = A_{11,h}$ , which justifies the use of  $A_{11,h}^{-1}$  in  $V_h^*$  in the theorem statement.

By Assumption 6(4),  $A_{11,h}$  is well-conditioned and  $\lambda_{\min}(A_{22,h}) > 0$ . By Lemma 2 (with the moment conditions in Assumption 6(1)–(3)),

$$\|\hat{A}_{11,h}^{-1}\| = O_p(1).$$

By the LLN for stationary processes with finite second moments (Assumption 6(2)–(3)),  $\hat{A}_{22,h} \rightarrow_p A_{22,h}$ , hence

$$\|\hat{A}_{22,h}^{-1}\| = O_p(1).$$

Furthermore, the empirical moment  $\widehat{A}_{12,h}$  obeys

$$\|\widehat{A}_{12,h}\| = O_p(\sqrt{J/(NT)}),$$

by the same second-moment argument as in Step 1 of the proof of Theorem 2 (using  $A_{12,h} = 0$  and the variance computation under m.d.s. exogeneity and i.i.d. cross-section).

**Step 2: Consistency of the long-run variance for the effective score.** Define the infeasible effective score (matching the convention in Theorem 2, with  $u_{h,t+h}$  the population residual rather than the OLS residual)

$$s_{h,t}^* := X_t \Phi(Z_{t-1})^\top u_{h,t+h} - q_{h,t} A_{22,h}^{-1} \mu_{Wu}, \quad q_{h,t} := X_t \Phi(Z_{t-1})^\top W_{t-1}, \quad \mu_{Wu} := \mathbb{E}[W_{t-1}^\top u_{h,t+h}].$$

By the population projection definition of  $\gamma_h$  in Assumption 6(2),  $\mu_{Wu} = 0$  identically, so the Frisch–Waugh correction term vanishes and

$$s_{h,t}^* = X_t \Phi(Z_{t-1})^\top u_{h,t+h},$$

matching the convention in Theorem 2. (We retain the correction in the empirical effective score  $\widehat{s}_{h,t}^*$  to make the OLS first-order condition  $\widehat{\mu}_{W\widehat{v}} = 0$  explicit; in finite sample, the empirical correction vanishes identically.)

Under Assumption 6(2),  $\{u_{h,t+h}\}_t$  is a linear MA process with absolutely summable impulse responses ( $\sum_l \|a_{h,l}\| < \infty$ ). Combined with the m.d.s. exogeneity of  $X_t$  in Assumption 6(1), the moment bounds in Assumption 6(2)–(3), and i.i.d. across  $i$ , the score process  $\{s_{h,t}^*\}_t$  is strictly stationary with

$$\mathbb{E}\|s_{h,t}^*\|^2 < \infty, \quad \sum_{k \in \mathbb{Z}} \|\mathbb{E}[s_{h,t}^* s_{h,t-k}^{*\top}]\| < \infty.$$

Hence the long-run covariance

$$\Omega_h^* = \frac{1}{N} \sum_{k \in \mathbb{Z}} \mathbb{E}[s_{h,t}^* s_{h,t-k}^{*\top}]$$

exists and is finite, and is the same as the long-run covariance defined in Theorem 2.

Let  $\widehat{\Omega}_h^{*,\text{inf}}$  denote the Bartlett HAC estimator constructed from  $\{s_{h,t}^*\}_{t=1}^{T-h}$ :

$$\begin{aligned} \widehat{\Omega}_h^{*,\text{inf}} &:= \frac{1}{NT} \sum_{t=1}^{T-h} s_{h,t}^* s_{h,t}^{*\top} \\ &\quad + \sum_{k=1}^L w_k \left( \frac{1}{NT} \sum_{t=k+1}^{T-h} (s_{h,t}^* s_{h,t-k}^{*\top} + s_{h,t-k}^* s_{h,t}^{*\top}) \right), \end{aligned}$$

with Bartlett weights  $w_k = 1 - k/(L+1)$ .

By standard HAC consistency arguments for stationary short-memory processes (Andrews, 1991; Newey and West, 1987), applied here in the sieve setting with growing dimension  $J$  (cf. Chen

and Christensen 2015),

$$\|\widehat{\Omega}_h^{\star, \text{inf}} - \Omega_h^{\star}\| \xrightarrow{p} 0,$$

provided the truncation lag  $L = L_{NT}$  satisfies the conditions in the theorem statement:  $L \rightarrow \infty$ ,  $L/T \rightarrow 0$ , and  $J^2/T + J^2L/(NT) \rightarrow 0$ . The last two conditions handle the growing sieve dimension by ensuring the variance of each entry of  $\widehat{\Omega}_h^{\star, \text{inf}}$  is controlled, and the truncation-bias conditions are absorbed by absolute summability of autocovariances under Assumption 6(2).

**Step 3: Feasible scores and the sandwich mapping.** Recall the population effective score (Step 2) and the empirical counterpart from the theorem statement:

$$s_{h,t}^{\star} = X_t \Phi(Z_{t-1})^{\top} u_{h,t+h}, \quad \widehat{s}_{h,t}^{\star} = X_t \Phi(Z_{t-1})^{\top} \widehat{v}_{h,t+h} - \widehat{q}_{h,t} \widehat{A}_{22,h}^{-1} \widehat{\mu}_{W\widehat{v}}.$$

By the OLS first-order conditions of the regression of  $Y_{t+h}$  on  $D_t = (\Phi(Z_{t-1})X_t, W_{t-1})$ ,

$$\widehat{\mu}_{W\widehat{v}} = \frac{1}{NT} \sum_{t=1}^{T-h} W_{t-1}^{\top} \widehat{v}_{h,t+h} = 0$$

identically in finite sample, so the empirical Frisch–Waugh correction term vanishes and

$$\widehat{s}_{h,t}^{\star} = X_t \Phi(Z_{t-1})^{\top} \widehat{v}_{h,t+h}.$$

Using  $\widehat{v}_{h,t+h} - u_{h,t+h} = (\widehat{v}_{h,t+h} - v_{h,t+h}) + (v_{h,t+h} - u_{h,t+h})$  with  $v_{h,t+h} - u_{h,t+h} = X_t R_h(Z_{t-1})$  (where  $R_h(Z_{t-1}) := (R_h(Z_{i,t-1}))_{i=1}^N$ ) and  $\widehat{v}_{h,t+h} - v_{h,t+h} = -D_t(\widehat{\theta}_h - \theta_h)$ , we obtain

$$\widehat{s}_{h,t}^{\star} - s_{h,t}^{\star} = \underbrace{X_t^2 \Phi(Z_{t-1})^{\top} R_h(Z_{t-1})}_{\mathbf{A}_t^{(\text{bias})}} - \underbrace{X_t \Phi(Z_{t-1})^{\top} D_t(\widehat{\theta}_h - \theta_h)}_{\mathbf{A}_t^{(\text{est})}}.$$

*Bound on the bias contribution  $\mathbf{A}_t^{(\text{bias})}$ .* For each coordinate  $j$ ,

$$A_{t,j}^{(\text{bias})} = X_t^2 \sum_{i=1}^N \phi_j(Z_{i,t-1}) R_h(Z_{i,t-1}).$$

By i.i.d. across  $i$  and finite second moments,

$$\mathbb{E}[(A_{t,j}^{(\text{bias})})^2] = \mathbb{E}[X_t^4] \cdot \mathbb{E}\left[\left(\sum_i \phi_j(Z_i) R_h(Z_i)\right)^2\right] = O(N \cdot \mathbb{E}[\phi_j(Z)^2 R_h(Z)^2] + N^2 \cdot \mathbb{E}[\phi_j(Z) R_h(Z)]^2).$$

For B-spline bases,  $\mathbb{E}[\phi_j(Z) R_h(Z)]^2 = O(J^{-2\kappa}/J^2)$  (by sieve localization), and  $\mathbb{E}[\phi_j(Z)^2 R_h(Z)^2] = O(J^{-2\kappa})$ . Hence  $\mathbb{E}[(A_{t,j}^{(\text{bias})})^2] = O(NJ^{-2\kappa} + N^2 J^{-2\kappa-2}) = O(NJ^{-2\kappa})$ .

Summing over  $j$  and averaging over  $t$ :

$$\frac{1}{NT} \sum_{t=1}^{T-h} \|\mathbf{A}_t^{(\text{bias})}\|_2^2 = \frac{1}{NT} \sum_t \sum_j (A_{t,j}^{(\text{bias})})^2.$$

By the per-coordinate bound  $\mathbb{E}[(A_{t,j}^{(\text{bias})})^2] = O(NJ^{-2\kappa})$  established above, summing over  $J$  coordinates gives  $\mathbb{E} \|\mathbf{A}_t^{(\text{bias})}\|_2^2 = O(NJ^{1-2\kappa})$ , hence

$$\frac{1}{NT} \sum_{t=1}^{T-h} \mathbb{E} \|\mathbf{A}_t^{(\text{bias})}\|_2^2 = \frac{1}{NT} \cdot T \cdot O(NJ^{1-2\kappa}) = O(J^{1-2\kappa}).$$

By Markov's inequality,

$$\frac{1}{NT} \sum_{t=1}^{T-h} \|\mathbf{A}_t^{(\text{bias})}\|_2^2 = O_p(J^{1-2\kappa}),$$

which is  $o_p(1)$  under  $\kappa > 1/2$  (Assumption 6(1)).

Setting  $v := \hat{\theta}_h - \theta_h$ ,

$$\frac{1}{NT} \sum_{t=1}^{T-h} \|\mathbf{A}_t^{(\text{est})}\|_2^2 = v^\top \left[ \frac{1}{NT} \sum_{t=1}^{T-h} X_t^2 D_t^\top \Phi(Z_{t-1}) \Phi(Z_{t-1})^\top D_t \right] v.$$

By the LLN under Assumption 6(1)–(5), the empirical second-moment matrix in brackets converges in probability to its expectation, which is of operator norm  $O(1)$  (combining moment regularity, the population  $A_{11,h}$ , and bounded fourth moments of  $X_t$ ). Hence

$$\frac{1}{NT} \sum_{t=1}^{T-h} \|\mathbf{A}_t^{(\text{est})}\|_2^2 \leq O_p(1) \cdot \|v\|_2^2 = O_p(\|\hat{\theta}_h - \theta_h\|_2^2).$$

Using the  $\ell^2$  rates from Theorem 2,

$$\|\hat{\theta}_h - \theta_h\|_2^2 = \|\hat{b}_h - b_h\|_2^2 + \|\hat{\gamma}_h - \gamma_h\|_2^2 = O_p(J/(NT) + J^{1-2\kappa}) + O_p(1/(NT)) = O_p(J/(NT) + J^{1-2\kappa}).$$

This is  $o_p(1)$  under Assumption 6(6) ( $J = o(NT)$ ) and  $\kappa > 1/2$  (which gives  $J^{1-2\kappa} \rightarrow 0$ ).

$\frac{1}{NT} \sum_t \|\hat{s}_{h,t}^\star - s_{h,t}^\star\|_2^2 = o_p(1)$ . By Cauchy–Schwarz, each cross-lag term in the HAC estimator satisfies

$$\begin{aligned} & \frac{1}{NT} \sum_{t=k+1}^{T-h} \|(\hat{s}_{h,t}^\star - s_{h,t}^\star) s_{h,t-k}^{\star\top}\| \\ & \leq \left( \frac{1}{NT} \sum_t \|\hat{s}_{h,t}^\star - s_{h,t}^\star\|_2^2 \right)^{1/2} \cdot \left( \frac{1}{NT} \sum_t \|s_{h,t}^\star\|_2^2 \right)^{1/2} = o_p(1), \end{aligned}$$

using  $\mathbb{E} \|s_{h,t}^*\|_2^2 < \infty$ . Summing over the  $L$  lags with Bartlett weights and using  $J^2 L / (NT) \rightarrow 0$ ,

$$\|\widehat{\Omega}_h^* - \widehat{\Omega}_h^{\star, \text{inf}}\| \xrightarrow{p} 0.$$

Combining with Step 2,  $\|\widehat{\Omega}_h^* - \Omega_h^*\| \xrightarrow{p} 0$ .

By Lemma 2 and the equivalence  $\widehat{\widehat{A}}_{11,h} = \widehat{A}_{11,h} + O_p(J/(NT))$  from the preamble,  $\widehat{A}_{11,h} \rightarrow_p A_{11,h}$  with  $\|\widehat{A}_{11,h}^{-1}\| = O_p(1)$ . The continuous mapping theorem yields

$$\|\widehat{V}_h^* - V_h^*\| = \left\| \widehat{\widehat{A}}_{11,h}^{-1} \widehat{\Omega}_h^* \widehat{\widehat{A}}_{11,h}^{-1} - A_{11,h}^{-1} \Omega_h^* A_{11,h}^{-1} \right\| = o_p(1).$$

By Cauchy-Schwarz and  $\sup_{z \in \mathcal{Z}} \|\phi(z)\| = O(\sqrt{J})$ ,

$$|\phi(z)^\top \widehat{V}_h^* \phi(z) - \phi(z)^\top V_h^* \phi(z)| \leq \|\phi(z)\|^2 \cdot \|\widehat{V}_h^* - V_h^*\| = O(J) \cdot o_p(1) = o_p(J),$$

which establishes the pointwise claim in the theorem.  $\square$

*Proof of Lemma 2.* Under the conditional homoskedasticity assumption,  $\mathbb{E}[X_t^2 \mid \mathcal{F}_{t-1}] = \sigma_X^2$ , and  $\phi(Z_{i,t-1})$  is  $\mathcal{F}_{t-1}$ -measurable (Assumption 6(2)). Iterating expectations,

$$\mathbb{E}[X_t^2 \phi(Z_{i,t-1}) \phi(Z_{i,t-1})^\top] = \mathbb{E}\left[\mathbb{E}[X_t^2 \mid \mathcal{F}_{t-1}] \phi(Z_{i,t-1}) \phi(Z_{i,t-1})^\top\right] = \sigma_X^2 \mathbb{E}[\phi \phi^\top] = \mathbb{E}[X_t^2] \mathbb{E}[\phi \phi^\top].$$

This justifies the decomposition

$$\begin{aligned} \widehat{A}_{11,h} - A_{11,h} &= \underbrace{\frac{1}{T-h} \sum_{t=1}^{T-h} (X_t^2 - \mathbb{E}[X_t^2]) \mathbb{E}[\phi \phi^\top]}_{\Delta_T} \\ &+ \underbrace{\frac{1}{N(T-h)} \sum_{i=1}^N \sum_{t=1}^{T-h} X_t^2 (\phi(Z_{i,t-1}) \phi(Z_{i,t-1})^\top - \mathbb{E}[\phi \phi^\top])}_{\Delta_{N,T}}. \end{aligned}$$

Assumption 6(4) gives  $\|\mathbb{E}[\phi \phi^\top]\| \leq C_2$ . The sequence  $\{X_t^2\}$  is strictly stationary with  $\mathbb{E}[X_t^4] < \infty$  by Assumptions 6(1) and (3). A standard law of large numbers yields  $T^{-1} \sum_{t=1}^{T-h} (X_t^2 - \mathbb{E}[X_t^2]) = O_p(T^{-1/2})$ , hence  $\|\Delta_T\| = O_p(T^{-1/2})$ .

Set  $f_{it} := X_t^2 [\phi(Z_{i,t-1}) \phi(Z_{i,t-1})^\top - \mathbb{E}[\phi \phi^\top]]$ . Each  $f_{it}$  has mean zero by conditional homoskedasticity and the factorization above, and

$$\|f_{it}\| \leq X_t^2 (\|\phi(Z_{i,t-1})\|^2 + \|\mathbb{E}[\phi \phi^\top]\|) \leq (C + C_2) X_t^2 J$$

by Assumption 6(5). For the matrix variance, the bounded conditional fourth-moment assumption

$\mathbb{E}[X_t^4 \mid \mathcal{F}_{t-1}] \leq C_4$  a.s. gives

$$\mathbb{E}[f_{it}f_{it}^\top] \preceq \mathbb{E}\left[\mathbb{E}[X_t^4 \mid \mathcal{F}_{t-1}] (\phi\phi^\top - \mathbb{E}[\phi\phi^\top])^2\right]$$

$$\preceq 2C_4 \mathbb{E}\left[\|\phi\|^2 \phi\phi^\top\right] + 2C_4 (\mathbb{E}[\phi\phi^\top])^2 \preceq O(J) \mathbb{E}[\phi\phi^\top] + O(1) I_J \preceq O(J) I_J,$$

using  $(\phi\phi^\top - \mathbb{E}[\phi\phi^\top])^2 \preceq 2(\phi\phi^\top)^2 + 2(\mathbb{E}[\phi\phi^\top])^2$ ,  $(\phi\phi^\top)^2 = \|\phi\|^2 \phi\phi^\top$ ,  $\|\phi\|^2 \leq CJ$  by Assumption 6(5), and  $\|\mathbb{E}[\phi\phi^\top]\| \leq C_2$  by Assumption 6(4).

Since  $\{f_{it}\}$  are i.i.d. across  $i$  and strictly stationary across  $t$ , cross-sectional averaging at fixed  $t$  yields  $\bar{f}_t := N^{-1} \sum_{i=1}^N f_{it}$  with  $\mathbb{E}[\bar{f}_t] = 0$  and matrix variance parameter  $O(J/N)$ . Applying matrix concentration for weakly dependent random matrices in the sieve setting (cf. Chen and Christensen (2015)) to the time-series  $\{\bar{f}_t\}_{t=1}^{T-h}$ ,

$$\|\Delta_{N,T}\| = O_p\left(\sqrt{\frac{J}{NT}} + \frac{J}{NT}\right),$$

where the rate is up to logarithmic factors absorbed under standard sieve growth conditions (Assumption 6(6)).

Combining the two bounds establishes the stated rate. The final claim follows from Weyl's inequality: under  $\lambda_{\min}(A_{11,h}) \geq c_1 > 0$ ,

$$\lambda_{\min}(\hat{A}_{11,h}) \geq \lambda_{\min}(A_{11,h}) - \|\hat{A}_{11,h} - A_{11,h}\| \geq c_1 - o_p(1) \geq \frac{c_1}{2}$$

with probability approaching one. □

**Corollary 2** (Higher-order rate for  $\hat{b}_h$ ). *Fix  $h \in \{0, 1, \dots, H\}$ . Suppose the assumptions of Lemma 2 hold and  $\lambda_{\min}(A_{11,h}) \geq c > 0$ . Then*

$$\|\hat{b}_h - b_h\|_2 = O_p(J^{1/2-\kappa}) + O_p\left(\sqrt{\frac{J}{NT}}\right) + O_p\left(\frac{J^{1/2-\kappa}}{T^{1/2}}\right) + O_p\left(\frac{1}{T^{1/2}}\sqrt{\frac{J}{NT}}\right),$$

where the last two terms are higher order relative to the first two under Assumption 6(6).

*Proof of Corollary 2.* By Assumption 6(1) and (2),  $A_{12,h} = \mathbb{E}[X_t \phi(Z_{i,t-1}) W_{i,t-1}^\top] = 0$ , hence we have  $\tilde{A}_{11,h} = A_{11,h}$ . The sample remainder  $\hat{\tilde{A}}_{11,h} - \hat{A}_{11,h} = -\hat{A}_{12,h} \hat{A}_{22,h}^{-1} \hat{A}_{21,h} = O_p(J/(NT))$ , which is dominated by the rate of  $R_{A,n}$  derived below and absorbed into it. Lemma 2 and the condition  $\lambda_{\min}(A_{11,h}) \geq c > 0$  yield the expansion

$$\hat{A}_{11,h}^{-1} = A_{11,h}^{-1} + R_{A,n}, \quad \|R_{A,n}\| = O_p\left(T^{-1/2} + \sqrt{\frac{J}{NT}} + \frac{J}{NT}\right), \quad \|A_{11,h}^{-1}\| = O(1).$$



From the normal equations applied to the sieve approximation (31),

$$\hat{b}_h - b_h = \hat{A}_{11,h}^{-1} (\mathbf{S}_{h,n}^{(\text{bias})} + \mathbf{S}_{h,n}^{(u)}),$$

where the score components are

$$\mathbf{S}_{h,n}^{(\text{bias})} := \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^{T-h} X_t^2 \phi(Z_{i,t-1}) R_h(Z_{i,t-1}), \quad \mathbf{S}_{h,n}^{(u)} := \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^{T-h} X_t \phi(Z_{i,t-1}) u_{i,t+h}.$$

The first term arises from the sieve approximation error  $R_h(z) = g_h(z) - \phi(z)^\top b_h$  and satisfies  $\|\mathbf{S}_{h,n}^{(\text{bias})}\|_2 = O_p(J^{1/2-\kappa})$  in  $\ell^2$  norm under Assumption 6 and the smoothness condition on  $g_h$ . The second term is the score from the LP error and satisfies  $\|\mathbf{S}_{h,n}^{(u)}\|_2 = O_p(\sqrt{J/(NT)})$  by the score bound established in the proof of Theorem 2.

Applying the triangle inequality,

$$\|\hat{b}_h - b_h\|_2 \leq \|A_{11,h}^{-1}\| (\|\mathbf{S}_{h,n}^{(\text{bias})}\|_2 + \|\mathbf{S}_{h,n}^{(u)}\|_2) + \|R_{A,n}\| (\|\mathbf{S}_{h,n}^{(\text{bias})}\|_2 + \|\mathbf{S}_{h,n}^{(u)}\|_2).$$

The first bracket gives the leading two terms,  $O_p(J^{1/2-\kappa})$  and  $O_p(\sqrt{J/(NT)})$ . Expanding the second bracket using  $\|R_{A,n}\| = O_p(T^{-1/2} + \sqrt{J/(NT)} + J/(NT))$  yields six cross-products. The two arising from the  $T^{-1/2}$  component of  $\|R_{A,n}\|$  give the third and fourth terms,  $O_p(J^{1/2-\kappa}/T^{1/2})$  and  $O_p(T^{-1/2}\sqrt{J/(NT)})$ . The remaining four cross-products are  $O_p(J^{1-\kappa}/\sqrt{NT})$ ,  $O_p(J/(NT))$ ,  $O_p(J^{3/2-\kappa}/(NT))$ , and  $O_p((J/(NT))^{3/2})$ , all of which are  $o_p(J^{1/2-\kappa}) + o_p(\sqrt{J/(NT)})$  under Assumption 6(6). Combining yields the four-term rate stated above.  $\square$

**Theorem 5** (Gaussian Approximation for High-Dimensional Stationary Time Series). *Let  $\{X_t\}_{t=1}^n$  be a  $p$ -dimensional mean-zero strictly stationary time series with elements  $X_{t,j}$ ,  $1 \leq j \leq p$ . Let  $m$  and  $M$  be positive integers with  $m \ll M$ , and set  $L := \lfloor \frac{n}{m+M} \rfloor$ . Suppose the following hold for some  $q > 4$  and  $\alpha > 1/2 - 1/q$ .*

- (i) (**Moments.**)  $\sup_{t,j} \mathbb{E}|X_{t,j}|^q < \infty$ .
- (ii) (**Functional dependence.**) Write  $X_{t,j,\{0\}}$  for the coupled series obtained by replacing  $\varepsilon_0$  with an i.i.d. copy  $\varepsilon_0^*$ , and let  $\delta_{i,r,j} := \|X_{i,j} - X_{i,j,\{0\}}\|_r$ . The functional-dependence-adjusted norm

$$\Psi_{r,\alpha} := \max_j \sup_{m \geq 0} (m+1)^\alpha \sum_{i \geq m} \delta_{i,r,j}$$

is assumed finite for  $r \in \{2, q\}$ , and satisfies  $\Theta_{q,\alpha} \lesssim p^{1/q} \Psi_{q,\alpha}$ .

- (iii) (**Long-run covariance.**) The matrix

$$\Sigma := \sum_{l \in \mathbb{Z}} \mathbb{E}[X_0 X_l^\top] = \mathbb{E}[X_0 X_0^\top] + \sum_{l=1}^{\infty} (\mathbb{E}[X_0 X_l^\top] + \mathbb{E}[X_l X_0^\top])$$

exists (with the series converging absolutely under condition (ii)) and satisfies  $0 < c \leq \lambda_{\min}(\Sigma)$  and  $\lambda_{\max}(\Sigma) \leq C < \infty$ .

(iv) (**Block-size conditions.**)

$$(\log p)^9 \ll L, \quad M \gg m^2 \vee m(\log p)^3, \quad m^\alpha \gg \log p, \quad M^{\alpha \wedge 1} \gg (\log p)^2.$$

(v) (**Moment exponent condition.**) The moment exponent  $q$  is taken sufficiently large that

$$p^{1/q} = o\left(L^{1/2-1/q}(\log p)^{-(1-1/q)}\right) \quad \text{and} \quad p^{1/q} = o\left(m^{\alpha-1/2+1/q} n^{1/2-1/q}\right).$$

Then, on a suitably enriched probability space, there exists  $Z \sim \mathcal{N}(0, \Sigma)$  such that

$$\mathbb{P}\left(\left\|\frac{1}{\sqrt{n}} \sum_{t=1}^n X_t - Z\right\|_\infty \geq \delta(p, L, m, q, \alpha)\right) \rightarrow 0,$$

where

$$\delta(p, L, m, q, \alpha) \lesssim L^{-1/6}(\log p)^{7/6} \vee m^{-\alpha} \log p \vee \sqrt{m/M} \log p \vee v(M)^{1/3}(\log p)^{2/3}, \quad (36)$$

with  $v(M) = M^{-1}$  for  $\alpha > 1$ ,  $(\log M)/M$  for  $\alpha = 1$ , and  $M^{-\alpha}$  for  $0 < \alpha < 1$ .

*Remark 4* (Discussion of rates). Under conditions (iv)–(v), the approximation error  $\delta(p, L, m, q, \alpha)$  in (36) reduces to four terms, each reflecting a distinct source of error in the blocking-and-coupling construction:

$$\delta \lesssim \underbrace{L^{-1/6}(\log p)^{7/6}}_{\text{(i) Gaussian approximation}} \vee \underbrace{m^{-\alpha} \log p}_{\text{(ii) small-block truncation}} \vee \underbrace{\sqrt{m/M} \log p}_{\text{(iii) small-block sum}} \vee \underbrace{v(M)^{1/3}(\log p)^{2/3}}_{\text{(iv) covariance matching}}.$$

Term (i) is the high-dimensional Gaussian approximation rate of Chernozhukov et al. (2017) for i.i.d. observations, with the full sample size  $n$  replaced by the effective block count  $L$ . The reduction from  $n$  to  $L = \lfloor n/(m+M) \rfloor$  reflects the effective sample size of the blocking scheme: Gaussian approximation is applied to the  $L$  approximately independent big-block sums rather than to the  $n$  original observations. Term (ii) arises from the  $m$ -dependent truncation of the original process and decreases as  $m$  grows under  $\alpha > 0$ . Term (iii) is the cost of removing the small blocks and vanishes under the block-ratio condition  $M \gg m^2$  in (iv). Term (iv) is the Gaussian-comparison cost of replacing the finite-sample covariance  $\Sigma_n$  by the limiting long-run covariance  $\Sigma$ , controlled by Zhang–Wu’s covariance error rate  $v(M)$ .

*Worked example:*  $L = \lfloor n^{2/3} \rfloor$ . Set  $L = \lfloor n^{2/3} \rfloor$ , implying  $m+M \asymp n^{1/3}$  and hence  $M \asymp n^{1/3}$  when  $m \ll M$ . Under condition (v), the four terms become

$$\delta \asymp n^{-1/9}(\log p)^{7/6} \vee m^{-\alpha} \log p \vee \sqrt{m/n^{1/3}} \log p \vee v(n^{1/3})^{1/3}(\log p)^{2/3}.$$

Balancing the small-block truncation (ii) and the Gaussian approximation (i) requires  $m^{-\alpha} \log p \lesssim n^{-1/9} (\log p)^{7/6}$ , i.e.,  $m \gg n^{1/(9\alpha)} (\log p)^{-1/(6\alpha)}$ . Combined with the block-size constraint  $M \gg m^2$  in (iv) (which gives  $m \ll n^{1/6}$ ), the admissible range is

$$n^{1/(9\alpha)} (\log p)^{-1/(6\alpha)} \ll m \ll n^{1/6},$$

which is non-vacuous whenever  $\alpha > 2/3$ . The covariance-matching term (iv) is dominated by (i) when  $v(n^{1/3}) \lesssim n^{-1/3} (\log p)^{3/2}$ ; for  $\alpha \geq 1$  this gives  $n^{-1/3} \lesssim n^{-1/3} (\log p)^{3/2}$ , automatic, and for  $\alpha < 1$  it gives  $n^{-\alpha/3} \lesssim n^{-1/3} (\log p)^{3/2}$ , i.e.,  $\alpha \geq 1$  or  $\log p$  moderate. Over the admissible range,

$$\delta \lesssim n^{-1/9} (\log p)^{7/6}.$$

*Effect of weak dependence.* When  $\{X_t\}$  is strongly weakly dependent (large  $\alpha$ ), the small-block truncation term (ii) becomes negligible for relatively small  $m$ , and the dominant contributions are from the Gaussian approximation (i) and the small-block sum (iii):

$$\delta \asymp L^{-1/6} (\log p)^{7/6} \vee \sqrt{m/M} \log p,$$

both of which vanish under conditions (iv) (in particular,  $M \gg m^2$ ). The covariance-matching term (iv) further sharpens to  $v(M) = M^{-1}$  for  $\alpha > 1$ , and is dominated by the other terms in this regime.

*Proof of Theorem 5.* The proof proceeds in three steps. Let  $\mathcal{B}_b := [(b-1)(M+m)+1, bM+(b-1)m]$  and  $\mathcal{S}_b := [bM+(b-1)m+1, b(M+m)]$  denote the  $b$ th big block and small block respectively, for  $b = 1, \dots, L$ . Define the  $m$ -dependent approximation of  $X_t$  by  $X_{t,m} := \mathbb{E}(X_t \mid \varepsilon_{t-m}, \dots, \varepsilon_t)$ , and set

$$Y_b := \sum_{t \in \mathcal{B}_b} X_t, \quad \tilde{Y}_b := \sum_{t \in \mathcal{B}_b} X_{t,m}, \quad b = 1, \dots, L.$$

**Step 1:  $m$ -dependence approximation and small-block removal.** We control the gap  $T_X - T_{Y,m}$  in a single stroke, where

$$T_X := \sum_{t=1}^n X_t, \quad T_{Y,m} := \sum_{b=1}^L \tilde{Y}_b = \sum_{b=1}^L \sum_{t \in \mathcal{B}_b} X_{t,m}.$$

This gap encodes both effects at once: replacing big-block summands by their  $m$ -dependent versions, and discarding the small-block observations.

Under the standing hypotheses  $q > 4 > 2$  and  $\alpha > 1/2 - 1/q$ , Lemma 7.1(i) of Zhang and Wu (2017) yields, for every  $y > 0$ ,

$$\mathbb{P}(\|T_X - T_{Y,m}\|_\infty \geq y) \lesssim f_1^*(y) + f_2^*(y),$$

with

$$f_1^*(y) = \frac{n m^{q/2-1-q\alpha} \Pi_{q,\alpha}^q}{y^q} + p \exp\left(-\frac{c y^2 m^{2\alpha}}{n \Psi_{2,\alpha}^2}\right),$$

$$f_2^*(y) = \frac{L m^q \Pi_{q,\alpha}^q}{y^q} + p \exp\left(-\frac{c y^2}{L m \Psi_{2,\alpha}^2}\right),$$

where  $\Pi_{q,\alpha} := \| |X|_\infty \|_{q,\alpha}$  denotes the  $L^\infty$  functional-dependence-adjusted norm. By condition (ii),  $\Pi_{q,\alpha}^q \lesssim p \Psi_{q,\alpha}^q$ , so the polynomial coefficients above are bounded respectively by  $p n m^{q/2-1-q\alpha} \Psi_{q,\alpha}^q$  and  $p L m^q \Psi_{q,\alpha}^q$ . The pair  $f_1^*$  captures the truncation gap  $T_X - T_{X,m}$ ; the pair  $f_2^*$  captures the small-block sum  $T_{X,m} - T_{Y,m}$ .

Pick

$$y = C \Psi_{2,\alpha} \left[ m^{-\alpha} \sqrt{n \log p} \vee \sqrt{L m \log p} \right] \vee C p^{1/q} \Psi_{q,\alpha} \left[ n^{1/q} m^{1/2-1/q-\alpha} \vee L^{1/q} m \right]$$

for a sufficiently large constant  $C > 0$ . With this choice, each of the four terms above is  $o(1)$ :

- $f_1^*$ , *Gaussian*:  $y \gtrsim \Psi_{2,\alpha} m^{-\alpha} \sqrt{n \log p}$  forces  $y^2 m^{2\alpha} / (n \Psi_{2,\alpha}^2) \gtrsim \log p$ , so  $p \exp(-c y^2 m^{2\alpha} / (n \Psi_{2,\alpha}^2)) \rightarrow 0$ .
- $f_1^*$ , *polynomial*:  $y \gtrsim p^{1/q} \sqrt{n} m^{1/2-1/q-\alpha} \Psi_{q,\alpha}$  ensures  $y^q \gtrsim p n^{q/2} m^{q/2-1-q\alpha} \Psi_{q,\alpha}^q$ , hence  $p n m^{q/2-1-q\alpha} \Psi_{q,\alpha}^q / y^q \rightarrow 0$ .
- $f_2^*$ , *Gaussian*:  $y \gtrsim \Psi_{2,\alpha} \sqrt{L m \log p}$  forces  $y^2 / (L m \Psi_{2,\alpha}^2) \gtrsim \log p$ , so the Gaussian tail of  $f_2^*$  vanishes.
- $f_2^*$ , *polynomial*:  $y \gtrsim p^{1/q} L^{1/q} m \Psi_{q,\alpha}$  gives  $p L m^q \Psi_{q,\alpha}^q / y^q \rightarrow 0$ .

Consequently  $\mathbb{P}(\|T_X - T_{Y,m}\|_\infty \geq y) \rightarrow 0$ .

Dividing  $y$  by  $\sqrt{n}$  and using  $L \asymp n/(M+m) \asymp n/M$  (since  $m \ll M$ ), the contribution to  $\delta(p, L, m, q, \alpha)$  in (36) is

$$\begin{aligned} \frac{y}{\sqrt{n}} &\lesssim \underbrace{m^{-\alpha} \sqrt{\log p} \Psi_{2,\alpha}}_{\text{trunc., Gaussian}} \vee \underbrace{\sqrt{m/M} \sqrt{\log p} \Psi_{2,\alpha}}_{\text{small-block, Gaussian}} \\ &\vee \underbrace{p^{1/q} m^{1/2-1/q-\alpha} n^{1/q-1/2} \Psi_{q,\alpha}}_{\text{trunc., polynomial}} \vee \underbrace{p^{1/q} L^{1/q} m n^{-1/2} \Psi_{q,\alpha}}_{\text{small-block, polynomial}}. \end{aligned}$$

The two Gaussian pieces are dominated by the  $m^{-\alpha} \log p$  and  $\sqrt{m/M} \log p$  terms appearing in (36). The two polynomial pieces are  $o(1)$  under condition (v): the truncation polynomial vanishes by the second clause  $p^{1/q} = o(m^{\alpha-1/2+1/q} n^{1/2-1/q})$ ; the small-block polynomial, using  $L \asymp n/M$  and  $M \gg m^2$  from (iv), reduces to  $p^{1/q} m M^{-1/q} n^{1/q-1/2}$ , which is  $o(1)$  under the same clause of (v).

Thus with probability tending to 1,

$$\left\| \frac{1}{\sqrt{n}} (T_X - T_{Y,m}) \right\|_\infty \lesssim m^{-\alpha} \log p \vee \sqrt{m/M} \log p,$$

contributing the second and third terms of  $\delta(p, L, m, q, \alpha)$  in (36).

**Step 2: Gaussian approximation of the independent block sums.**

By construction,  $\{\tilde{Y}_b\}_{b=1}^L$  are independent: each  $\tilde{Y}_b = \sum_{t \in \mathcal{B}_b} X_{t,m}$  depends only on  $\{\varepsilon_s : s \in \mathcal{B}'_b\}$  with  $\mathcal{B}'_b = [\min \mathcal{B}_b - m, \max \mathcal{B}_b]$ , and the small blocks of size  $m$  between consecutive big blocks ensure  $\mathcal{B}'_b \cap \mathcal{B}'_{b+1} = \emptyset$ . Define  $Z_b \sim \mathcal{N}(0, \mathbb{E}[\tilde{Y}_b \tilde{Y}_b^\top])$  independently. Apply Theorem 3.1 of Chernozhukov et al. (2016) to the normalized independent block sums  $L^{-1/2} \tilde{Y}_{b,j}$ ,  $1 \leq j \leq p$ : for every Borel set  $A \subset \mathbb{R}$  and  $\delta > 0$ ,

$$\mathbb{P}\left(\max_{1 \leq j \leq p} \left| L^{-1/2} \sum_{b=1}^L \tilde{Y}_{b,j} \right| \in A\right) - \mathbb{P}\left(\max_{1 \leq j \leq p} \left| L^{-1/2} \sum_{b=1}^L Z_{b,j} \right| \in A^{C_7 \delta}\right) \leq g(\delta), \quad (37)$$

where

$$g(\delta) := \frac{C_8 (\log p)^2}{\delta^3 \sqrt{L}} \{L^* + M_{L,1}(\delta) + M_{L,2}(\delta)\},$$

with universal constants  $C_7, C_8 > 0$ , and

$$\begin{aligned} L^* &:= \max_{1 \leq j \leq p} \frac{1}{L} \sum_{b=1}^L \mathbb{E} |\tilde{Y}_{b,j}|^3, \\ M_{L,1}(\delta) &:= \frac{1}{L} \sum_{b=1}^L \mathbb{E} \left[ \max_{1 \leq j \leq p} |\tilde{Y}_{b,j}|^3 \mathbf{1} \left\{ \max_{1 \leq j \leq p} |\tilde{Y}_{b,j}| \geq \frac{\delta \sqrt{L}}{\log p} \right\} \right], \\ M_{L,2}(\delta) &:= \frac{1}{L} \sum_{b=1}^L \mathbb{E} \left[ \max_{1 \leq j \leq p} |Z_{b,j}|^3 \mathbf{1} \left\{ \max_{1 \leq j \leq p} |Z_{b,j}| \geq \frac{\delta \sqrt{L}}{\log p} \right\} \right]. \end{aligned}$$

We bound each term using the following standard inequality.

**Lemma 3** (Burkholder 1988; Rio 2009). *Let  $q > 1$ ,  $q' = \min(q, 2)$ , and let  $M_n = \sum_{t=1}^n \xi_t$  where  $\xi_t \in \mathcal{L}^q$  are martingale differences. Then  $\|M_n\|_q^{q'} \leq K_q^{q'} \sum_{t=1}^n \|\xi_t\|_q^{q'}$ , where  $K_q = \max((q-1)^{-1}, \sqrt{q-1})$ .*

By Wu's m.d.s. decomposition,  $X_{t,j} = \sum_{s \leq t} \mathcal{P}_s X_{t,j}$  with  $\mathcal{P}_s = \mathbb{E}[\cdot | \mathcal{F}_s] - \mathbb{E}[\cdot | \mathcal{F}_{s-1}]$ , and the projections satisfy  $\|\mathcal{P}_s X_{t,j}\|_q \leq \delta_{t-s,q,j}$ . Applying Lemma 3 with  $q' = 2$  to the m.d.s. representation of  $\tilde{Y}_{b,j} = \sum_{t \in \mathcal{B}_b} X_{t,j,m}$  and using Minkowski's inequality across projections,

$$\|\tilde{Y}_{b,j}\|_q \leq K_q \sqrt{M} \sum_{s \geq 0} \delta_{s,q,j} \leq K_q \sqrt{M} \Psi_{q,\alpha}.$$

Therefore, at  $q = 3$  (which exists since  $q \geq 4$  in condition (i)),

$$L^* = \max_{1 \leq j \leq p} \frac{1}{L} \sum_{b=1}^L \mathbb{E} |\tilde{Y}_{b,j}|^3 \lesssim M^{3/2} \Psi_{q,\alpha}^3.$$

Setting  $v_{L,q} := \max_b (\mathbb{E}[\max_j |\tilde{Y}_{b,j}|^q])^{1/q} \lesssim p^{1/q} M^{1/2} \Psi_{q,\alpha}$  (the  $p^{1/q}$  from the union bound across coordinates), a Markov-type truncation argument following the proof of Theorem 2.1 in Chernozhukov

et al. (2016) gives

$$M_{L,1}(\delta) \leq \frac{(\log p)^{q-3}}{(\delta\sqrt{L})^{q-3}} \mathbb{E} \left[ \max_{1 \leq j \leq p} |\tilde{Y}_{b,j}|^q \right] \lesssim \frac{(\log p)^{q-3}}{(\delta\sqrt{L})^{q-3}} p M^{q/2} \Psi_{q,\alpha}^q.$$

For the Gaussian term, since  $Z_{b,j} \sim \mathcal{N}(0, \Sigma_{b,jj})$  with  $\Sigma_{b,jj} = \mathbb{E}[\tilde{Y}_{b,j}^2] \lesssim M \Psi_{2,\alpha}^2$ , an integration-by-parts argument using the Gaussian tail bound  $\mathbb{P}(|Z| > t) \leq 2(t\sqrt{2\pi})^{-1} \exp(-t^2/(2\sigma^2))$  for  $Z \sim \mathcal{N}(0, \sigma^2)$  yields

$$M_{L,2}(\delta) \lesssim M^{3/2} \Psi_{2,\alpha}^3 \left( \frac{\delta\sqrt{L}}{\sqrt{M} \Psi_{2,\alpha} \log p} \right)^2 \exp \left( -\frac{1}{2} \left\{ \frac{\delta\sqrt{L}}{\sqrt{M} \Psi_{2,\alpha} \log p} \right\}^2 \right).$$

Setting

$$T_2 := M^{1/2} L^{-1/6} (\log p)^{2/3} \Psi_{q,\alpha} \vee p^{1/q} M^{1/2} (\log p)^{1-1/q} L^{-(q-2)/(2q)} \Psi_{q,\alpha},$$

one verifies that  $g(\delta) \rightarrow 0$  for  $\delta \asymp T_2$  under condition (iv) ( $L^{-1/2} (\log p)^{3/2} \sqrt{M} \log(p \vee L) \ll T_2$  holds since  $\log(p \vee L) \ll L^{1/3}$ ). From (37),

$$\mathbb{P} \left( \left\| L^{-1/2} \sum_{b=1}^L (\tilde{Y}_b - Z_b) \right\|_{\infty} \geq T_2 \right) \rightarrow 0.$$

After rescaling to the  $n^{-1/2} \sum_t X_t$  scale (dividing by  $\sqrt{M}$ , since  $L^{-1/2} \sum_b \tilde{Y}_b = \sqrt{M} \cdot n^{-1/2} \sum_t X_{t,m}$  on  $\bigcup_b \mathcal{B}_b$ ), Step 2 contributes

$$T_2/\sqrt{M} \asymp L^{-1/6} (\log p)^{2/3} \Psi_{q,\alpha} \vee p^{1/q} (\log p)^{1-1/q} L^{-1/2+1/q} \Psi_{q,\alpha}$$

to the rate  $\delta(p, L, m, q, \alpha)$  in (36).

### Step 3: Approximation of Gaussian block sums by the limiting Gaussian.

Let  $\{X_t^z\}$  be a stationary Gaussian process with the same autocovariance structure as  $\{X_t\}$ :  $\mathbb{E}[X_t^z (X_{t-l}^z)^\top] = \mathbb{E}[X_t X_{t-l}^\top]$  for all  $l$ , so that  $n^{-1/2} \sum_{t=1}^n X_t^z \sim \mathcal{N}(0, \Sigma_n)$  with  $\Sigma_n \rightarrow \Sigma$  under condition (iii). Step 3 controls (a) the truncation error from removing the small blocks for  $\{X_t^z\}$  and (b) the covariance-matching error between the block-sum Gaussians  $Z_b$  and the limit  $Z \sim \mathcal{N}(0, \Sigma)$ .

(a) *Truncation error for the Gaussian process.* Let  $\tilde{X}_t^z := X_t^z - X_{t,m}^z$  denote the truncation residual of the Gaussian process, where  $X_{t,m}^z$  is the  $m$ -dependent approximation of  $X_t^z$  defined analogously to  $X_{t,m}$ . Since  $\sum_{b=1}^L Z_b \stackrel{d}{=} \sum_{b=1}^L \sum_{t \in \mathcal{B}_b} X_{t,m}^z$  (matched covariances by construction of  $Z_b$  in Step 2), Step 3(a) controls

$$S_n^z := n^{-1/2} \sum_{t=1}^n X_t^z - n^{-1/2} \sum_{b=1}^L \sum_{t \in \mathcal{B}_b} X_{t,m}^z = n^{-1/2} \sum_{t=1}^n \tilde{X}_t^z - n^{-1/2} \sum_{t \in \bigcup_b \mathcal{B}_b} X_{t,m}^z.$$

Both pieces of  $S_n^z$  are linear functionals of the Gaussian process  $\{X_t^z\}$ , hence each coordinate of  $S_n^z$

is itself centered Gaussian. We bound the variance of each coordinate.

By stationarity and the orthogonality structure of the projection decomposition, for each  $j$ ,

$$\mathbf{Var}\left(n^{-1/2} \sum_{t=1}^n \tilde{X}_{t,j}^z\right) \leq \|X_{:,j}^z - X_{:,j,m}^z\|_{2,\alpha}^2 \lesssim m^{-2\alpha} \Psi_{2,\alpha}^2,$$

using  $\sum_{l>m} \delta_{l,2,j}^z \leq m^{-\alpha} \Psi_{2,\alpha}$  and the Gaussian moment equivalence  $\|X_{t,j}^z\|_q \lesssim \sqrt{q} \|X_{t,j}^z\|_2$ , which extends  $\Psi_{q,\alpha}^z \lesssim \sqrt{q} \Psi_{2,\alpha}$ .

The total cardinality of  $\bigcup_b \mathcal{S}_b$  is  $Lm$ , and by an analogous variance computation (with  $Lm \Psi_{2,\alpha}^2$  playing the role of the total variance contribution),

$$\mathbf{Var}\left(n^{-1/2} \sum_{t \in \bigcup_b \mathcal{S}_b} X_{t,m,j}^z\right) \lesssim \frac{Lm}{n} \Psi_{2,\alpha}^2 \asymp \frac{m}{M} \Psi_{2,\alpha}^2$$

using  $L \asymp n/M$  (since  $m \ll M$ ).

Since each coordinate of  $S_n^z$  is centered Gaussian with variance  $\sigma_j^2 \lesssim m^{-2\alpha} \Psi_{2,\alpha}^2 \vee (m/M) \Psi_{2,\alpha}^2$ , by the standard Gaussian maximal inequality,

$$\mathbb{P}(\|S_n^z\|_\infty \geq y') \leq 2p \exp\left(-\frac{(y')^2}{c \sigma_{\max}^2}\right), \quad \sigma_{\max}^2 \lesssim m^{-2\alpha} \Psi_{2,\alpha}^2 \vee \frac{m}{M} \Psi_{2,\alpha}^2.$$

Setting

$$\tilde{T}_{3,\text{trunc}} := m^{-\alpha} \sqrt{\log p} \Psi_{2,\alpha} \vee \sqrt{m/M} \sqrt{\log p} \Psi_{2,\alpha},$$

we obtain  $\mathbb{P}(\|S_n^z\|_\infty \geq \tilde{T}_{3,\text{trunc}}) \rightarrow 0$ . The contribution to  $\delta(p, L, m, q, \alpha)$  from Step 3(a) is therefore  $\tilde{T}_{3,\text{trunc}}$ , which after the  $\sqrt{\log p}$  inflation in the conversion to  $\rho_K$  (see (39)) becomes

$$\tilde{T}_{3,\text{trunc}} \sqrt{\log p} \asymp m^{-\alpha} \log p \Psi_{2,\alpha} \vee \sqrt{m/M} \log p \Psi_{2,\alpha},$$

matching the second and third terms of  $\delta$  in (36).

(b) *Covariance matching.* Step 3(a) has bridged from  $n^{-1/2} \sum_{b=1}^L Z_b$  to  $n^{-1/2} \sum_{t=1}^n X_t^z \sim \mathcal{N}(0, \Sigma_n)$ , where

$$\Sigma_n := \mathbf{Var}\left(n^{-1/2} \sum_{t=1}^n X_t\right) = \sum_{|l|<n} \left(1 - \frac{|l|}{n}\right) \mathbb{E}[X_0 X_l^\top]$$

is the finite-sample autocovariance matrix. It remains to close the gap to the limiting covariance  $\Sigma$ . Under condition (ii), absolute summability of autocovariances and Corollary 5.3 of Zhang and Wu (2017) give

$$\Delta := \max_{1 \leq j, k \leq p} |\Sigma_{n,jk} - \sigma_{jk}| = O(v(M)),$$

where  $v(M) = M^{-1}$  if  $\alpha > 1$ ,  $(\log M)/M$  if  $\alpha = 1$ , and  $M^{-\alpha}$  if  $0 < \alpha < 1$ . By the Gaussian comparison inequality (Chernozhukov et al., 2015, Theorem 1) applied to the centered Gaussian vectors  $n^{-1/2} \sum_{t=1}^n X_t^z \sim \mathcal{N}(0, \Sigma_n)$  and  $Z \sim \mathcal{N}(0, \Sigma)$ ,

$$\rho_K\left(\frac{1}{\sqrt{n}} \sum_{t=1}^n X_t^z, Z\right) \lesssim \Delta^{1/3} (\log p)^{2/3} = v(M)^{1/3} (\log p)^{2/3}.$$

Under condition (iv),  $v(M)^{1/3}(\log p)^{2/3} \rightarrow 0$ , so this final covariance-matching error vanishes and  $Z \sim \mathcal{N}(0, \Sigma)$  serves as the limiting Gaussian.

**Combining the three steps via coupling.** For random vectors  $V, W \in \mathbb{R}^p$ , write

$$\rho_K(V, W) := \sup_{u \in \mathbb{R}} \left| \mathbb{P}(\|V\|_\infty \leq u) - \mathbb{P}(\|W\|_\infty \leq u) \right|.$$

*Anti-concentration.* By the anti-concentration inequality for maxima of Gaussian vectors (Chernozhukov et al., 2015, Theorem 3), applied to  $Z \sim \mathcal{N}(0, \Sigma)$  with  $\min_j \sigma_{jj} \geq c > 0$  from condition (iii),

$$\sup_{u \in \mathbb{R}} \mathbb{P}(\|Z\|_\infty - u \leq \epsilon) \lesssim \epsilon \sqrt{\log p}. \quad (38)$$

Combined with Chernozhukov et al. (2016, Lemma 2.1), every tail bound of the form  $\mathbb{P}(\|V - W\|_\infty \geq \epsilon) \leq \eta$  where the law of  $W$  satisfies (38) yields

$$\rho_K(V, W) \lesssim \epsilon \sqrt{\log p} + \eta. \quad (39)$$

*Chain of Kolmogorov-distance bounds.* Steps 1, 2, 3(a), and 3(b) yield, respectively,

$$\begin{aligned} \rho_K\left(\frac{1}{\sqrt{n}} \sum_t X_t, \frac{1}{\sqrt{n}} \sum_b \tilde{Y}_b\right) &\lesssim T_1 \sqrt{\log p} + o(1) && \text{(Step 1),} \\ \rho_K\left(\frac{1}{\sqrt{n}} \sum_b \tilde{Y}_b, \frac{1}{\sqrt{n}} \sum_b Z_b\right) &\lesssim (T_2/\sqrt{M}) \sqrt{\log p} + o(1) && \text{(Step 2),} \\ \rho_K\left(\frac{1}{\sqrt{n}} \sum_b Z_b, \frac{1}{\sqrt{n}} \sum_t X_t^z\right) &\lesssim \tilde{T}_{3,\text{trunc}} \sqrt{\log p} + o(1) && \text{(Step 3(a)),} \\ \rho_K\left(\frac{1}{\sqrt{n}} \sum_t X_t^z, Z\right) &\lesssim v(M)^{1/3}(\log p)^{2/3} && \text{(Step 3(b)),} \end{aligned}$$

where the first three lines apply (39) to the tail bounds produced in the corresponding steps, and the fourth line follows from the Gaussian comparison (Chernozhukov et al., 2015, Theorem 1) with covariance gap  $\Delta = O(v(M))$  from condition (iii) and Corollary 5.3 of Zhang and Wu (2017). Here  $\tilde{T}_{3,\text{trunc}} := m^{-\alpha} \sqrt{\log p} \Psi_{2,\alpha} \vee \sqrt{m/M} \sqrt{\log p} \Psi_{2,\alpha}$  denotes the tail-bound size for Step 3(a) computed via the direct Gaussian-max bound (see remark below (36)).

*Triangle inequality.* By the triangle inequality for  $\rho_K$ ,

$$\rho_K\left(\frac{1}{\sqrt{n}} \sum_{t=1}^n X_t, Z\right) \lesssim (T_1 + T_2/\sqrt{M} + \tilde{T}_{3,\text{trunc}}) \sqrt{\log p} + v(M)^{1/3}(\log p)^{2/3} + o(1).$$

Substituting the expressions for  $T_1$ ,  $T_2/\sqrt{M}$ , and  $\tilde{T}_{3,\text{trunc}}$  established in Steps 1–3 and collecting terms, the right-hand side is bounded by  $\delta(p, L, m, q, \alpha)$  as in (36).

*Coupling.* By the conditional Strassen theorem in the form stated as Chernozhukov et al. (2016, Lemma 4.2), the preceding Kolmogorov-distance bound between the laws of  $\frac{1}{\sqrt{n}} \sum_{t=1}^n X_t$  and  $Z$  on  $\mathbb{R}^p$  (equipped with the  $\|\cdot\|_\infty$  norm) can be realized as a coupling: on a suitably enriched probability



space, there exists  $Z \sim \mathcal{N}(0, \Sigma)$  such that

$$\mathbb{P}\left(\left\|\frac{1}{\sqrt{n}} \sum_{t=1}^n X_t - Z\right\|_{\infty} \geq \delta(p, L, m, q, \alpha)\right) \rightarrow 0,$$

which is the assertion of Theorem 5.  $\square$

*Proof of Theorem 4.* We establish both claims in five steps. Throughout, fix  $h$  and write  $n = T - h \asymp T$ . Define the (infeasible) rescaled stochastic and bias scores

$$\mathbf{S}_{h,t}^{(u)} := \frac{1}{\sqrt{N}} X_t \Phi(Z_{t-1})^\top u_{h,t+h} \in \mathbb{R}^J, \quad \mathbf{S}_{h,t}^{(R)} := \frac{1}{\sqrt{N}} X_t^2 \Phi(Z_{t-1})^\top R_h(Z_{t-1}) \in \mathbb{R}^J,$$

where  $R_h(Z_{t-1}) := (R_h(Z_{i,t-1}))_{i=1}^N$  stacks the sieve approximation residuals. The residualized response  $v_{h,t+h} := Y_{t+h} - D_t \theta_h$  then admits the decomposition

$$v_{h,t+h} = u_{h,t+h} + X_t R_h(Z_{t-1}),$$

with  $\mu_{Wu} = \mathbb{E}[W_{t-1}^\top u_{h,t+h}] = 0$  by the population projection definition of  $\gamma_h$ .

Hence  $\frac{1}{\sqrt{NT}} \sum_{t=1}^{T-h} X_t \Phi(Z_{t-1})^\top v_{h,t+h} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} (\mathbf{S}_{h,t}^{(u)} + \mathbf{S}_{h,t}^{(R)})$ . The first sum drives the asymptotic distribution; the second is asymptotically negligible under the undersmoothing condition (Step 2).

Define the long-run covariance of  $\{\mathbf{S}_{h,t}^{(u)}\}$  and the asymptotic variance by

$$\Omega_h^* := \sum_{k \in \mathbb{Z}} \mathbb{E}[\mathbf{S}_{h,t}^{(u)} (\mathbf{S}_{h,t-k}^{(u)})^\top], \quad V_h^* := A_{11,h}^{-1} \Omega_h^* A_{11,h}^{-1}.$$

**Step 1 (Sieve decomposition and uniform rate).** Decompose the estimation error pointwise:

$$\widehat{g}_h(z) - g_h(z) = \phi(z)^\top (\widehat{b}_h - b_h) - R_h(z), \quad R_h(z) := g_h(z) - \phi(z)^\top b_h.$$

Since  $g_h \in \mathcal{H}^\kappa(\mathcal{Z})$  with  $\kappa > 1/2$  and  $\mathcal{G}_M$  is the Hölder ball of Assumption 6(1), classical sieve approximation theory gives  $\sup_{z \in \mathcal{Z}} |R_h(z)| = O(J^{-\kappa})$ .

By Assumption 6(1),  $\mathbb{E}[X_t | \mathcal{F}_{t-1}] = 0$ , so  $A_{12,h} = 0$  and  $\widetilde{A}_{11,h} = A_{11,h}$ . Writing  $g_h(Z_{i,t-1}) = \phi(Z_{i,t-1})^\top b_h + R_h(Z_{i,t-1})$  and substituting into the model  $Y_{i,t+h} = g_h(Z_{i,t-1}) X_t + W_{i,t-1}^\top \gamma_h + u_{i,t+h}$ , the normal equations give the linearization

$$\widehat{b}_h - b_h = \widehat{A}_{11,h}^{-1} (\mathbf{S}_{h,n}^{(\text{bias})} + \mathbf{S}_{h,n}^{(u)}),$$

where

$$\mathbf{S}_{h,n}^{(\text{bias})} := \frac{1}{NT} \sum_{t=1}^{T-h} X_t^2 \Phi(Z_{t-1})^\top R_h(Z_{t-1}), \quad \mathbf{S}_{h,n}^{(u)} := \frac{1}{NT} \sum_{t=1}^{T-h} X_t \Phi(Z_{t-1})^\top u_{h,t+h},$$

with  $R_h(Z_{t-1}) := (R_h(Z_{i,t-1}))_{i=1}^N \in \mathbb{R}^N$  stacking the sieve approximation residuals. (Contributions from  $W_{t-1}^\top \gamma_h$  are absorbed into the  $\widehat{A}_{12,h} \widehat{A}_{22,h}^{-1}$  correction, which is  $O_p(\sqrt{J/(NT)})$  under  $A_{12,h} = 0$

and hence dominated by  $\mathbf{S}_{h,n}^{(u)}$ ; cf. proof of Theorem 2.)

*Bias rate.* For each coordinate  $j$ ,

$$\begin{aligned} |\mathbf{S}_{h,n,j}^{(\text{bias})}| &\leq \frac{1}{NT} \sum_{t=1}^{T-h} X_t^2 \sum_{i=1}^N |\phi_j(Z_{i,t-1})| \cdot |R_h(Z_{i,t-1})| \\ &\lesssim \sup_{j,z} |\phi_j(z)| \cdot \sup_z |R_h(z)| \cdot \frac{1}{T} \sum_t X_t^2 = O_p(J^{-\kappa}), \end{aligned}$$

using  $\sup_{j,z} |\phi_j(z)| = O(1)$  for B-spline bases (cf. Newey 1997) and  $\mathbb{E}[X_t^2] < \infty$ . Hence the per-coordinate sup-norm rate is

$$\|\mathbf{S}_{h,n}^{(\text{bias})}\|_\infty = O_p(J^{-\kappa}), \quad \|\mathbf{S}_{h,n}^{(\text{bias})}\|_2 = O_p(J^{1/2-\kappa}).$$

*Stochastic rate.* For each coordinate  $j$ ,

$$\text{Var}(\mathbf{S}_{h,n,j}^{(u)}) = \frac{1}{(NT)^2} \sum_t \mathbb{E} \left[ X_t^2 \left( \sum_i \phi_j(Z_{i,t-1}) u_{i,t+h} \right)^2 \right] = O\left(\frac{1}{NT}\right),$$

using m.d.s. exogeneity in time, i.i.d. across  $i$ , and finite variances in Assumption 6(2)–(3). Summing over  $J$  coordinates and applying Markov's inequality,

$$\|\mathbf{S}_{h,n}^{(u)}\|_2 = O_p(\sqrt{J/(NT)}).$$

*Coefficient rate.* By Lemma 2,  $\|\hat{A}_{11,h}^{-1}\| = O_p(1)$ . Combining the above,

$$\|\hat{b}_h - b_h\|_2 \leq \|\hat{A}_{11,h}^{-1}\| (\|\mathbf{S}_{h,n}^{(\text{bias})}\|_2 + \|\mathbf{S}_{h,n}^{(u)}\|_2) = O_p(J^{1/2-\kappa}) + O_p(\sqrt{J/(NT)}).$$

*Sup-norm rate (leading bound).* Using  $\sup_z \|\phi(z)\| \leq C\sqrt{J}$  from Assumption 6(5),

$$\sup_{z \in \mathcal{Z}} |\phi(z)^\top (\hat{b}_h - b_h)| \leq C\sqrt{J} \|\hat{b}_h - b_h\|_2 = O_p\left(\frac{J}{\sqrt{NT}}\right) + O_p(J^{1-\kappa}).$$

*Sup-norm rate (Riesz-kernel sharpening for the bias).* The above sup-norm bound on the bias contribution can be sharpened using the sieve reproducing kernel  $K_J(z, z') := \phi(z)^\top A_{11,h}^{-1} \phi(z')$ . Indeed,

$$\phi(z)^\top A_{11,h}^{-1} \mathbf{S}_{h,n}^{(\text{bias})} = \frac{1}{NT} \sum_t X_t^2 \sum_i K_J(z, Z_{i,t-1}) R_h(Z_{i,t-1}),$$

so that

$$\begin{aligned} \sup_z |\phi(z)^\top A_{11,h}^{-1} \mathbf{S}_{h,n}^{(\text{bias})}| &\leq \sup_z |R_h(z)| \cdot \sup_z \frac{1}{NT} \sum_{t,i} X_t^2 |K_J(z, Z_{i,t-1})| \\ &= O_p(J^{-\kappa}) \cdot O_p(\sqrt{J}) = O_p(J^{1/2-\kappa}), \end{aligned}$$

using  $\sup_z \frac{1}{N} \sum_i |K_J(z, Z_{i,t-1})| = O_p(\sqrt{J})$  for the B-spline kernel (a standard sieve result; see, e.g., Newey 1997; Chen and Christensen 2015). Combining,

$$\sup_z |\phi(z)^\top (\hat{b}_h - b_h)| = O_p\left(\frac{J}{\sqrt{NT}}\right) + O_p(J^{1/2-\kappa}).$$

*Higher-order cross terms.* Substituting the Neumann series expansion from Lemma 2,  $\hat{A}_{11,h}^{-1} = A_{11,h}^{-1} + R_{A,n}$  with  $\|R_{A,n}\| = O_p(T^{-1/2} + \sqrt{J/(NT)} + J/(NT))$ , and applying the Riesz-kernel argument to  $R_{A,n}$  as well, the cross terms are

$$\begin{aligned} \sup_z |\phi(z)^\top R_{A,n} \mathbf{S}_{h,n}^{(\text{bias})}| &\leq \sqrt{J} \cdot \|R_{A,n}\| \cdot \|\mathbf{S}_{h,n}^{(\text{bias})}\|_\infty = O_p(T^{-1/2} J^{1/2-\kappa}), \\ \sup_z |\phi(z)^\top R_{A,n} \mathbf{S}_{h,n}^{(u)}| &\leq \sqrt{J} \cdot \|R_{A,n}\| \cdot \|\mathbf{S}_{h,n}^{(u)}\|_2 = O_p(T^{-1/2} \frac{J}{\sqrt{NT}}) + O_p(\frac{J^{3/2}}{NT}). \end{aligned}$$

Adding  $\sup_z |R_h(z)| = O(J^{-\kappa})$ , which is dominated by  $O_p(J^{1/2-\kappa})$ ,

$$\sup_{z \in \mathcal{Z}} |\hat{g}_h(z) - g_h(z)| = O_p\left(\frac{J}{\sqrt{NT}}\right) + O_p(J^{1/2-\kappa}) + O_p(T^{-1/2} J^{1/2-\kappa}) + O_p\left(T^{-1/2} \frac{J}{\sqrt{NT}}\right) + O_p\left(\frac{J^{3/2}}{NT}\right),$$

yielding part (a) of Theorem 4.

**Step 2 (Linear representation).** Write  $\hat{A}_{11,h}^{-1} = A_{11,h}^{-1} + R_{A,n}$ . Then

$$\sqrt{NT} (\hat{b}_h - b_h) = A_{11,h}^{-1} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \mathbf{S}_{h,t}^{(u)} \right) + r_{h,n},$$

where  $\mathbf{S}_{h,t}^{(u)} := N^{-1/2} X_t \Phi(Z_{t-1})^\top u_{h,t+h}$  is the rescaled stochastic score, and

$$r_{h,n} = R_{A,n} \left( \frac{1}{\sqrt{T}} \sum_t \mathbf{S}_{h,t}^{(u)} \right) + \sqrt{NT} A_{11,h}^{-1} \mathbf{S}_{h,n}^{(\text{bias})}.$$

*First term: Neumann remainder times stochastic score.* For each coordinate  $j$ ,  $T^{-1/2} \sum_t S_{h,t,j}^{(u)} = O_p(1)$  by stationarity and absolute summability of autocovariances of  $\{S_{h,t,j}^{(u)}\}$  (Assumption 6(2)). Summing over the  $J$  coordinates and applying Markov,

$$\left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \mathbf{S}_{h,t}^{(u)} \right\|_2 = O_p(\sqrt{J}).$$

By Lemma 2,  $\|R_{A,n}\| = O_p(T^{-1/2} + \sqrt{J/(NT)} + J/(NT))$ . Hence

$$\|R_{A,n} T^{-1/2} \sum_t \mathbf{S}_{h,t}^{(u)}\|_2 \leq \|R_{A,n}\| \cdot \|T^{-1/2} \sum_t \mathbf{S}_{h,t}^{(u)}\|_2 = O_p\left(\sqrt{\frac{J}{T}} + \frac{J}{\sqrt{NT}} + \frac{J^{3/2}}{NT}\right).$$

Under Assumption 6(6) ( $J^2/(NT) \rightarrow 0$  and, in the proof's asymptotic regime,  $J = o(T)$ ), all three terms are  $o_p(1)$ .

*Second term: bias amplification.* By Lemma 2,  $\|\hat{A}_{11,h}^{-1}\| = O_p(1)$ . Using the  $\ell^2$ -norm bias rate  $\|\mathbf{S}_{h,n}^{(\text{bias})}\|_2 = O_p(J^{1/2-\kappa})$  from Step 1,

$$\sqrt{NT} \|\hat{A}_{11,h}^{-1} \mathbf{S}_{h,n}^{(\text{bias})}\|_2 \leq O_p(1) \cdot \sqrt{NT} \|\mathbf{S}_{h,n}^{(\text{bias})}\|_2 = O_p(\sqrt{NT} J^{1/2-\kappa}) = o_p(1)$$

under the undersmoothing condition  $\sqrt{NT} J^{1/2-\kappa} \rightarrow 0$  in hypothesis (iv) of Theorem 4. Hence  $\|r_{h,n}\|_2 = o_p(1)$ .

*Sup-norm form.* By Step 4 (using the Riesz  $\ell^1$ -localization  $\sup_z \|\tilde{\phi}(z)\|_1 = O(J^\rho)$  from (41)),

$$\sup_{z \in \mathcal{Z}} |\phi(z)^\top r_{h,n}| \leq \sup_z \|\tilde{\phi}(z)\|_1 \cdot \|A_{11,h}\| \cdot \|r_{h,n}\|_2 \lesssim J^\rho \|r_{h,n}\|_2 = o_p(J^\rho).$$

For the anti-concentration step in Step 5 we will need the stronger control  $\sup_z |\phi(z)^\top r_{h,n}| = o_p(1/\sqrt{\log J})$ , which holds under hypothesis (iv) provided

$$\sqrt{NT} J^{\rho+1/2-\kappa} (\log J)^{1/2} = o(1).$$

For B-splines ( $\rho = 0$ ), this reduces to  $\sqrt{NT} J^{1/2-\kappa} (\log J)^{1/2} = o(1)$ , which is implied by the undersmoothing condition  $\sqrt{NT} J^{1/2-\kappa} \rightarrow 0$  in hypothesis (iv) up to the  $(\log J)^{1/2}$  factor. For non-localized bases ( $\rho = 1/2$ ), the condition becomes  $\sqrt{NT} J^{1-\kappa} (\log J)^{1/2} = o(1)$ , requiring marginally stronger undersmoothing.

**Step 3 (Verifying the conditions of Theorem 5 for  $\{\mathbf{S}_{h,t}^{(u)}\}$ ).** Apply Theorem 5 with  $\{X_t^{\text{Thm}}\} = \{\mathbf{S}_{h,t}^{(u)}\}$ ,  $p = J$ ,  $n = T - h \asymp T$ . Recall  $\mathbf{S}_{h,t}^{(u)} = N^{-1/2} X_t \Phi(Z_{t-1})^\top u_{h,t+h}$  with  $j$ th coordinate  $S_{h,t,j}^{(u)} = N^{-1/2} \sum_{i=1}^N X_t \phi_j(Z_{i,t-1}) u_{i,t+h}$ .

*Condition (i): moments.* Write  $S_{h,t,j}^{(u)} = X_t T_{t,j}$ , with  $T_{t,j} := N^{-1/2} \sum_{i=1}^N \phi_j(Z_{i,t-1}) u_{i,t+h}$ . By Assumption 6(1),  $\{(Z_{i,t-1}, u_{i,t+h})\}_{i=1}^N$  are i.i.d. across  $i$  at fixed  $t$ , and each summand has mean zero by m.d.s. exogeneity (conditional on the time-series  $\sigma$ -field  $\mathcal{F}_{t-1}$ ). Applying Rosenthal's inequality across  $i$  to  $\sqrt{N} T_{t,j} = \sum_i \phi_j(Z_{i,t-1}) u_{i,t+h}$  at order  $2q$ :

$$\mathbb{E} |T_{t,j}|^{2q} \lesssim N^{1-q} \mathbb{E} |\phi_j(Z_{i,t-1}) u_{i,t+h}|^{2q} + (\mathbb{E} [\phi_j(Z_{i,t-1})^2 u_{i,t+h}^2])^q.$$

For B-spline bases,  $\sup_{j,z} |\phi_j(z)| = O(1)$  (Newey, 1997), and  $\mathbb{E} |u_{i,t+h}|^{2q} < \infty$  from hypothesis (iii). Hence both terms are  $O(1)$ :

$$\sup_{t,j} \mathbb{E} |T_{t,j}|^{2q} = O(1).$$

By Cauchy-Schwarz and  $\mathbb{E} |X_t|^{2q} < \infty$  from hypothesis (iii),

$$\mathbb{E} |S_{h,t,j}^{(u)}|^q \leq (\mathbb{E} |X_t|^{2q})^{1/2} (\mathbb{E} |T_{t,j}|^{2q})^{1/2} = O(1),$$

so  $\sup_{t,j} \mathbb{E} |S_{h,t,j}^{(u)}|^q = O(1)$ , verifying Condition (i) of Theorem 5. *Condition (ii): functional depen-*

dence. Under Assumption 6(2), the residual admits the MA representation

$$u_{i,t+h} = \sum_{l=0}^{\infty} a_{h,l} \eta_{i,t+h-l}, \quad \sum_{l=0}^{\infty} l^{1+\delta} \|a_{h,l}\| < \infty,$$

where  $\{\eta_{it}\}$  is i.i.d. across  $i$  and  $t$  with mean zero and finite  $2q$ -th moments by hypothesis (iii). To compute the functional-dependence measure of  $\{\mathbf{S}_{h,t}^{(u)}\}$  in the time direction, replace the time-0 cross-section shock vector  $(\eta_{i,0})_{i=1}^N$  with an independent copy  $(\eta_{i,0}^*)_{i=1}^N$ . Each  $u_{i,t+h}$  changes by  $a_{h,t+h}^\top (\eta_{i,0} - \eta_{i,0}^*)$ , differently for each  $i$ . The coupled difference is therefore

$$S_{h,t,j}^{(u)} - S_{h,t,j}^{(u),*} = N^{-1/2} \sum_{i=1}^N X_t \phi_j(Z_{i,t-1}) a_{h,t+h}^\top (\eta_{i,0} - \eta_{i,0}^*),$$

which is a sum of  $N$  independent mean-zero terms across  $i$ .

By Rosenthal's inequality across  $i$ , conditional on  $X_t$  and  $\{Z_{i,t-1}\}_{i=1}^N$ :

$$\|S_{h,t,j}^{(u)} - S_{h,t,j}^{(u),*}\|_q \lesssim \|X_t\|_{2q} \cdot \sup_{j,z} |\phi_j(z)| \cdot \|a_{h,t+h}\| \cdot \|\eta_0 - \eta_0^*\|_{2q},$$

using  $\sup_{j,z} |\phi_j(z)| = O(1)$  for B-spline bases and finite  $2q$ -th moments of  $X_t$  and  $\eta$ . Hence

$$\|S_{h,t,j}^{(u)} - S_{h,t,j}^{(u),*}\|_q \lesssim \|a_{h,t+h}\|,$$

uniformly in  $j$  and independently of  $N$ .

Summing over  $t$  with the weighting in  $\Theta_{q,\alpha}$ ,

$$\Theta_{q,\alpha} := \max_j \sum_{t=0}^{\infty} (t+1)^\alpha \|S_{h,t,j}^{(u)} - S_{h,t,j}^{(u),*}\|_q \lesssim \sum_{t=0}^{\infty} (t+1)^\alpha \|a_{h,t}\| < \infty$$

for any  $\alpha \leq 1 + \delta$ , by the summability hypothesis. The analogous norm  $\Psi_{2,\alpha}$  is finite by the same argument with  $q = 2$ . The standing condition  $\Pi_{q,\alpha} \lesssim p^{1/q} \Psi_{q,\alpha}$  in Theorem 5 also holds since  $\Theta_{q,\alpha} = O(1)$ .

*Condition (iii): long-run covariance.* The long-run covariance  $\Omega_h^* = \sum_{k \in \mathbb{Z}} \mathbb{E}[\mathbf{S}_{h,t}^{(u)} \mathbf{S}_{h,t-k}^{(u)\top}]$  exists by the absolute summability of autocovariances established in Assumption 6(2) and Ibragimov and Linnik (1971, Theorem 18.6.5). By Assumption 6(2),  $\lambda_{\min}(\Omega_h^*) \geq c > 0$ , and  $\lambda_{\max}(\Omega_h^*) \leq C < \infty$  by the moment bounds of Condition (i).

*Condition (iv): block-size conditions.* Set  $L = \lfloor T^{2/3} \rfloor$ ,  $m = \lfloor T^{1/(9\alpha)} \rfloor$ ,  $M = \lfloor T^{1/3}/2 \rfloor$ . We have  $m + M \asymp T^{1/3}$  and  $m \ll M$  provided  $\alpha > 2/3$  (so that  $1/(9\alpha) < 1/6$  — see below). We verify each clause of Theorem 5's condition (iv) with  $p = J$ :

$$\begin{aligned}
(\log J)^9 &\ll L = T^{2/3} && \text{requires } (\log J)^9 \ll T^{2/3}, \\
M \gg m^2 &\iff T^{1/3} \gg T^{2/(9\alpha)} && \text{requires } \alpha > 2/3, \\
M \gg m(\log p)^3 &\iff T^{1/3} \gg T^{1/(9\alpha)}(\log T)^3 && \text{automatic for } \alpha > 1/3 \text{ and any moderate } \log T, \\
m^\alpha &\gg \log p = \log J, && \text{i.e. } T^{1/9} \gg \log J, \text{ since } J = o(T^{1/2}), \\
M^{\alpha \wedge 1} &\gg (\log p)^2 = (\log J)^2, && \text{automatic for any moderate } \log J.
\end{aligned}$$

The binding requirement is  $\alpha > 2/3$ , consistent with the worked example in Remark 4 of Theorem 5. The clause  $(\log J)^9 \ll T^{2/3}$  is implied by  $\log J \ll T^{2/27}$ , which holds for any polynomial  $J$  in  $T$ . *Identifying  $\alpha$  in Theorem 5.* By Assumption 6(2), the MA coefficients of  $u_{h,t+h}$  satisfy  $\sum_l l^{1+\delta} \|a_{h,l}\| < \infty$  for some  $\delta > 0$ . As shown above,  $\delta_{t,q,j} \lesssim \|a_{h,t+h}\|$ , so  $\sum_{l>m} \delta_{l,q,j} \leq (m+1)^{-(1+\delta)} \sum_l l^{1+\delta} \|a_{h,l}\| = O(m^{-(1+\delta)})$ . Hence the functional-dependence-adjusted norm  $\Psi_{q,\alpha}$  in Theorem 5 is finite for any  $\alpha \leq 1 + \delta$ . We choose  $\alpha = 1 + \delta > 1 > 2/3$ , so the worked-example block-size choice  $L = T^{2/3}$  in Theorem 5 (which requires  $\alpha > 2/3$ ) is admissible. *Condition (v): polynomial-tail.* Both clauses of condition (v) of Theorem 5 reduce, after substituting  $L = T^{2/3}$ ,  $m = T^{1/(9\alpha)}$ , and  $p = J$ , to the hypothesis  $J^{1/2}(\log J)^{7/6} = o(T^{1/9})$  in hypothesis (iv) of Theorem 4, after using  $q$  sufficiently large.

*Application of Theorem 5.* Under the verified conditions, by Remark 4 of Theorem 5,

$$\delta(J, L, m, q, \alpha) \lesssim T^{-1/9}(\log J)^{7/6} \rightarrow 0$$

under hypothesis (iv). Theorem 5 then yields, on a suitably enriched probability space, a centered Gaussian vector  $Z^\dagger \sim \mathcal{N}(0, \Omega_h^*)$  such that

$$\mathbb{P}\left(\left\|\frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \mathbf{s}_{h,t}^{(u)} - Z^\dagger\right\|_\infty \geq \delta(J, L, m, q, \alpha)\right) \rightarrow 0. \quad (40)$$

**Step 4 (From  $\ell^\infty$  to  $\sup_{z \in \mathcal{Z}}$ ).** Define

$$\mathbb{G}_h(z) := \sqrt{NT} \phi(z)^\top (\widehat{b}_h - b_h), \quad Z_h^{\text{int}}(z) := \phi(z)^\top A_{11,h}^{-1} (\Omega_h^*)^{1/2} \mathcal{N}_J,$$

both at the  $\sqrt{NT}$ -rescaled scale, where  $\mathcal{N}_J \sim \mathcal{N}(0, I_J)$ . By Step 2,

$$\mathbb{G}_h(z) = \phi(z)^\top A_{11,h}^{-1} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \mathbf{s}_{h,t}^{(u)} \right) + \phi(z)^\top r_{h,n}.$$

*Riesz-kernel bound on the basis transform.* We assume the sieve basis satisfies the Riesz  $\ell^1$ -

localization

$$\sup_{z \in \mathcal{Z}} \|\tilde{\phi}(z)\|_1 = O(J^\rho), \quad \tilde{\phi}(z) := A_{11,h}^{-1} \phi(z), \quad (41)$$

for some  $\rho \geq 0$  depending on the basis. For B-splines,  $\rho = 0$  by the banded inverse property of the gram matrix (Chen and Christensen 2015); for general orthonormal polynomial bases, typically  $\rho = 1/2$  (arising from the Cauchy–Schwarz bound  $\|\tilde{\phi}(z)\|_1 \leq \sqrt{J} \|\tilde{\phi}(z)\|_2 = O(\sqrt{J})$  in the absence of localization).

*Remainder negligibility.* By Step 2,  $\|r_{h,n}\|_2 = o_p(1)$  under hypothesis (iv). Using (41),

$$\sup_{z \in \mathcal{Z}} |\phi(z)^\top r_{h,n}| \leq \sup_z \|\tilde{\phi}(z)\|_1 \cdot \|A_{11,h} r_{h,n}\|_\infty \lesssim J^\rho \|r_{h,n}\|_2 = o_p(J^\rho).$$

For the anti-concentration step in Step 5 we will need  $\sup_z |\phi(z)^\top r_{h,n}| = o_p(1/\sqrt{\log J})$ , which holds under hypothesis (iv) provided  $\sqrt{NT} J^{\rho+1/2-\kappa} (\log J)^{1/2} = o(1)$ .

*Lipschitz modulus of  $\mathbb{G}_h$  and  $Z_h^{\text{int}}$ .* By hypothesis (ii) of Theorem 4, the basis is  $O(J)$ -Lipschitz on  $\mathcal{Z}$  with the localization  $\|\phi(z) - \phi(z')\|_2 \leq C J |z - z'|$  (the factor  $J$  arises since each individual basis function has Lipschitz constant  $\asymp J$  and only  $O(1)$  are active at each  $z$ , for B-splines and similar localized bases). Combined with  $\|A_{11,h}^{-1}\| = O(1)$  and  $\|T^{-1/2} \sum_t \mathbf{S}_{h,t}^{(u)}\|_2 = O_p(\sqrt{J})$  from Step 2,

$$|\mathbb{G}_h(z) - \mathbb{G}_h(z')| \leq \|\phi(z) - \phi(z')\|_2 \cdot \|A_{11,h}^{-1}\| \cdot \|T^{-1/2} \sum_t \mathbf{S}_{h,t}^{(u)}\|_2 = O_p(J^{3/2}) |z - z'|.$$

For the auxiliary Gaussian process, the increment variance satisfies

$$\text{Var}(Z_h^{\text{int}}(z) - Z_h^{\text{int}}(z')) = (\phi(z) - \phi(z'))^\top A_{11,h}^{-1} \Omega_h^* A_{11,h}^{-1} (\phi(z) - \phi(z')) \leq C J^2 |z - z'|^2,$$

so by Dudley’s entropy bound (van der Vaart and Wellner 1996, Corollary 2.2.8), for any subset of  $\mathcal{Z}$  of diameter  $\varepsilon$ ,

$$\mathbb{E} \sup_{|z-z'| \leq \varepsilon} |Z_h^{\text{int}}(z) - Z_h^{\text{int}}(z')| \leq C J \varepsilon \sqrt{\log(1/\varepsilon)}.$$

*Net construction.* Construct an  $\varepsilon$ -net  $\{z_\ell\}_{\ell=1}^{N_\varepsilon}$  of  $\mathcal{Z}$  with  $N_\varepsilon = O(\varepsilon^{-1})$ . Choose  $\varepsilon = J^{-3}$ , so  $N_\varepsilon = O(J^3)$ . The oscillation of  $\mathbb{G}_h$  between net points satisfies

$$\max_\ell \sup_{|z-z_\ell| \leq \varepsilon} |\mathbb{G}_h(z) - \mathbb{G}_h(z_\ell)| = O_p(J^{3/2} \cdot J^{-3}) = O_p(J^{-3/2}) = o_p(1),$$

and analogously

$$\mathbb{E} \max_\ell \sup_{|z-z_\ell| \leq \varepsilon} |Z_h^{\text{int}}(z) - Z_h^{\text{int}}(z_\ell)| \lesssim J \cdot J^{-3} \sqrt{\log J} = J^{-2} \sqrt{\log J} = o(1).$$

Hence

$$\left| \sup_{z \in \mathcal{Z}} |\mathbb{G}_h(z)| - \max_{\ell} |\mathbb{G}_h(z_\ell)| \right| = o_p(1), \quad \left| \sup_{z \in \mathcal{Z}} |Z_h^{\text{int}}(z)| - \max_{\ell} |Z_h^{\text{int}}(z_\ell)| \right| = o_p(1).$$

*Coupling transfer at net points.* For each  $\ell$ , the linear functional  $\phi(z_\ell)^\top A_{11,h}^{-1} = \tilde{\phi}(z_\ell)^\top$  satisfies  $\|\tilde{\phi}(z_\ell)\|_1 = O(J^\rho)$  by (41). Combining with the strong approximation (40) and Step 2,

$$\begin{aligned} |\mathbb{G}_h(z_\ell) - Z_h^{\text{int}}(z_\ell)| &\leq |\tilde{\phi}(z_\ell)^\top (T^{-1/2} \sum_t \mathbf{S}_{h,t}^{(u)} - Z^\dagger)| + |\phi(z_\ell)^\top r_{h,n}| \\ &\leq \|\tilde{\phi}(z_\ell)\|_1 \cdot \|T^{-1/2} \sum_t \mathbf{S}_{h,t}^{(u)} - Z^\dagger\|_\infty + o_p(J^\rho) \\ &= O(J^\rho) \cdot O_p(\delta) + o_p(J^\rho) = O_p(J^\rho \delta) + o_p(J^\rho), \end{aligned}$$

where  $\delta = \delta(J, L, m, q, \alpha) \lesssim T^{-1/9}(\log J)^{7/6} \rightarrow 0$  under hypothesis (iv). Taking the maximum over  $\ell$ ,

$$\left| \max_{\ell} |\mathbb{G}_h(z_\ell)| - \max_{\ell} |Z_h^{\text{int}}(z_\ell)| \right| \leq \max_{\ell} |\mathbb{G}_h(z_\ell) - Z_h^{\text{int}}(z_\ell)| = O_p(J^\rho \delta) + o_p(J^\rho) = o_p(1)$$

provided  $J^\rho \delta = o(1)$ , which holds under the strengthened sieve growth condition  $J^\rho (\log J)^{7/6} = o(T^{1/9})$ . For B-splines ( $\rho = 0$ ), this reduces to  $(\log J)^{7/6} = o(T^{1/9})$ ; for orthonormal polynomial bases ( $\rho = 1/2$ ), it becomes the condition  $J^{1/2}(\log J)^{7/6} = o(T^{1/9})$  stated in hypothesis (iv) of Theorem 4.

Triangle inequality across the three preceding displays yields

$$\left| \sup_{z \in \mathcal{Z}} |\mathbb{G}_h(z)| - \sup_{z \in \mathcal{Z}} |Z_h^{\text{int}}(z)| \right| = o_p(1).$$

### Step 5 (Identification of the limit and confidence bands).

The auxiliary Gaussian process  $Z_h^{\text{int}}(z)$  from Step 4 has covariance kernel

$$\mathbb{E}[Z_h^{\text{int}}(z) Z_h^{\text{int}}(z')] = \phi(z)^\top A_{11,h}^{-1} \Omega_h^* A_{11,h}^{-1} \phi(z') = \phi(z)^\top V_h^* \phi(z'),$$

using  $V_h^* = A_{11,h}^{-1} \Omega_h^* A_{11,h}^{-1}$  from the convention in Theorem 3. Define

$$Z_h(z) := \phi(z)^\top A_{11,h}^{-1} (\Omega_h^*)^{1/2} \mathcal{N}_J, \quad \mathcal{N}_J \sim \mathcal{N}(0, I_J),$$

so that  $Z_h^{\text{int}} \stackrel{d}{=} Z_h$  as Gaussian processes on  $\mathcal{Z}$ , with covariance kernel  $\mathbb{E}[Z_h(z) Z_h(z')] = \sigma_h^{*2}(z, z') := \phi(z)^\top A_{11,h}^{-1} \Omega_h^* A_{11,h}^{-1} \phi(z')$  at unit scale (matching the asymptotic variance  $\sigma_h^{*2}(z) = \sigma_h^{*2}(z, z)$  in the pointwise CLT of Theorem 2).

*Coupling between  $\sqrt{NT} \sup_z |\hat{g}_h - g_h|$  and  $\sup_z |Z_h|$ .* By Step 4 and the distributional identity  $Z_h^{\text{int}} \stackrel{d}{=} Z_h$ ,

$$\left| \sup_{z \in \mathcal{Z}} |\mathbb{G}_h(z)| - \sup_{z \in \mathcal{Z}} |Z_h(z)| \right| = o_p(1).$$



Substituting  $\mathbb{G}_h(z) = \sqrt{NT} \phi(z)^\top (\hat{b}_h - b_h) = \sqrt{NT}(\hat{g}_h(z) - g_h(z)) + \sqrt{NT} R_h(z)$  and using the triangle inequality  $|\sup_z |\hat{g}_h - g_h| - \sup_z |\phi^\top (\hat{b}_h - b_h)|| \leq \sup_z |R_h(z)|$ ,

$$\sqrt{NT} \sup_{z \in \mathcal{Z}} |\hat{g}_h(z) - g_h(z)| = \sup_{z \in \mathcal{Z}} |\mathbb{G}_h(z)| + O(\sqrt{NT} J^{-\kappa}).$$

Under the undersmoothing condition  $\sqrt{NT} J^{1/2-\kappa} \rightarrow 0$  in hypothesis (iv) of Theorem 4, we have  $\sqrt{NT} J^{-\kappa} = o(J^{-1/2}) = o(1)$ , so the bias contribution is  $o(1)$ . Combining,

$$\left| \sqrt{NT} \sup_{z \in \mathcal{Z}} |\hat{g}_h(z) - g_h(z)| - \sup_{z \in \mathcal{Z}} |Z_h(z)| \right| = o_p(1).$$

The distribution of  $\sup_{z \in \mathcal{Z}} |Z_h(z)|$  is continuous by non-degeneracy of  $\Omega_h^*$  (Assumption 6(2)) and sample-path continuity of  $Z_h(\cdot)$  (established in Step 4 via the Gaussian-process modulus). For any  $\delta > 0$ , the standard coupling-to-Kolmogorov inequality gives

$$\begin{aligned} & \sup_{t \in \mathbb{R}} \left| \mathbb{P} \left( \sqrt{NT} \sup_{z \in \mathcal{Z}} |\hat{g}_h(z) - g_h(z)| \leq t \right) - \mathbb{P} \left( \sup_{z \in \mathcal{Z}} |Z_h(z)| \leq t \right) \right| \\ & \leq \mathbb{P} \left( \left| \sqrt{NT} \sup_z |\hat{g}_h - g_h| - \sup_z |Z_h| \right| > \delta \right) + \sup_{t \in \mathbb{R}} \mathbb{P} \left( \left| \sup_z |Z_h| - t \right| \leq \delta \right). \end{aligned}$$

The first term tends to zero for any fixed  $\delta > 0$  by the preceding display.

The second term is the Lévy concentration function of  $\sup_z |Z_h(z)|$ . By the anti-concentration inequality of Chernozhukov et al. (2015, Theorem 3), applied to the centered Gaussian process  $Z_h$  on  $\mathcal{Z}$  with finite-dimensional projections indexed by the net  $\{z_\ell\}_{\ell=1}^{N_\varepsilon}$  from Step 4 (with  $N_\varepsilon = O(J^3)$ , so  $\log N_\varepsilon = O(\log J)$ ),

$$\sup_{t \in \mathbb{R}} \mathbb{P} \left( \left| \sup_z |Z_h(z)| - t \right| \leq \delta \right) \leq C \delta \sqrt{1 \vee \log(N_\varepsilon \cdot \sigma_{\max}/\delta)} \leq C \delta \sqrt{\log(J/\delta)},$$

where  $\sigma_{\max}^2 := \sup_z \sigma_h^{*2}(z) \leq \lambda_{\max}(\Omega_h^*) \cdot \sup_z \|A_{11,h}^{-1} \phi(z)\|_2^2 = O(J)$ . This is a standard result for maxima of finite-dimensional Gaussian vectors with bounded variance.

Choose  $\delta_n = (\log J)^{-1}$ . Then:

- The first term: by the preceding display,  $\mathbb{P}(|\dots| > \delta_n) \rightarrow 0$  since  $\delta_n \rightarrow 0$  and the LHS is  $o_p(1)$ .
- The second term:  $\delta_n \sqrt{\log(J/\delta_n)} = (\log J)^{-1} \sqrt{\log J + \log \log J} = O((\log J)^{-1/2}) \rightarrow 0$ .

Both terms vanish.

Combining,

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P} \left( \sqrt{NT} \sup_{z \in \mathcal{Z}} |\hat{g}_h(z) - g_h(z)| \leq t \right) - \mathbb{P} \left( \sup_{z \in \mathcal{Z}} |Z_h(z)| \leq t \right) \right| \rightarrow 0,$$

which establishes part (b) of Theorem 4.

Define the critical value

$$c_{h,1-\alpha}^* := \inf \left\{ c > 0 : \mathbb{P} \left( \sup_z |Z_h(z)| \leq c \right) \geq 1 - \alpha \right\}.$$

By the Kolmogorov-distance bound above with  $t = c_{h,1-\alpha}^*$ ,

$$\mathbb{P} \left( \sqrt{NT} \sup_z |\hat{g}_h(z) - g_h(z)| \leq c_{h,1-\alpha}^* \right) \rightarrow 1 - \alpha,$$

i.e., the band  $\hat{g}_h(z) \pm c_{h,1-\alpha}^*/\sqrt{NT}$  has asymptotic coverage  $1 - \alpha$ . □

## D Appendix D: Additional Simulation Results

Figure 7 repeats the Monte Carlo comparison in Figure 1 of the main text, but under a Fourier type DGP

$$g(z) = 0.8 \sin(2\pi z) + 2 \cos(2\pi z) - 0.5 \sin(4\pi z) + \cos(4\pi z),$$

where  $z$  is normalized to  $[0, 1]$  by the transformation  $(z + 4.65)/9.3$ . The pattern closely mirrors the results for the cubic DGP in the main text. The sieve estimator recovers the nonlinear shape of the true Fourier IRF well, whereas the linear-interaction LP continues to impose a monotone profile and therefore fails to capture the curvature in the DGP. Under the Fourier DGP, the sieve estimate is slightly less accurate near the boundaries of the state space at the short horizons and exhibits slightly higher variance, but it still substantially outperforms the linear benchmark. Overall, it leads to the same substantive conclusion: when the underlying state dependence is nonlinear, flexible sieve estimation delivers meaningful gains.

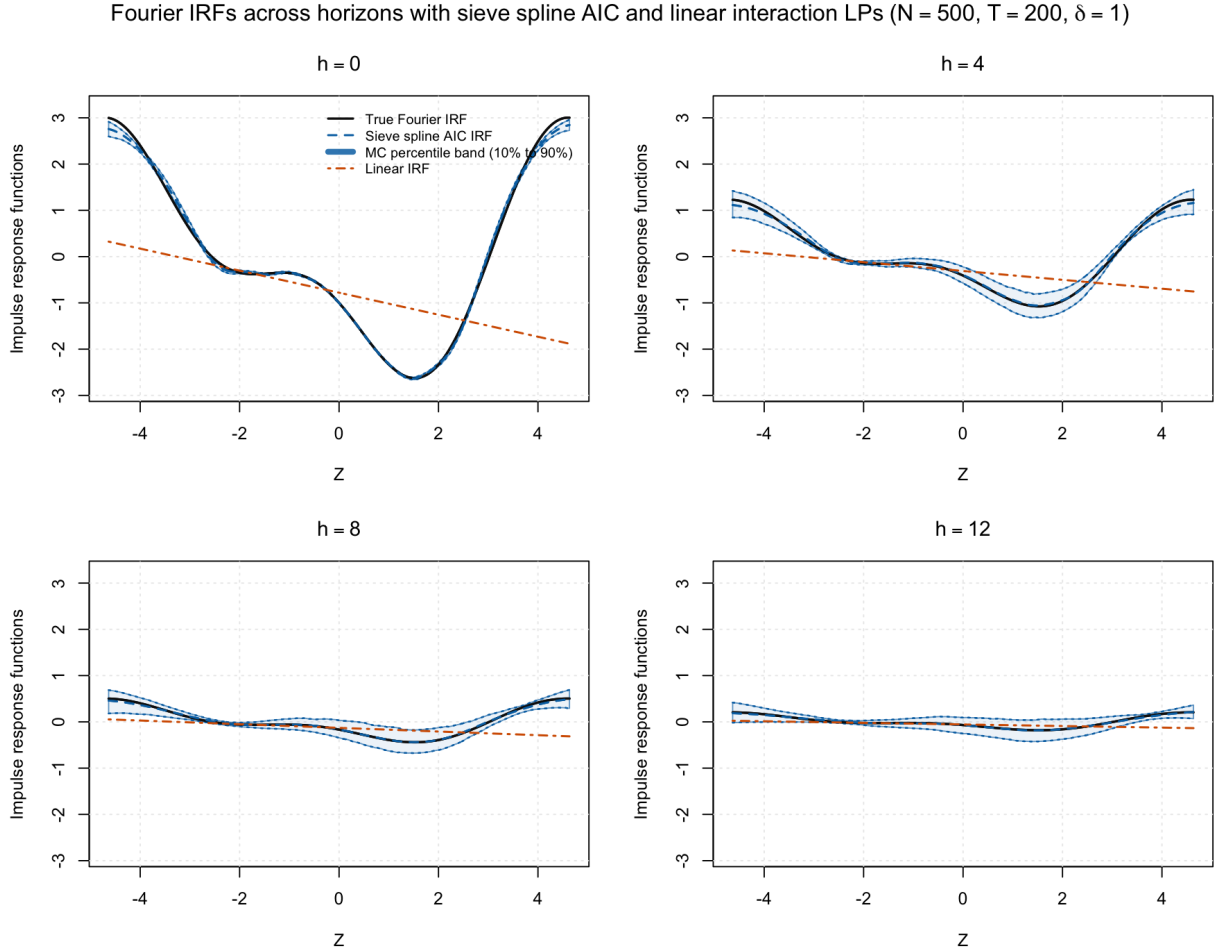


Figure 7: Monte Carlo comparison of sieve and linear IRFs under the Fourier DGP

Table 6: Uniform-band performance of sieve across selector choices at horizons  $h \in \{0, 4\}$  under nominal coverage  $1 - \alpha = 0.95$

| $N$                                        | $T$ | Oracle |      | AIC  |       | GCV  |       | LASSO |       |
|--------------------------------------------|-----|--------|------|------|-------|------|-------|-------|-------|
|                                            |     | Cov    | Wid  | Cov  | Wid   | Cov  | Wid   | Cov   | Wid   |
| <i>Panel A: Horizon <math>h = 0</math></i> |     |        |      |      |       |      |       |       |       |
| 1                                          | 40  | 0.69   | 3.32 | 0.55 | 3.19  | 0.56 | 3.14  | 0.60  | 3.54  |
|                                            | 120 | 0.78   | 2.69 | 0.62 | 2.88  | 0.62 | 2.85  | 0.61  | 4.08  |
|                                            | 200 | 0.82   | 2.36 | 0.68 | 2.61  | 0.68 | 2.61  | 0.68  | 4.29  |
| 100                                        | 40  | 0.87   | 0.85 | 0.73 | 1.02  | 0.73 | 1.01  | 0.65  | 2.78  |
|                                            | 120 | 0.91   | 0.66 | 0.84 | 0.84  | 0.84 | 0.84  | 0.80  | 2.62  |
|                                            | 200 | 0.92   | 0.56 | 0.87 | 0.69  | 0.87 | 0.69  | 0.80  | 2.48  |
| 500                                        | 40  | 0.84   | 0.52 | 0.76 | 0.67  | 0.76 | 0.67  | 0.72  | 2.10  |
|                                            | 120 | 0.93   | 0.40 | 0.83 | 0.53  | 0.83 | 0.53  | 0.82  | 1.95  |
|                                            | 200 | 0.94   | 0.35 | 0.84 | 0.47  | 0.84 | 0.47  | 0.80  | 1.79  |
| <i>Panel B: Horizon <math>h = 4</math></i> |     |        |      |      |       |      |       |       |       |
| 1                                          | 40  | 0.57   | 8.84 | 0.41 | 13.09 | 0.45 | 10.90 | 0.55  | 10.54 |
|                                            | 120 | 0.82   | 4.72 | 0.63 | 5.51  | 0.68 | 5.22  | 0.72  | 8.91  |
|                                            | 200 | 0.78   | 3.87 | 0.66 | 4.56  | 0.68 | 4.32  | 0.70  | 8.21  |
| 100                                        | 40  | 0.83   | 1.46 | 0.74 | 1.70  | 0.74 | 1.70  | 0.68  | 5.42  |
|                                            | 120 | 0.88   | 1.04 | 0.82 | 1.34  | 0.82 | 1.34  | 0.82  | 4.25  |
|                                            | 200 | 0.89   | 0.89 | 0.83 | 1.16  | 0.83 | 1.16  | 0.83  | 3.74  |
| 500                                        | 40  | 0.80   | 0.88 | 0.72 | 1.12  | 0.72 | 1.12  | 0.76  | 3.50  |
|                                            | 120 | 0.87   | 0.66 | 0.79 | 0.83  | 0.79 | 0.83  | 0.86  | 3.04  |
|                                            | 200 | 0.92   | 0.55 | 0.83 | 0.78  | 0.83 | 0.78  | 0.85  | 2.70  |

Table 7: Uniform-band performance of sieve across selector choices at horizons  $h \in \{8, 12\}$  under nominal coverage  $1 - \alpha = 0.95$

|                                             |     | Oracle |      | AIC  |       | GCV  |      | LASSO |       |
|---------------------------------------------|-----|--------|------|------|-------|------|------|-------|-------|
| $N$                                         | $T$ | Cov    | Wid  | Cov  | Wid   | Cov  | Wid  | Cov   | Wid   |
| <i>Panel A: Horizon <math>h = 8</math></i>  |     |        |      |      |       |      |      |       |       |
| 1                                           | 120 | 0.74   | 5.28 | 0.63 | 9.75  | 0.65 | 6.78 | 0.74  | 10.10 |
|                                             | 200 | 0.84   | 4.29 | 0.71 | 6.09  | 0.76 | 5.25 | 0.74  | 7.87  |
| 100                                         | 40  | 0.75   | 1.58 | 0.67 | 2.07  | 0.67 | 2.06 | 0.70  | 7.43  |
|                                             | 120 | 0.87   | 1.12 | 0.80 | 1.48  | 0.80 | 1.48 | 0.86  | 4.64  |
|                                             | 200 | 0.91   | 0.94 | 0.83 | 1.27  | 0.83 | 1.26 | 0.78  | 3.98  |
| 500                                         | 40  | 0.81   | 0.96 | 0.72 | 1.32  | 0.72 | 1.32 | 0.69  | 4.05  |
|                                             | 120 | 0.86   | 0.69 | 0.83 | 0.88  | 0.83 | 0.88 | 0.86  | 3.18  |
|                                             | 200 | 0.89   | 0.58 | 0.81 | 0.77  | 0.81 | 0.77 | 0.83  | 2.79  |
| <i>Panel B: Horizon <math>h = 12</math></i> |     |        |      |      |       |      |      |       |       |
| 1                                           | 120 | 0.72   | 5.69 | 0.54 | 12.06 | 0.59 | 6.67 | 0.67  | 10.29 |
|                                             | 200 | 0.82   | 4.52 | 0.69 | 7.76  | 0.69 | 5.44 | 0.79  | 9.12  |
| 100                                         | 40  | 0.74   | 1.71 | 0.67 | 2.16  | 0.67 | 2.16 | 0.71  | 9.78  |
|                                             | 120 | 0.89   | 1.14 | 0.81 | 1.51  | 0.81 | 1.51 | 0.82  | 4.99  |
|                                             | 200 | 0.90   | 0.94 | 0.82 | 1.19  | 0.82 | 1.19 | 0.83  | 4.07  |
| 500                                         | 40  | 0.72   | 1.02 | 0.62 | 1.37  | 0.62 | 1.36 | 0.68  | 4.62  |
|                                             | 120 | 0.91   | 0.70 | 0.84 | 0.93  | 0.84 | 0.93 | 0.87  | 3.09  |
|                                             | 200 | 0.89   | 0.59 | 0.80 | 0.79  | 0.80 | 0.79 | 0.86  | 2.65  |

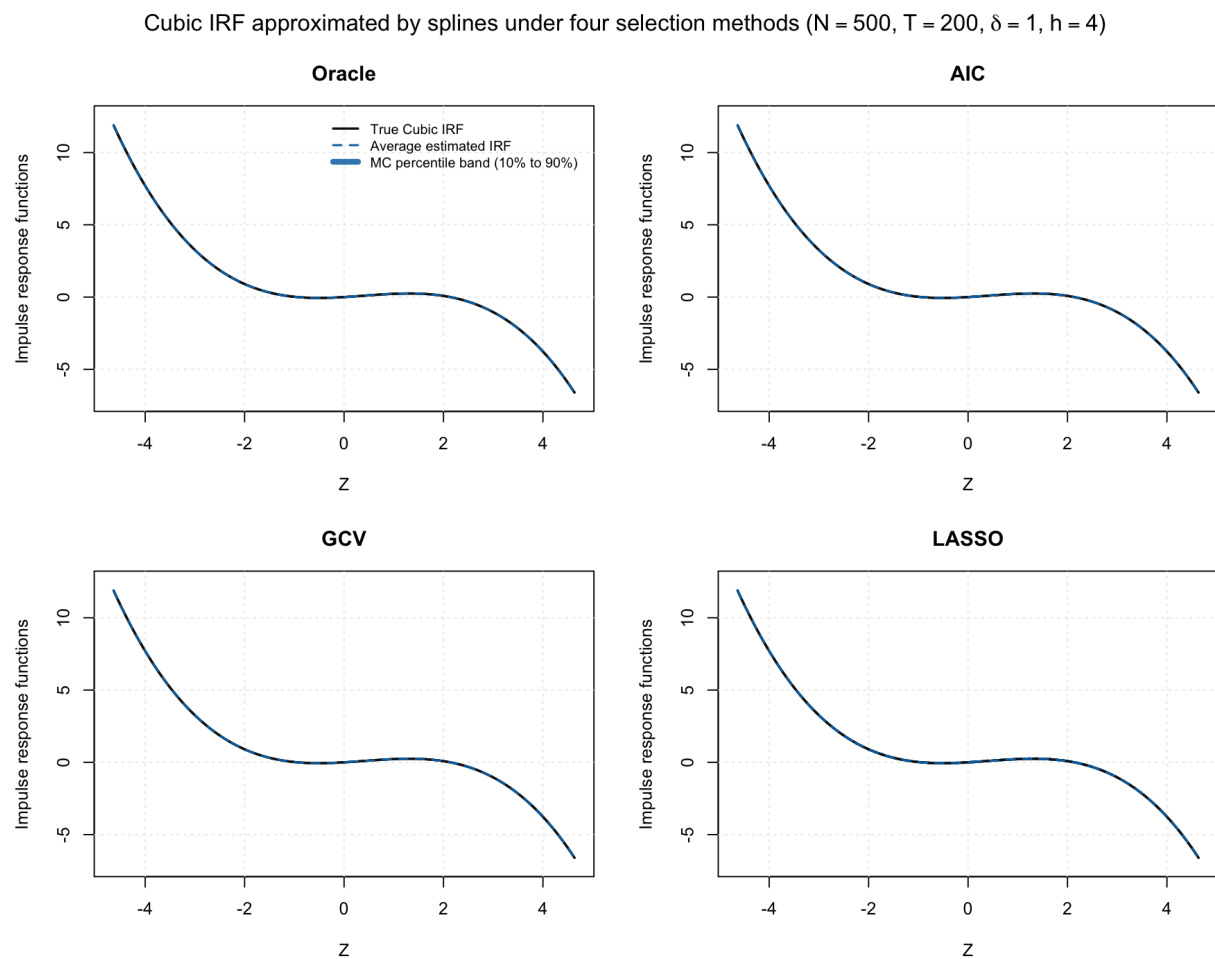


Figure 8: Monte Carlo comparison of sieve IRFs across selector choices under the cubic DGP over the full-support

## References

- Almuzara, M., Arellano, M., Blundell, R., and Bonhomme, S. (2025). Nonlinear micro income processes with macro shocks. Technical Report 1162, Federal Reserve Bank of New York.
- Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–858.
- Atkeson, A. G., Eisfeldt, A. L., and Weill, P.-O. (2017). Measuring the financial soundness of us firms, 1926–2012. *Research in Economics*, 71(3):613–635.
- Auclert, A. (2019). Monetary policy and the redistribution channel. *American Economic Review*, 109(6):2333–2367.
- Bayer, C. and Luetticke, R. (2020). Solving discrete time heterogeneous agent models with aggregate risk and many idiosyncratic states by perturbation. *Quantitative Economics*, 11(4):1253–1288.
- Bousquet, P. (2025). Uncovering nonlinearities with local projections. Unpublished manuscript.
- Burkholder, D. L. (1988). Sharp inequalities for martingales and stochastic integrals. In *Colloque Paul Lévy sur les Processus Stochastiques*, volume 157–158 of *Astérisque*, pages 75–94. Société Mathématique de France.
- Chen, X. and Christensen, T. M. (2015). Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions. *Journal of Econometrics*, 188(2):447–465.
- Chen, X., Liao, Y., and Wang, W. (2024). Inference on time series nonparametric conditional moment restrictions using nonlinear sieves. *Journal of Econometrics*, page 105920.
- Chernozhukov, V., Chetverikov, D., and Kato, K. (2014). Gaussian approximation of suprema of empirical processes. *The Annals of Statistics*, 42(4):1564 – 1597.
- Chernozhukov, V., Chetverikov, D., and Kato, K. (2015). Comparison and anti-concentration bounds for maxima of gaussian random vectors. *Probability Theory and Related Fields*, 162(1–2):47–70.
- Chernozhukov, V., Chetverikov, D., and Kato, K. (2016). Empirical and multiplier bootstraps for suprema of empirical processes of increasing complexity, and related gaussian couplings. *Stochastic Processes and their Applications*, 126(12):3632–3651.
- Chernozhukov, V., Chetverikov, D., and Kato, K. (2017). Central limit theorems and bootstrap in high dimensions. *The Annals of Probability*, 45(4):2309–2352.
- Cloyne, J., Ferreira, C., Froemel, M., and Surico, P. (2023). Monetary policy, corporate finance, and investment. *Journal of the European Economic Association*, 21(6):2586–2634.

- Cui, W., Xie, R. C., and Zhang, R. (2024). Risk-taking with financing constraints. Working paper.
- David, J. and Gourio, F. (2023). The rise of intangible investment and the transmission of monetary policy. *Chicago Fed Letter*, 482.
- DeVore, R. A. and Lorentz, G. G. (1993). *Constructive Approximation*, volume 303 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin.
- Duffie, D. (2011). *Measuring corporate default risk*. Oxford University Press.
- Duffie, D., Saita, L., and Wang, K. (2007). Multi-period corporate default prediction with stochastic covariates. *Journal of financial economics*, 83(3):635–665.
- Gilchrist, S. and Zakrajšek, E. (2012). Credit spreads and business cycle fluctuations. *American economic review*, 102(4):1692–1720.
- Gonçalves, S., Herrera, A. M., Kilian, L., and Pesavento, E. (2024a). Nonparametric local projections. Research Department Working Paper 2414, Federal Reserve Bank of Dallas. Latest posted version: November 15, 2024.
- Gonçalves, S., Herrera, A. M., Kilian, L., and Pesavento, E. (2024b). State-dependent local projections. *Journal of Econometrics*, 244(2):105702.
- Ibragimov, I. A. and Linnik, Y. V. (1971). *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff, Groningen.
- Jeenas, P. (2023). Firm balance sheet liquidity, monetary policy shocks, and investment dynamics. *Working Paper*.
- Jordà, Ò. and Taylor, A. M. (2025). Local projections. *Journal of Economic Literature*, 63(1):59–110.
- Kitagawa, T., Wang, W., and Xu, M. (2025). Nonlinearity in dynamic causal effects: Making the bad into the good, and the good into the great? *Journal of Business & Economic Statistics*, 43(4):770–777.
- Kolesár, M. and Plagborg-Møller, M. (2024). Dynamic causal effects in a nonlinear world: the good, the bad, and the ugly. *arXiv preprint arXiv:2411.10415*.
- Lu, J., Kolar, M., and Liu, H. (2020). Kernel meets sieve: Post-regularization confidence bands for sparse additive model. *Journal of the American Statistical Association*, 115(532):2084–2099.
- Newey, W. K. (1997). Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics*, 79:147–168.
- Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.



- Ottonello, P. and Winberry, T. (2020). Financial heterogeneity and the investment channel of monetary policy. *Econometrica*, 88(6):2473–2502.
- 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.
- Rio, E. (2009). Moment inequalities for sums of dependent random variables under projective conditions. *Journal of Theoretical Probability*, 22(1):146–163.
- Schaefer, S. M. and Strebulaev, I. A. (2008). Structural models of credit risk are useful: Evidence from hedge ratios on corporate bonds. *Journal of Financial Economics*, 90(1):1–19.
- Schumaker, L. L. (1981). *Spline Functions: Basic Theory*. John Wiley & Sons, New York.
- van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer-Verlag, New York.
- Winkler, V. (2026). When and why state-dependent local projections work. *arXiv preprint arXiv:2601.01622*.
- Zhang, D. and Wu, W. B. (2017). Gaussian approximation for high dimensional time series. *The Annals of Statistics*, 45(5):1895 – 1919.