Econometric Theory, **42**, 2026, 1151–1193.
doi:10.1017/S0266466625100157

SLOW MOVERS IN PANEL DATA

YUYA SASAKI
Vanderbilt University

TAKUYA URA
University of California Davis

Panel data often contain stayers and *slow movers*. The literature proposes an estimator for the average partial effects (APEs) for this setting without a formal theory. The literature is also silent about inference in the presence of stayers and many slow movers. We contribute to this state of the art. First, we develop an asymptotic theory to guarantee that such an estimator is consistent in the presence of stayers and slow movers. Second, we propose its standard error. Third, we relax the existing assumption to allow for "many" slow movers. Fourth, we generalize the existing estimator. Fifth, we establish that this generalized estimator can achieve larger extents of bias reduction and hence faster convergence rates. Simulation studies demonstrate that the conventional 95% confidence interval covers the true value of the APE with 37%–93% frequencies whereas our proposed one achieves 93%–96% coverage frequencies. Using the U.S. Panel Study of Income Dynamics, we find that estimates of the marginal propensity to consume based on our generalized estimator remarkably differ in values from those of the existing estimators. Moreover, the generalized estimator achieves more than three times as small standard errors as those of the existing robust estimator.

1. INTRODUCTION

Panel datasets used in economics often contain many individuals/families/firms with no or little within-variations. We refer to units with no within-variation as stayers, units with little within-variation as *slow movers*,[1] and units with non-trivial

We have benefited from very useful comments by Peter C. B. Phillips (Editor), Martin Weidner (Co-Editor), two anonymous referees, Cheng Hsiao, Ivana Komunjer, M. Hashem Pesaran, Valentin Verdier, seminar participants at CREST-PSE, Georgetown University, London School of Economics, Northwestern University, Ohio State University, Singapore Management University, Texas A&M University, University of Bologna, University of California at Santa Cruz, University of Glasgow, University of North Carolina at Chapel Hill, and the University of Southern California, and participants at numerous conferences. Y.S. thanks Brian and Charlotte Grove for financial support. We are responsible for all the remaining errors. Address correspondence to Yuya Sasaki, Department of Economics, Vanderbilt University, Nashville, TN, USA, e-mail: yuya.sasaki@vanderbilt.edu.

[1] We thank Hashem Pesaran for providing us with the terminology of slow movers. A mathematical characterization of "slow movers" is presented in Section 2.3. In that section, we define the slow movers as the cross-sectional observations for which the determinant of the covariate matrix is less than a threshold value h (in absolute value), where h is a constant depending on the sample size.

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

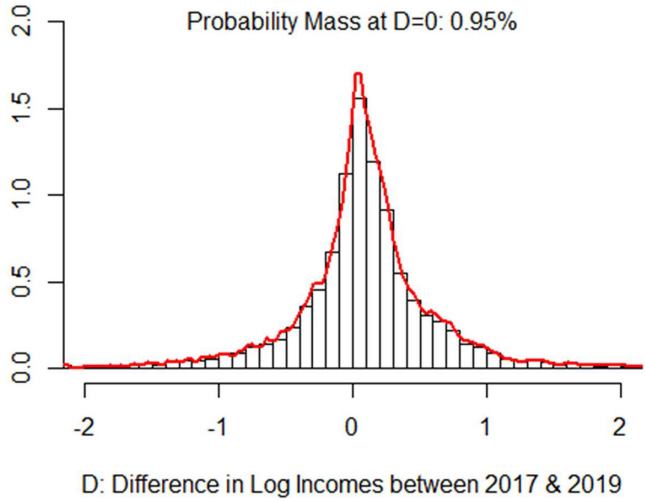


FIGURE 1. Empirical distribution of the difference D in log total family incomes between 2017 and 2019 in the U.S. PSID. The chart displays a histogram and a kernel density plot. About 0.95% of the sample has exactly zero difference ($D = 0$). The displayed kernel density plot does not use the information about the mass with $D = 0$, and it therefore represents the continuous part of the distribution of D. The sample consists of 7,814 households with positive total family income less than or equal to 250K U.S. dollars in both 2017 and 2019.

within-variation as movers. Figure 1 illustrates a distribution of the log total family income in the U.S. Panel Study of Income Dynamics (PSID). In this figure, about 0.95% of the sample are stayers, and the large spike of the kernel density plot also indicates that there are many slow movers.[2]

Also illustrated are the eight possible types of distributions displayed in Figure 2. Panels (a), (c), (e), and (g) illustrate cases with no stayers, while Panels (b), (d), (f), and (h) illustrate cases with stayers. Panels (a) and (b) illustrate cases with moderate densities of slow movers, while Panels (c)–(h) illustrate cases with more slow movers. Although Figure 1 most closely resembles Figure 2(d) among the eight panels in Figure 2, we cannot rule out Figure 2(f) or (h) as potential population distributions to generate the empirical distribution displayed in Figure 1.[3]

[2]This kernel density plot does not use the information about the mass of $D = 0$.

[3]The red curve shown in Figure 1 is continuous indeed, but this is an artifact of the convolution with a continuous kernel. We cannot rule out a discontinuous density as in Figure 2(f) as a data-generating process for Figure 1. Similarly, the red curve shown in Figure 1 is bounded indeed, but this is because any finite sample kernel density estimate is finite. We cannot rule out an unbounded density as in Figure 2(h) as a data-generating process for Figure 1.

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

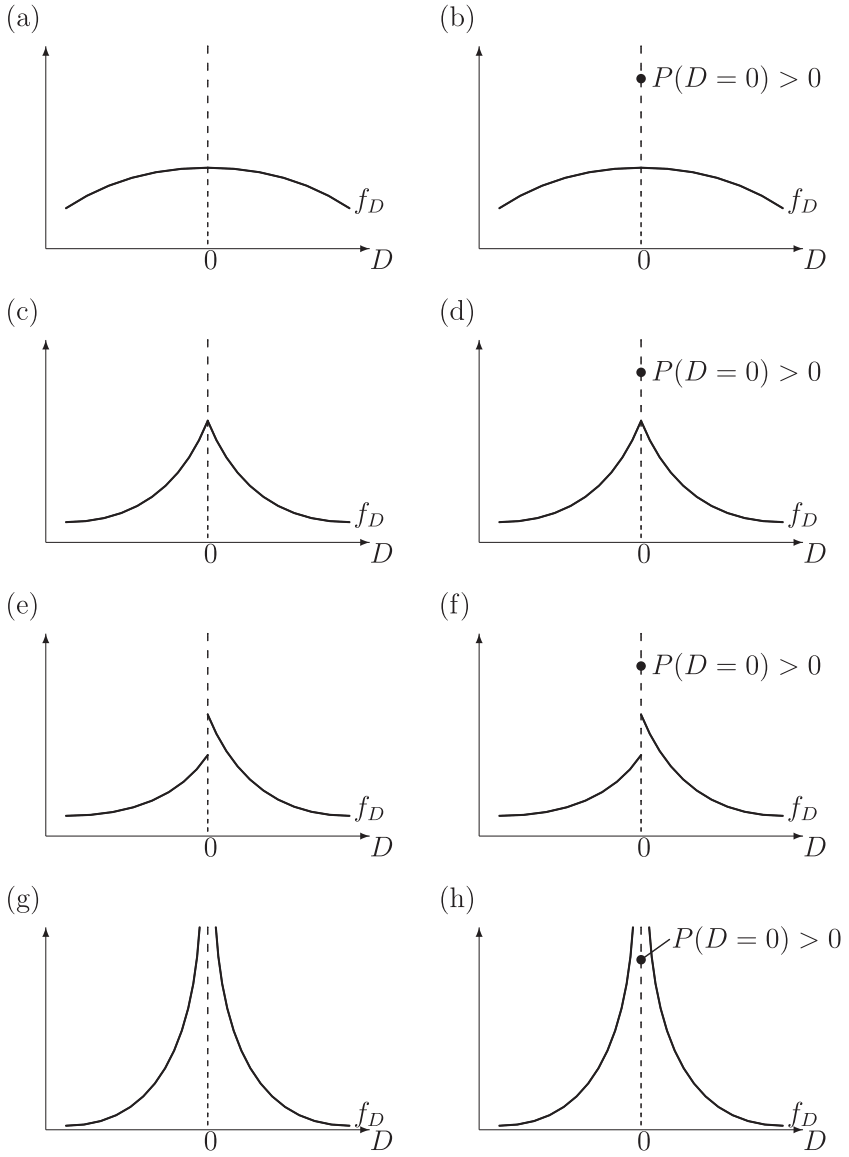


FIGURE 2. Illustration of eight distribution models to which the method of robust inference proposed in this article robustly applies in a unified manner: (a) a smooth, bounded, and locally symmetric distribution of D with no mass of $D = 0$; (b) a smooth, bounded, and locally symmetric distribution of D with a mass of $D = 0$; (c) a non-smooth, bounded, and locally symmetric distribution of D with no mass of $D = 0$; (d) a non-smooth, bounded, and locally symmetric distribution of D with a mass of $D = 0$; (e) a non-smooth, bounded, and locally asymmetric distribution of D with no mass of $D = 0$; (f) a non-smooth, bounded, and locally asymmetric distribution of D with a mass of $D = 0$; (g) an unbounded distribution of D with no mass of $D = 0$; and (h) an unbounded distribution of D with a mass of $D = 0$. The densities of the absolutely continuous part of the distribution are represented by solid curves, while the probability mass of the singular part is represented by a black circle.

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

Therefore, an ideal econometric method should robustly accommodate all the data-generating processes represented by the eight panels in Figure 2.

We highlight issues concerning stayers and slow movers in the context of correlated random coefficient models in short panels (Chamberlain, 1992). With no within-variations, we cannot identify the partial effects for the stayers. With small within-variations, we can only weakly identify the partial effects for slow movers. There is no identification problem for the movers. In the conventional constant-coefficient panel models, the partial effects identifiable from the movers identify the population parameter thanks to its homogeneity. On the other hand, in random coefficient models, the partial effects identifiable from the movers do not identify the population average partial effects (APEs) due to the heterogeneity in coefficients—the APEs among the movers generally differ from those among the stayers and slow movers. The presence of many slow movers, therefore, incurs identification challenges in the correlated random coefficient models unlike the constant coefficient models.

As an important benefit, correlated random coefficient models in short panels can accommodate endogenous regressors and nonseparable error terms at the same time. We briefly demonstrate this feature here by introducing a simplified model—we will consider a more general model in the next section. Suppose that we observe a panel dataset of an outcome variable Y_t and a scalar regressor X_t with two time periods ($t = 1, 2$). Consider the conditional expectation $E[Y_t \mid X_1, X_2] = (\beta_0(X_1, X_2) + \delta_{0t}) + X_t(\beta_1(X_1, X_2) + \delta_{1t})$. This model allows for time and individual unobserved heterogeneity in both the intercept and slope on the right-hand side of the above equation. Notably, the unknown individual heterogeneity $(\beta_0(X_1, X_2), \beta_1(X_1, X_2))$ can be correlated with the observed covariates (X_1, X_2) since it is a function of (X_1, X_2).

Chamberlain (1992) establishes a $\sqrt{N}$-consistent estimation of the APEs in the absence of stayers and slow movers. Graham and Powell (2012) investigate the identification and estimation of the APEs in the presence of slow movers. Such a case with slow movers considered by Graham and Powell (2012) is characterized by Figure 2(a). Graham and Powell (2012) also propose their second estimator which may potentially work in the presence of stayers as in Figure 2(b), but no formal theory has been developed for this second estimator. The literature is also silent about estimation and inference under the cases with "many" slow movers represented by Figure 2(c)–(h).

We make the following contributions to this state of the art in the literature. First, we develop an asymptotic theory for the second estimator of Graham and Powell (2012), for which formal theory has been missing in the literature. Second, as a consequence, we propose its standard error to allow for inference based on this estimator. Third, we show that this estimator in fact works under weaker conditions than those considered in the existing literature. In particular, we allow for many slow movers as in Figure 2(c)–(h), in addition to the two baseline cases, Figure 2(a) and (b), considered in the literature. Fourth, we generalize the second estimator of Graham and Powell (2012) as a local linear estimator to a higher-order

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

local polynomial estimator. Fifth, we show that this generalized estimator achieves larger extents of bias reduction and hence faster convergence rates.[4]

The default estimator of Graham and Powell (2012) is based on trimming the slow movers. While trimming is useful for mitigating variances from weak identification, it incurs the price of biased estimates for failing to account for the partial effects among the slow movers. This price may be negligible when there are only few slow movers as in Figure 2(a) and (b). If there are many slow movers as in Figure 1 or Figure 2(c)–(h), on the other hand, then the trimming bias can be too large to ignore. We therefore infer the partial effects of the trimmed slow movers by extrapolating the conditional expectation function of the partial effects from the movers to the slow movers. If a researcher is willing to assume that this conditional expectation function is smooth, then a higher-order polynomial approximation of this function leads to more precise bias correction, and this idea motivates our proposal of the local polynomial estimator as a device of improved bias reduction.

This article is related to several branches of the econometrics literature. Random coefficient panel models have been extensively studied at least since Swamy (1970)—see Hsiao (2014, Sect. 6.2) and Hsiao and Pesaran (2008) for a review. Among others, we particularly consider *correlated* random coefficient panel models studied by Chamberlain (1982, 1992), Wooldridge (2005), Murtazashvili and Wooldridge (2008), Arellano and Bonhomme (2012), Graham and Powell (2012), Fernández-Val and Lee (2013), Graham et al. (2018), Laage (2024), Verdier (2020), and Holland et al. (2023). Also related is the literature on non-separable panel models with correlated random effects, such as Altonji and Matzkin (2005), Hoderlein and White (2012), Chernozhukov et al. (2013), and Chernozhukov et al. (2015). The debiasing feature of our estimator is also related to the bias correction methods (e.g., Calonico, Cattaneo, and Titiunik, 2014) in kernel estimation. On the technical side, we use Karamata's theorem (Resnick, 2007, Thm. 2.1) from the literature on heavy-tailed distributions as a main technical device to develop nonparametric econometric theories under possible unboundedness, local asymmetry, or non-smoothness of a density function.

The article proceeds as follows. Section 2 introduces the model and briefly reviews the identification and estimation strategies in the existing literature. Section 3 proposes the method of unified estimation and robust inference. Section 4 provides theoretical results for the proposed method. We present Monte Carlo simulations and an empirical application in Sections 5 and 6, respectively. Sections 2–6 focus on the case in which the number T of time periods equals the number p of regressors. Section 7 presents an extension to the case where T is larger than p. Section 8 concludes. The appendix collects proofs of the main theorems, auxiliary lemmas, and their proofs.

[4]This feature appears in Sasaki and Ura (2022) which uses a similar bias correction technique in a different context. We cannot apply their results directly in the framework of this article. In particular, the assumptions of Sasaki and Ura (2022) exclude the presence of stayers.

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

2. CORRELATED RANDOM COEFFICIENT PANEL DATA MODELS

Following Chamberlain (1992), we consider the correlated random coefficient panel data model

$$Y_t = \mathbf{X}_t' b_t(A, U_t) \tag{1}$$

for $t = 1, \ldots, T$, where Y_t denotes a scalar outcome, $\mathbf{X}_t$ denotes a p-dimensional column vector of regressors, A denotes a correlated unobserved heterogeneity, and U_t denotes an idiosyncratic error. The main parameter of interest is the APE, $E[b_t(A, U_t)]$, i.e., the effect of an exogenous change in $\mathbf{X}_t$ on Y_t. With these notations, the panel model (1) is equipped with the following assumption.

Assumption 1. (i) U_t can be decomposed as $U_t = (U_{1t}, U_{2t})$, and $b_t(A, U_t)$ is decomposed as $b_t(A, U_t) = b^*(A, U_t) + d_t(U_{2t})$, for $t = 1, \ldots, T$. (ii) Given $(\mathbf{X}_1, \ldots, \mathbf{X}_T, A)$, the random variables U_t and U_s have the same conditional distribution. (iii) U_{2t} and $(\mathbf{X}_1, \ldots, \mathbf{X}_T, A)$ are independent. (iv) $E[b_t(A, U_t) \mid \mathbf{X}_1, \ldots, \mathbf{X}_T]$ exists for $t = 1, \ldots, T$. (v) $E[d_1(U_{21})] = 0$.

This assumption is standard in the literature (e.g., Chamberlain, 1992; Graham and Powell, 2012). Under this assumption, the parameter of interest, $E[b_t(A, U_t)]$, can be decomposed as $E[b_t(A, U_t)] = E[b_1(A, U_1)] + E[d_t(U_{2t})]$. The first part $E[b_1(A, U_1)]$ is the APE at $t = 1$. The second part $E[d_t(U_{2t})]$ is the mean time shift from the first period to the t-th period in the APE. Therefore, the identification for $E[b_t(A, U_t)]$ boils down to the identification for $(\boldsymbol{\beta}', \boldsymbol{\delta}')'$, where $\boldsymbol{\beta} = E[b_1(A, U_1)]$ and $\boldsymbol{\delta} = (E[d_2(U_{22})]', \ldots, E[d_T(U_{2T})]')'$.

Graham and Powell (2012) provide the identification of $(\boldsymbol{\beta}', \boldsymbol{\delta}')'$. We focus on the case of $T = p$ in which $E[b_t(A, U_t)]$ cannot be regularly identified—an extension to the case of $T > p$ is discussed in Section 7. Assumption 1 and the panel model (1) imply

$$E[\mathbf{X}^*\mathbf{Y} \mid \mathbf{X}] = \mathbf{X}^*\mathbf{W}\boldsymbol{\delta} + D\boldsymbol{\beta}(\mathbf{X}) \text{ with } \boldsymbol{\beta} = E[\boldsymbol{\beta}(\mathbf{X})], \tag{2}$$

where $\mathbf{Y} = (Y_1, \ldots, Y_T)'$, $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_T)'$, $D = \det(\mathbf{X})$, $\boldsymbol{\beta}(\mathbf{X}) = E[b^*(A, U_t)|\mathbf{X}]$, $\mathbf{X}^*$ is the adjoint of $\mathbf{X}$, and

$$\mathbf{W} = \begin{pmatrix} 0_p & \mathbf{X}_2 & & 0_p \\ & & \ddots & \\ 0_p & 0_p & & \mathbf{X}_T \end{pmatrix}'.$$

Recall that the adjoint $\mathbf{X}^*$ of $\mathbf{X}$ is defined by $\left((-1)^{i+j}|\mathbf{X}_{-i,-j}|\right)_{i,j}$, where $|\mathbf{X}_{-i,-j}|$ is the determinant of the (i,j)-minor $\mathbf{X}_{-i,-j}$ of $\mathbf{X}$ obtained by deleting row i and column j of $\mathbf{X}$. Taking the conditional expectation of (2) given $D = 0$, one identifies $\boldsymbol{\delta}$ by

$$\boldsymbol{\delta} = E[(\mathbf{X}^*\mathbf{W})'\mathbf{X}^*\mathbf{W} \mid D = 0]^{-1} E[(\mathbf{X}^*\mathbf{W})'\mathbf{X}^*\mathbf{Y} \mid D = 0]. \tag{3}$$

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

The identification analysis for $\boldsymbol{\beta}$ in Graham and Powell (2012) branches into two separate cases depending on the probability of the event $D = 0$. The first case, $P(D = 0) = 0$, represented by Figure 2(a), is presented in Section 2.1. The second case, $P(D = 0) \neq 0$, represented by Figure 2(b), is presented in Section 2.2. In both cases, the conditional distribution of D given $D \neq 0$ needs to be continuous near 0 (cf. Assumption 2).

2.1. Identification and Estimation Under $P(D = 0) = 0$

When $P(D = 0) = 0$, the parameter $\boldsymbol{\beta}$ is equal to $E[\boldsymbol{\beta}(\mathbf{X}) \mid D \neq 0]$. Hence, (2) yields the identifying equality $\boldsymbol{\beta} = E[D^{-1}\mathbf{X}^*(\mathbf{Y} - \mathbf{W}\boldsymbol{\delta})|D \neq 0]$. Taking a regularized sample counterpart, Graham and Powell (2012, Prop. 1.2) propose to estimate $\boldsymbol{\beta}$ by

$$\hat{\boldsymbol{\beta}}^M = \frac{E_N[\mathbf{1}\{|D| > h_N\}D^{-1}\mathbf{X}^*(\mathbf{Y} - \mathbf{W}\hat{\boldsymbol{\delta}})]}{E_N[\mathbf{1}\{|D| > h_N\}]}, \tag{4}$$

where $\hat{\boldsymbol{\delta}}$ is an estimator of $\boldsymbol{\delta}$, E_N is the sample mean operator, and $h_N > 0$ is a tuning parameter that converges to 0 as $N \to \infty$.[5] Here, the superscript "M" in $\hat{\boldsymbol{\beta}}^M$ is used to emphasize that it estimates the APE for the "movers" at $t = 1$. The standard error for this estimator $\hat{\boldsymbol{\beta}}^M$ and the supporting asymptotic theory are presented by Graham and Powell (2012) under the assumption covering the case with a smooth and bounded density of D without a mass of stayers represented by Figure 2(a).

2.2. Identification and Estimation Under $P(D = 0) > 0$

Graham and Powell (2012, Sect. 3.2) separately consider the case with a point mass of stayers, i.e., $P(D = 0) > 0$. They derive the following constructive identifying formula:

$$\boldsymbol{\beta} = P(D \neq 0)E[\boldsymbol{\beta}(\mathbf{X}) \mid D \neq 0] + P(D = 0)\mathbf{m}^{(1)}(0), \tag{5}$$

where $\mathbf{m}^{(k)}(u)$ is the k-th derivative of the function $\mathbf{m}(u)$ defined by

$$\mathbf{m}(u) = E[\mathbf{X}^*(\mathbf{Y} - \mathbf{W}\boldsymbol{\delta}) \mid D = u]. \tag{6}$$

Intuitively, the second term of (5) extrapolates the conditional partial effects $\mathbf{m}^{(1)}(u)$ of the movers $u > 0$ to the limit subpopulation $u = 0$ of the stayers to infer the partial effects for the stayers. Based on (5), they propose to estimate $\boldsymbol{\beta}$ by the mixture estimator

$$\hat{\boldsymbol{\beta}}_1 = E_N[\mathbf{1}\{|D| > h_N\}]\hat{\boldsymbol{\beta}}^M + E_N[\mathbf{1}\{|D| \leq h_N\}]\hat{\mathbf{m}}^{(1)}(0), \tag{7}$$

[5] Although the framework is different, $\hat{\boldsymbol{\beta}}^M$ can be regarded as the mean group estimator (cf. Pesaran and Smith, 1995; Hsiao, Pesaran, and Tahmiscioglu, 1999). We can interpret $D^{-1}\mathbf{X}^*(\mathbf{Y} - \mathbf{W}\hat{\boldsymbol{\delta}})$ as the individual-specific estimator for $\boldsymbol{\beta}$.

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

where $\hat{\mathbf{m}}^{(1)}(0)$ is an estimator for $\mathbf{m}^{(1)}(0)$. Such an estimator is expected to be effective in models with a mass of stayers represented by Figure 2(b). However, with no standard error and no supporting theories, the literature has not formally confirmed if $\hat{\boldsymbol{\beta}}_1$ indeed works.

2.3. Our Contributions to This Literature

As discussed in the previous two sections, there exist two separate estimators for $\boldsymbol{\beta}$ depending on whether the probability of $D = 0$ is zero or not. For the case of $P(D = 0) = 0$ as represented by Figure 2(a), the estimator $\hat{\boldsymbol{\beta}}^M$ given in (4) is consistent for $\boldsymbol{\beta}$. The literature also proposes its standard error along with formal asymptotic theories. On the other hand, in the presence of a mass of stayers, i.e., $P(D = 0) > 0$ as represented by Figure 2(b), it is speculated that the estimator $\hat{\boldsymbol{\beta}}_1$ might work, but the literature is silent about its standard error and formal supporting theories for whether it indeed works.

In light of this state of the art, we make the following five contributions to this literature. First, we provide asymptotic theories for the estimator $\widehat{\boldsymbol{\beta}}_1$. Second, as a consequence, we prove its standard error to enable inference for $\boldsymbol{\beta}$ in the presence of the slow movers (characterized by $0 < |D| < h_N$) and stayers (characterized by $|D| = 0$) as in Figure 2(a) and (b). Third, while the existing literature does not allow for the presence of many slow movers, our theory is built on more relaxed assumptions by allowing for "many" slow movers to accommodate all the eight cases represented by Figure 2(a)–(h) rather than just Figure 2(a) and (b). Fourth, we generalize the local linear estimator $\widehat{\boldsymbol{\beta}}_1$ to a local polynomial estimator $\widehat{\boldsymbol{\beta}}_L$ of arbitrary orders $L \geq 1$. Fifth, we show that this generalized estimator $\widehat{\boldsymbol{\beta}}_L$ with higher order L entails larger extents of bias reduction and hence faster convergence rates.

3. THE METHOD OF ESTIMATION AND INFERENCE

3.1. The Estimator

For an integer $L \geq 1$, our proposed estimator $\hat{\boldsymbol{\beta}}_L$ of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}}_L = E_N[\mathbf{1}\{|D| > h_N\} D^{-1} \mathbf{X}^* (\mathbf{Y} - \mathbf{W}\hat{\boldsymbol{\delta}})] + \hat{\boldsymbol{\gamma}}\hat{\mathbf{h}}, \tag{8}$$

where the three first-step components are

$$\begin{aligned}
\hat{\boldsymbol{\delta}} =& E_N[(\mathbf{D}'_{0:L} E_N[\mathbf{D}_{0:L}\mathbf{D}'_{0:L}]^{-1}\mathbf{e}_1)(\mathbf{X}^*\mathbf{W})'\mathbf{X}^*\mathbf{W}]^{-1} \\
& E_N[(\mathbf{D}'_{0:L} E_N[\mathbf{D}_{0:L}\mathbf{D}'_{0:L}]^{-1}\mathbf{e}_1)(\mathbf{X}^*\mathbf{W})'\mathbf{X}^*\mathbf{Y}], \\
\hat{\boldsymbol{\gamma}} =& E_N\left[\mathbf{X}^*(\mathbf{Y} - \mathbf{W}\hat{\boldsymbol{\delta}})\mathbf{D}'_{1:L}\right] E_N\left[\mathbf{D}_{1:L}\mathbf{D}'_{1:L}\right]^{-1}, \qquad \text{and} \\
\hat{\mathbf{h}} =& E_N[\mathbf{1}\{|D| \leq h_N\}(\ 1 \quad \cdots \quad D^{L-1}\ )']
\end{aligned}$$

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

with the auxiliary notations $\mathbf{e}_1 = (1\ 0\ \cdots\ 0)'$, $\mathbf{D}_{0:L} = \mathbf{1}\{|D| \le h_N\}(1\ D\ \cdots\ D^L)'$, and $\mathbf{D}_{1:L} = \mathbf{1}\{|D| \le h_N\}(D\ \cdots\ D^L)'$.

3.1.1. *$\hat{\boldsymbol{\beta}}_L$ as a Generalization of $\hat{\boldsymbol{\beta}}_1$.* We propose (8) as a generalization of the existing local linear estimator (7) to the local polynomial estimator of degree L. To see this, note that $\hat{\boldsymbol{\beta}}_L$ can be rewritten as

$$\hat{\boldsymbol{\beta}}_L = E_N[\mathbf{1}\{|D| > h_N\}]\hat{\boldsymbol{\beta}}^M + \sum_{l=1}^{L} \frac{\hat{\mathbf{m}}^{(l)}(0)}{l!} E_N[D^{l-1}\mathbf{1}\{|D| \le h_N\}], \tag{9}$$

where $\hat{\boldsymbol{\beta}}^M$ is defined in (4) and $(\hat{\mathbf{m}}^{(1)}(0)/1!\ \cdots\ \hat{\mathbf{m}}^{(L)}(0)/L!) := \hat{\boldsymbol{\gamma}}$. Thus, this estimator $\hat{\boldsymbol{\beta}}_L$ defined in (8) reduces to the existing local linear estimator $\hat{\boldsymbol{\beta}}_1$ defined in (7) in the special case of $L = 1$, in particular justifying the consistency in the notations between (7) and (8).

3.1.2. *Robustness.* Recall that Graham and Powell (2012) show that $\hat{\boldsymbol{\beta}}^M$ works under the case represented by Figure 2(a). They also propose $\hat{\boldsymbol{\beta}}_1$ possibly for the case represented by Figure 2(b), although no formal supporting theory has been developed. Thus, the literature has been silent about whether any estimator would work for the cases represented by Figure 2(b)–(h). In Section 4, we show that the generalized estimator $\hat{\boldsymbol{\beta}}_L$, including $\hat{\boldsymbol{\beta}}_1$ as a special case, in fact works robustly under all of the eight cases represented by Figure 2(a)–(h).

3.1.3. *Bias Reduction.* In addition to the robustness, there is a bias advantage of considering $\hat{\boldsymbol{\beta}}_L$ over $\hat{\boldsymbol{\beta}}^M$. For simplicity, suppose that there is no stayer. In this case, $\hat{\boldsymbol{\beta}}^M$ is consistent, but the information about slow movers is still useful for the purpose of correcting the bias of $\hat{\boldsymbol{\beta}}^M$, and thus for the objective of improving the accuracy of estimation and inference. The local constant estimator $\hat{\boldsymbol{\beta}}^M$ in (4) captures only the $E[\boldsymbol{\beta}(\mathbf{X}) \mid |D| > h_N]$ part in the approximation

$$\begin{aligned}\boldsymbol{\beta} &= E[\boldsymbol{\beta}(\mathbf{X})\mathbf{1}\{|D| > h_N\}] + O(P(|D| \le h_N)) \\ &= P(|D| > h_N)E[\boldsymbol{\beta}(\mathbf{X}) \mid |D| > h_N] + O(P(|D| \le h_N)).\end{aligned}$$

Consequently, $\hat{\boldsymbol{\beta}}^M$ generates a bias of order $O(P(|D| \le h_N))$ as an estimator for $\boldsymbol{\beta}$. In contrast, based on (8) and (9), the generalized estimator $\hat{\boldsymbol{\beta}}_L$ estimates its population counterpart $\mathbf{b}_L$ defined by

$$\begin{aligned}\mathbf{b}_L &= E[\boldsymbol{\beta}(\mathbf{X})\mathbf{1}\{|D| > h_N\}] + \boldsymbol{\gamma}\mathbf{h} = E[\boldsymbol{\beta}(\mathbf{X})\mathbf{1}\{|D| > h_N\}] \\ &\quad + \sum_{l=1}^{L} \frac{\mathbf{m}^{(l)}(0)}{l!} E[D^{l-1}\mathbf{1}\{|D| \le h_N\}],\end{aligned}$$

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

where $\mathbf{h} = E[\mathbf{1}\{|D| \leq h_N\}(1 \ \cdots \ D^{L-1})']$ and $\boldsymbol{\gamma} = (\mathbf{m}^{(1)}(0)/1! \ \cdots \ \mathbf{m}^{(L)}(0)/L!)$. This estimand $\mathbf{b}_L$ approximates $\boldsymbol{\beta}$ better than just the leading term $E[\boldsymbol{\beta}(\mathbf{X})\mathbf{1}\{|D| > h_N\}]$ because the additional terms $\boldsymbol{\gamma}\mathbf{h}$ is based on

$$\boldsymbol{\beta} = E[\boldsymbol{\beta}(\mathbf{X})\mathbf{1}\{|D| > h_N\}] + \sum_{l=1}^{L} \frac{\mathbf{m}^{(l)}(0)}{l!} E[D^{l-1}\mathbf{1}\{|D| \leq h_N\}] + O\left(E[D^L\mathbf{1}\{|D| \leq h_N\}]\right),$$

which we formally show in Theorem 1 in Section 4. Consequently, $\hat{\boldsymbol{\beta}}_L$ generates a smaller bias of order $O\left(E[D^L\mathbf{1}\{|D| \leq h_N\}]\right)$, as opposed to $O\left(P(|D| \leq h_N)\right)$. When $L \geq 1$, this bias is smaller than the bias of $\hat{\boldsymbol{\beta}}^M$.[6] Furthermore, the bias tends to be smaller as the degree L of the polynomial becomes larger.

Thanks to this bias reduction feature of the generalized estimator $\hat{\boldsymbol{\beta}}_L$, our requirement regarding the convergence rate of the trimming threshold h_N is weaker than that for $\hat{\boldsymbol{\beta}}^M$. When we use $\hat{\boldsymbol{\beta}}^M$ as an estimator for $\boldsymbol{\beta}$, we would need to use a small bandwidth h_N satisfying $Nh_N^3 \to 0$ to undersmooth the asymptotic bias as in the empirical application of Graham and Powell (2012). In contrast, we can choose a large bandwidth h_N, e.g., a value that is proportional to $N^{-1/(2L+1)}$. This idea is analogous to the robust bias correction method developed by Calonico, Cattaneo, and Titiunik (2014) in the context of regression discontinuity designs or nonparametric regressions more broadly in general. The weaker requirement on h_N allows us to achieve a faster convergence rate in estimation by taking h_N larger. At the same time, the validity for the statistical inference is insensitive to h_N because our inference is valid for a larger class of h_N than that based on $\hat{\boldsymbol{\beta}}^M$ given in (4). With this said, we want to mention a price that we need to pay to enjoy these bias benefits. Namely, we need to assume more smoothness on the conditional mean function $\mathbf{m}$ as formally stated in Assumption 3 than just one-time differentiability.

3.2. Standard Error

Based on the estimator proposed in Section 3.1, we can conduct robust inference about the parameters. In particular, we consider the APE $E[b_t(A, U_t)] = E[b_1(A, U_1)] + E[d_t(U_{2t})]$, which can be succinctly written as $\boldsymbol{\theta} = \boldsymbol{\beta} + \mathbf{R}\boldsymbol{\delta}$ using a $p \times (p \cdot (T-1))$ matrix $\mathbf{R}$ extracting time effects for a desired period t. For instance, when $p = T = 2$, $\mathbf{R}$ is the 2×2 matrix of zeros for $t = 1$ and $\mathbf{R}$ is the 2×2 identity matrix for $t = 2$.

[6] Graham and Powell (2012, p. 2125) discuss a possibility of bias correction. Indeed, our proposed estimator in the special case of $L = 1$ reduces to their bias-corrected estimator. However, Graham and Powell (2012) leave a formal analysis for future research. This article fills the gap by providing a formal asymptotic analysis for the proposed estimator with $L = 1$ as well as with general $L \geq 1$.

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

The estimator for the APE $\boldsymbol{\theta}$ is $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\beta}}_L + \mathbf{R}\hat{\boldsymbol{\delta}}$. We estimate the influence function for this estimator by

$$\hat{\boldsymbol{\xi}} = \mathbf{1}\{|D| > h_N\} D^{-1}\mathbf{X}^*(\mathbf{Y} - \mathbf{W}\hat{\boldsymbol{\delta}}) - E_N[\mathbf{1}\{|D| > h_N\} D^{-1}\mathbf{X}^*(\mathbf{Y} - \mathbf{W}\hat{\boldsymbol{\delta}})] \tag{10}$$

$$+ (\mathbf{X}^*(\mathbf{Y} - \mathbf{W}\hat{\boldsymbol{\delta}}) - \hat{\boldsymbol{\gamma}}\mathbf{D}_{1:L})\mathbf{D}'_{1:L} E_N\left[\mathbf{D}_{1:L}\mathbf{D}'_{1:L}\right]^{-1}\hat{\mathbf{h}} \tag{11}$$

$$+ \hat{\mathbf{Q}}\hat{\mathbf{V}}^{-1}\left(\mathbf{D}'_{0:L} E_N\left[\mathbf{D}_{0:L}\mathbf{D}'_{0:L}\right]^{-1}\mathbf{e}_1\right)(\mathbf{X}^*\mathbf{W})'\mathbf{X}^*(\mathbf{Y} - \mathbf{W}\hat{\boldsymbol{\delta}}), \tag{12}$$

where

$$\hat{\mathbf{V}} = E_N\left[\left(\mathbf{D}'_{0:L} E_N\left[\mathbf{D}_{0:L}\mathbf{D}'_{0:L}\right]^{-1}\mathbf{e}_1\right)(\mathbf{X}^*\mathbf{W})'\mathbf{X}^*\mathbf{W}\right], \text{ and}$$

$$\hat{\mathbf{Q}} = \mathbf{R} - E_N\left[\left(\mathbf{1}\{|D| > h_N\} D^{-1} + \mathbf{D}'_{1:L} E_N\left[\mathbf{D}_{1:L}\mathbf{D}'_{1:L}\right]^{-1}\hat{\mathbf{h}}\right)\mathbf{X}^*\mathbf{W}\right].$$

While the above formulas may appear sophisticated, we can view them intuitively as follows. Line (10) represents an estimate of the unit influence for the trimmed mean estimation, that is, the unit influence for the $E_N[\mathbf{1}\{|D| > h_N\} D^{-1}\mathbf{X}^*(\mathbf{Y} - \mathbf{W}\hat{\boldsymbol{\delta}})]$ part of the estimator (8). Line (11) represents an estimate of the unit influence for the estimation of the local polynomial coefficients $\boldsymbol{\gamma}$ by $\hat{\boldsymbol{\gamma}}$. Finally, Line (12) represents an estimate of the unit influence for the estimation of the mean time effects $\boldsymbol{\delta}$ by $\hat{\boldsymbol{\delta}}$. In total, Equations (10)–(12) capture the unit influence for all the estimated components of $\hat{\boldsymbol{\theta}}$.

In Section 4, we provide a sufficient condition for the asymptotic normality

$$\sqrt{N} E_N\left[\hat{\boldsymbol{\xi}}\hat{\boldsymbol{\xi}}'\right]^{-1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \to_d \mathcal{N}\left(0, \mathbf{I}_p\right)$$

as $h_N \to 0$ and $N \to \infty$. This result allows us to conduct a robust statistical inference on the APE $\boldsymbol{\theta} = E[b_t(A, U_t)]$.

3.3. Summary of the Procedure

The whole procedure of estimation and inference presented in Sections 3.1 and 3.2 is summarized as an algorithm below.

Step 1. Compute
$\hat{\boldsymbol{\delta}} = E_N[(\mathbf{D}'_{0:L} E_N[\mathbf{D}_{0:L}\mathbf{D}'_{0:L}]^{-1}\mathbf{e}_1)(\mathbf{X}^*\mathbf{W})'\mathbf{X}^*\mathbf{W}]^{-1} E_N[(\mathbf{D}'_{0:L} E_N[\mathbf{D}_{0:L}\mathbf{D}'_{0:L}]^{-1}\mathbf{e}_1)$
$(\mathbf{X}^*\mathbf{W})'\mathbf{X}^*\mathbf{Y}]$.

Step 2. Compute $\hat{\boldsymbol{\gamma}} = E_N\left[\mathbf{X}^*(\mathbf{Y} - \mathbf{W}\hat{\boldsymbol{\delta}})\mathbf{D}'_{1:L}\right] E_N\left[\mathbf{D}_{1:L}\mathbf{D}'_{1:L}\right]^{-1}$.

Step 3. Compute $\hat{\mathbf{h}} = E_N[\mathbf{1}\{|D| \le h_N\}(\ 1 \quad \cdots \quad D^{L-1}\ )']$.

Step 4. Compute $\hat{\boldsymbol{\beta}}_L = E_N[\mathbf{1}\{|D| > h_N\} D^{-1}\mathbf{X}^*(\mathbf{Y} - \mathbf{W}\hat{\boldsymbol{\delta}})] + \hat{\boldsymbol{\gamma}}\hat{\mathbf{h}}$.

Step 5. Compute $\hat{\mathbf{V}} = E_N\left[\left(\mathbf{D}'_{0:L} E_N\left[\mathbf{D}_{0:L}\mathbf{D}'_{0:L}\right]^{-1}\mathbf{e}_1\right)(\mathbf{X}^*\mathbf{W})'\mathbf{X}^*\mathbf{W}\right]$.

Step 6. Compute $\hat{\mathbf{Q}} = \mathbf{R} - E_N\left[\left(\mathbf{1}\{|D| > h_N\} D^{-1} + \mathbf{D}'_{1:L} E_N\left[\mathbf{D}_{1:L}\mathbf{D}'_{1:L}\right]^{-1}\hat{\mathbf{h}}\right)\mathbf{X}^*\mathbf{W}\right]$.

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

Step 7. Compute

$$\hat{\boldsymbol{\zeta}} = \mathbf{1}\{|D| > h_N\} D^{-1} \mathbf{X}^* (\mathbf{Y} - \mathbf{W}\hat{\boldsymbol{\delta}}) - E_N[\mathbf{1}\{|D| > h_N\} D^{-1} \mathbf{X}^* (\mathbf{Y} - \mathbf{W}\hat{\boldsymbol{\delta}})]$$
$$+ (\mathbf{X}^* (\mathbf{Y} - \mathbf{W}\hat{\boldsymbol{\delta}}) - \hat{\boldsymbol{\gamma}}' \mathbf{D}_{1:L}) \mathbf{D}'_{1:L} E_N \left[\mathbf{D}_{1:L} \mathbf{D}'_{1:L}\right]^{-1} \hat{\mathbf{h}}$$
$$+ \hat{\mathbf{Q}} \hat{\mathbf{V}}^{-1} \left(\mathbf{D}'_{0:L} E_N \left[\mathbf{D}_{0:L} \mathbf{D}'_{0:L}\right]^{-1} \mathbf{e}_1\right) (\mathbf{X}^* \mathbf{W})' \mathbf{X}^* (\mathbf{Y} - \mathbf{W}\hat{\boldsymbol{\delta}}).$$

Step 8. Report $\hat{\boldsymbol{\beta}}_L$ and $E_N\left[\hat{\boldsymbol{\zeta}}\hat{\boldsymbol{\zeta}}'\right]^{1/2} / \sqrt{N}$ as an estimate and its standard error, respectively.

4. ASYMPTOTIC PROPERTIES OF THE ESTIMATOR $\hat{\theta}$

In this section, we present asymptotic properties about $\hat{\boldsymbol{\theta}}$ to provide a theoretical guarantee that the method introduced in Section 3 works. We start with stating assumptions on the data-generating process. We use the information near $D = 0$ to estimate $\boldsymbol{\beta}$, and we therefore impose the following assumption, which generalizes the assumptions in Graham and Powell (2012).

Assumption 2. (i) $P(D \neq 0) > 0$. (ii) The conditional distribution of D given $D \neq 0$ has a density ϕ with respect to the Lebesgue measure in a neighborhood of zero. (iii) ϕ is bounded away from zero in a neighborhood of zero.[7] (iv) There are positive numbers α_1, α_2 such that the conditional distribution of $|D|^{-1}$ given $D > 0$ (resp., $D < 0$) has a regularly varying tail with index α_1 (resp., α_2).[8]

In this assumption, we allow for $P(D = 0) > 0$ as well as $\lim_{u \to 0} \phi(u) = \infty$, whereas it requires the absolute continuity of the conditional distribution of D "given $D \neq 0$." It can be empirically relevant that many stayers and many slow movers may exist as discussed in Section 1 with Figures 1 and 2. This feature is different from Graham and Powell (2012), where they treat the case with no mass and the case with a mass separately and they assume that ϕ is smooth and bounded around zero, as in Figure 2(a),(b). Assumption 2 encompasses all of the eight cases illustrated in Figure 2.

To control for a potential unboundedness or non-smoothness of the density ϕ near 0 as in Figure 2(c)–(h), we impose the regularly varying tail condition.[9] The tails are regularly varying with $\alpha_1 = \alpha_2 = 1$ in the special case where ϕ is continuous around zero as in Graham and Powell (2012).[10] It is also the case

[7] The nonparametric identification of $\boldsymbol{\delta}$ based on (3) requires that 0 is in the support of D. Following Assumption 1.2. (i) of Graham and Powell (2012), we assume a stronger condition of $\phi(0) > 0$. By this assumption, we can construct an $O(1/\sqrt{Nh_N})$-consistent estimator for $\boldsymbol{\delta}$.

[8] In other words, $\lim_{t \to \infty} \frac{P(|D|^{-1} \geq tu | D > 0)}{P(|D|^{-1} \geq t | D > 0)} = u^{-\alpha_1}$ and $\lim_{t \to \infty} \frac{P(|D|^{-1} \geq tu | D < 0)}{P(|D|^{-1} \geq t | D < 0)} = u^{-\alpha_2}$ for every $u > 0$.

[9] The regularly varying tail condition holds for a large class of distributions for D and has been used in different contexts. For example, Chaudhuri and Hill (2016) use this condition for the inverse probability weighting estimation of the average treatment effect.

[10] Leading examples are when $D \mid D \neq 0$ follows $N(0,1)$ or t_k with $k \geq 1$.

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

when $\phi(t)$ is proportional to $|t|^{\alpha-1}$ near zero for some $\alpha > 0$. We require that the density of D is bounded away from zero in a neighborhood of zero, so $\max\{\alpha_1, \alpha_2\} \leq 1$.[11]

Since our proposed estimator uses local polynomial regressions, we impose the following smoothness conditions on a few functions.

Assumption 3. (i) The mappings, $u \mapsto E[\mathbf{X}^*(\mathbf{Y} - \mathbf{W}\boldsymbol{\delta}) \mid D = u]$ and $u \mapsto E[(\mathbf{X}^*\mathbf{W})'\mathbf{X}^*(\mathbf{Y} - \mathbf{W}\boldsymbol{\delta}) \mid D = u]$, are $(L+1)$-times differentiable at 0 whose $(L+1)$-th derivative is bounded in a neighborhood of zero. (ii) The mappings, $u \mapsto E\left[(\mathbf{X}^*\mathbf{W})'\mathbf{X}^*\mathbf{W} \mid D = u\right]$ and $u \mapsto Var(\mathbf{X}^*(\mathbf{Y} - \mathbf{W}\boldsymbol{\delta}) \mid D = u)$, are differentiable at 0 whose derivative is bounded in a neighborhood of zero.

This smoothness condition is closely related to Assumption 1.2 (iii) in Graham and Powell (2012). While they impose such a smoothness condition primarily to ensure the ignorability of trimming bias, we further leverage our smoothness condition to correct the trimming bias.

It is important to emphasize that the smoothness index, L, is an inherent part of the model assumption rather than a tuning parameter. Therefore, debating the choice of L in practical applications is not meaningful.[12]

In addition to the smoothness assumption, we assume i.i.d. sampling across individuals, bounded moments, and invertibility for matrices. Graham and Powell (2012) also impose these assumptions.

Assumption 4. (i) $\{(\mathbf{Y}_i, \mathbf{X}_i)\}_{i=1}^N$ are independently and identically distributed. (ii) $E[Z] < \infty$ and $E[Z \mid D = u]$ is bounded in a neighborhood of zero for $Z = \|\mathbf{X}^*\mathbf{W}\|^4$, $\|\mathbf{X}^*(\mathbf{Y} - \mathbf{W}\boldsymbol{\delta})\|^4$, $\|(\mathbf{X}^*\mathbf{W})'\mathbf{X}^*(\mathbf{Y} - \mathbf{W}\boldsymbol{\delta})\|^4$. (iii) $E[(\mathbf{X}^*\mathbf{W})'\mathbf{X}^*\mathbf{W} \mid D = 0]$ and $Var(\mathbf{X}^*(\mathbf{Y} - \mathbf{W}\boldsymbol{\delta}) \mid D = 0)$ are invertible.

Now, we present the two main theoretical results. First, we characterize the bias of $\hat{\boldsymbol{\beta}}_L$ as an estimator for $\boldsymbol{\beta}$. The term $\mathbf{b}_L - \boldsymbol{\beta}$ captures the bias, and we show that the bias can be smaller as L becomes larger.

THEOREM 1. *Under $T = p$ and Assumptions 1–3,* $\mathbf{b}_L = \boldsymbol{\beta} + O\left(E[D^L \mathbf{1}\{|D| \leq h_N\}]\right)$.

Together with this bias reduction result, we can characterize the asymptotic distribution for $\hat{\boldsymbol{\theta}}$ as an estimator for $\boldsymbol{\theta}$.

[11] For example, when $\phi(t) = \alpha|t|^{\alpha-1}/2$ for $t \in [-1, 1]$, the density function ϕ is bounded away from zero near zero if and only if $\alpha \leq 1$. As other notable counterexamples, the density function ϕ is zero at zero if $1/D$ given $D \neq 0$ follows $N(0, 1)$ or t_k with $k \geq 2$.

[12] It may be natural to use simulations to answer the question of what happens if a researcher assumes a level of smoothness that is overly optimistic, i.e., a violation of the smoothness assumption by the true underlying model. However, we could not come up with a data-generating process that violates a certain order of the smoothness assumption due to the convolution relationship between (X_1, X_2) and D. We hypothesize that a violation of the smoothness assumption at a higher order may not result in significant underperformance of the estimator, similar to the lack of power in tests for higher-order smoothness. This question is complex and, therefore, we leave it for future research.

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

THEOREM 2. *Suppose that there is a positive constant c such that*

$$Nh_N^{2+c} \to \infty \text{ and } Nh_N^{2(L+1)} \to 0 \tag{13}$$

as $h_N \to 0$ *and* $N \to \infty$. *Under* $T = p$ *and Assumptions 1–4,*

$$\sqrt{N} E_N \left[\hat{\boldsymbol{\zeta}} \hat{\boldsymbol{\zeta}}' \right]^{-1/2} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \to_d \mathcal{N}\left(0, \mathbf{I}_p\right)$$

as $h_N \to 0$ *and* $N \to \infty$.

The condition (13) on the tuning parameter h_N can be satisfied for example if we take h_N to be proportional to $N^{-1/(2L+1)}$.

Remark 1. As briefly discussed in Section 3.1.3, the bias benefits of $\hat{\boldsymbol{\beta}}_L$ come at the cost of the smoothness conditions imposed in Assumption 3. While these smoothness conditions in a deleted neighborhood of 0 may be plausible in applications, they may not be plausible in a neighborhood of 0 for certain applications as the stayers could be very different in characteristics from the slow movers. Graham and Powell (2012, p. 2130) also mention such a limitation. In case a researcher suspects that the stayers and slow movers may be very different, an alternative approach would be to investigate the conditional APEs given non-stayers, i.e., $D \neq 0$. This alternative parameter can be analyzed by using the same method as the one proposed above after dropping the mass of stayers with $D = 0$.[13]

Remark 2. To fix ideas about Remark 1, consider the model (1) where the regressors $\mathbf{X}_t' = (1, X_{1,t}, X_{2,t})$ consist of a continuous variable $X_{1,t}$ and a binary variable $X_{2,t}$. In this setting, those with time-invariant binary variable (i.e., $X_{2,1} = X_{2,2} = X_{2,3}$) will constitute a subpopulation of stayers. This subpopulation may well be very different from the subpopulation of slow movers even though the values of D are close to each other between these subpopulations. Note that the slow movers have time varying $X_{2,t}$'s.[14] This example illustrates a case in which the smoothness condition in Assumption 3 can be implausible. Researchers with such a model may alternatively consider the conditional APEs given the movers in terms of the binary regressor. This alternative parameter can be analyzed by using the same method as the one proposed above after dropping the mass of stayers in terms of the binary regressor.

Remark 3. Practitioners may be interested in a restricted model in which not all the coefficients are heterogeneous. Specifically, consider the model $Y_t = \mathbf{V}_t' \boldsymbol{\zeta}_0 + \mathbf{X}_t' b_t(A, U_t)$ with $\mathbf{V}_t$ independent of $(A, \mathbf{U})$. In this framework, Graham and Powell (2012, Sect. 3.4) argue that $\boldsymbol{\zeta}_0$ is $\sqrt{N}$ estimable thanks to Robinson (1988), and

[13] This argument can be generalized to the case when these smoothness conditions are imposed except for a subset of the stayers. For example, one may impose Assumption 3 except for the individuals with $\mathbf{X}_1 = \cdots = \mathbf{X}_T$. In such case, we can investigate the conditional APEs for those individuals satisfying Assumption 3, and we can use the one proposed above after dropping the mass of stayers with $\mathbf{X}_1 = \cdots = \mathbf{X}_T$.

[14] Slow movers are those with $((X_{2,1}, X_{2,2}, X_{2,3}) = (0,1,1), (1,0,0)$ & $X_{1,2} \approx X_{1,3})$, $((X_{2,1}, X_{2,2}, X_{2,3}) = (0,0,1), (1,1,0)$ & $X_{1,1} \approx X_{1,2})$, or $((X_{2,1}, X_{2,2}, X_{2,3}) = (1,0,1), (0,1,0)$ & $X_{1,1} \approx X_{1,3})$.

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

propose to use its estimate $\widehat{\boldsymbol{\zeta}}$ and replace $\mathbf{Y}$ by $\mathbf{Y}-\mathbf{V}\widehat{\boldsymbol{\zeta}}$. They conjecture that their baseline asymptotic theory will remain valid with this replacement because the main estimator converges at a lower rate than $\sqrt{N}$. Likewise, we conjecture that our asymptotic theory will also remain valid with this replacement as well, although formal analysis is left for future research.

Remark 4. When $P(D=0)=0$ and ϕ is bounded, the bias term in Theorem 1 is in the order of h_N^{L+1}. In this case, the convergence rate for $\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}$ in Theorem 2 is $1/\sqrt{Nh_N}$ (cf. Lemmas B11 and B12 in the appendix). The optimal choice of h_N based on the asymptotic mean square error is $N^{-1/(2L+3)}$ and the resulting convergence rate is $N^{-(L+1)/(2L+3)}$. The estimator can have a faster convergence rate when we impose Assumption 3 with a larger value of L. When $L \geq 1$, this rate is faster than the convergence rate of $N^{-1/3}$ in Graham and Powell (2012, p. 2125).

5. SIMULATION STUDIES

In this section, we use extensive simulations to demonstrate the robust performance of the unified estimator $\hat{\boldsymbol{\beta}}_L$ proposed in this article. In the first set of simulation exercises, we consider various data-generating processes with no mass of stayers, i.e., $P(D=0)=0$. The objective of this exercise is to compare the performance of the unified estimator $\hat{\boldsymbol{\beta}}_L$ with that of the conventional estimator $\hat{\boldsymbol{\beta}}^M$ when both of them are consistent. In the second set of simulation exercises, we consider various data-generating processes with a nontrivial mass of stayers, i.e., $P(D=0)>0$. The objective of this exercise is to demonstrate the robustness of the unified estimator $\hat{\boldsymbol{\beta}}_L$ even in the presence of stayers.

5.1. Data-Generating Process

Consider the correlated random coefficient panel model (Graham and Powell, 2012):

$$Y_t = \mathbf{X}_t' b_t(A, U_t), \qquad t = 1, 2,$$

where $\mathbf{X}_t = (1, X_t)'$ is a bivariate explanatory variable. The random coefficients are decomposed as $b_t(A, U_t) = b^*(A, U_t) + d_t(U_{2t})$, where the endogenous effects $b^*(A, U_t)$ and the time effects $d_t(U_{2t})$ are given by

$$b^*(A, U_{11t}, U_{12t}) = (A + U_{11t}, A + U_{12t})' \qquad \text{and}$$
$$d_t(U_{21t}, U_{22t}) = (U_{21t} + 0.5 \cdot \mathbf{1}\{t=2\}, U_{22t} + 0.5 \cdot \mathbf{1}\{t=2\})',$$

respectively. The correlated unobserved heterogeneity A is generated by

$$A \mid \varepsilon \sim N(\rho\sigma_A(1+\varepsilon), \sigma_A^2),$$

where ε is a non-negative random variable with $P(\varepsilon \leq t) = \pi_0 + (1-\pi_0)t^\alpha$ for every $t \in [0,1]$. The idiosyncratic parts $(U_{11t}, U_{12t}, U_{21t}, U_{22t})'$ are generated

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

independently as

$$U_{11t}, U_{12t}, U_{21t}, U_{22t} \sim N(0, \sigma_U^2).$$

The explanatory variables (X_1, X_2) are generated according to

$$X_1 \sim N(0,1) \text{ and } X_2 = X_1 + \lambda\varepsilon,$$

where λ is a Rademacher random variable.

The roles played by the main parameters in this setup are as follows. The correlation parameter ρ determines the extent of endogeneity. The mixture parameter $\pi_0 = P(D = 0)$ determines the probability mass of exact stayers. The power parameter α determines the concentration of movers near $D = 0$. Specifically, the smaller α is, the more slow movers will exist. The signal-to-noise ratio σ_A/σ_U determines the relative strength of the correlated unobserved heterogeneity to the idiosyncratic errors. In this setup, the APE $t = 1$ is

$$\boldsymbol{\beta} = E[b^*(A, U_{11t}, U_{12t})] = \left(\pi_0 + (1-\pi_0)\frac{2\alpha+1}{\alpha+1}\right)\rho\sigma_A\,(1,1)',$$

and the average time effects are

$$\delta_{0t} = E[d_1(U_{21t}, U_{22t})] = \begin{cases} (0.00, 0.00)' & \text{if } t = 1 \\ (0.05, 0.05)' & \text{if } t = 2 \end{cases}.$$

These numbers are consistent with the normalization $\delta_{01} = (0.0, 0.0)'$ imposed by Graham and Powell (2012). In our simulation analysis, we compare results across various values of the correlation parameter $\rho \in \{0.5, 1.0\}$, the stayer probability parameter $\pi_0 \in \{0.0, 0.1\}$, and the concentration parameter $1/\alpha \in \{1, 2, 3, 4\}$. We set $\sigma_A = \sigma_U = 0.1$ throughout the simulation exercises.

5.2. Simulation Procedure

For each realization $\{(Y_{i1}, X_{i1}', Y_{i2}, X_{i2}')'\}_{i=1}^n$ generated by the process described above, we first compute the tuning parameter h_N via the plug-in rule $h_N = \frac{1}{2}\min\{\hat{\sigma}_D, \hat{\tau}_D/1.34\}N^{-1/(2L+1)}$, where $\hat{\sigma}_D$ and $\hat{\tau}_D$ denote the sample standard deviation and the sample inter-quartile range of D, respectively. This choice rule coincides with that of Graham and Powell (2012, p. 2138) when $L = 1$. After h_N is selected, the population average effect $\boldsymbol{\beta}$ is estimated by the conventional estimator $\hat{\boldsymbol{\beta}}^M$ and the proposed unified estimator $\hat{\boldsymbol{\beta}}_L$. Finally, the standard errors of $\hat{\boldsymbol{\beta}}^M$ and $\hat{\boldsymbol{\beta}}_L$ are estimated. The estimated standard errors together with the parameter estimates are in turn used to construct estimated 95% confidence intervals for the parameter $\boldsymbol{\beta}$, with the boundaries defined as the 2.5-th and 97.5-th percentiles of the estimated asymptotic normal distribution.

Simulations are run across the alternative sample sizes: $N \in \{500, 1000\}$. Each set of simulations runs 2,500 iterations of data generation, h_N-choice, parameter estimation, and standard error estimation. The average, the bias, the standard

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

deviation, and the root mean squared error of the estimates, $\hat{\boldsymbol{\beta}}_L$ and $\hat{\boldsymbol{\delta}}$, are computed over 2,500 iterations for each simulation set. In addition, the coverage frequencies of the true parameter value $\boldsymbol{\beta}$ by the estimated 90% and 95% confidence intervals are computed over 2,500 iterations for each simulation set.

5.3. Simulation Results

5.3.1. *With No Exact Stayers.* We first focus on the subset of simulation results based on the data-generating processes with no exact stayers, i.e., $\pi_0 = 0$. Recall that the first coordinate $\boldsymbol{\beta}_{00}$ of the two-dimensional parameter vector $\boldsymbol{\beta}$ represents the additive heterogeneity while the second $\boldsymbol{\beta}_{01}$ of it represents the heterogeneous marginal effect of X_t. Appendix C in the Supplementary Material presents simulation results for both of the coordinates. Since the results are very similar between these two coordinates, however, we focus on the second coordinate $\boldsymbol{\beta}_{01}$ in the main text for conciseness. We also focus on the sample size of $N = 1,000$ in the main text while Appendix C presents results across the alternative sample sizes: $N \in \{500, 1000\}$. Table 1 summarizes simulation results with the combinations, $\pi_0 = 0$, $1/\alpha \in \{1, 2, 3, 4\}$, and $\rho \in \{0.5, 1.0\}$, of the data-generating parameters. In this table, we use the degree $L = 2$ of local polynomials.

The contrast between panels (I) and (II) of Table 1 summarizes the aforementioned observations regarding the bias of the conventional estimator $\hat{\boldsymbol{\beta}}^M$ and the unbiasedness of the unified estimator $\hat{\boldsymbol{\beta}}_L$, but now across a larger set of data-generating parameters. Likewise, a comparison between panels (III) and (IV) of Table 1 implies that the root mean square of the unified estimator $\hat{\boldsymbol{\beta}}_L$ is no worse than that of the conventional estimator $\hat{\boldsymbol{\beta}}^M$ in general, and also that the former is strictly better as the sample size becomes larger. Finally, a comparison between panels (V) and (VI) of Table 1 evidences that the undercoverage by the confidence intervals associated with the conventional estimator $\hat{\boldsymbol{\beta}}^M$ tends to become worse as $1/\alpha$ increases and/or ρ increases, while the confidence intervals associated with the unified estimator $\hat{\boldsymbol{\beta}}_L$ robustly achieves the desired coverage probability regardless of the values of the data-generating parameters, $1/\alpha \in \{1, 2, 3, 4\}$ and $\rho \in \{0.5, 1.0\}$. Table 2 shows a counterpart of the right panels of Table 1, but with the degrees $L = 1$ and $L = 3$ of local polynomials instead of $L = 2$. The key patterns of the results displayed in Table 2 remain the same as those displayed in Table 1.

Summarizing the simulation results so far based on the data-generating designs with no exact stayers ($\pi_0 = 0$), we make the following points. First, the conventional estimator $\hat{\boldsymbol{\beta}}^M$ is biased even in the absence of exact stayers, while the unified estimator $\hat{\boldsymbol{\beta}}_L$ is almost unbiased. Second, the bias correction by the unified estimator $\hat{\boldsymbol{\beta}}_L$ is achieved without much sacrificing the variance in comparison with the conventional estimator $\hat{\boldsymbol{\beta}}^M$. As a consequence, the root mean square error

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

TABLE 1. Simulation results for the conventional estimator $\hat{\boldsymbol{\beta}}^M$ and the unified estimator $\hat{\boldsymbol{\beta}}_L$ based on 2500 Monte Carlo iterations

Conventional estimator versus unified estimator with large bandwidths and $L = 2$

(I) Bias: $\hat{\boldsymbol{\beta}}^M$

		$1/\alpha$			
$N = 1{,}000$		1	2	3	4
ρ	0.5	0.002	0.004	0.005	0.005
	1.0	0.003	0.009	0.010	0.010

(II) Bias: $\hat{\boldsymbol{\beta}}_L$

		$1/\alpha$			
$N = 1{,}000$		1	2	3	4
ρ	0.5	0.000	0.000	0.000	0.000
	1.0	0.000	0.000	0.000	0.000

(III) RMSE: $\hat{\boldsymbol{\beta}}^M$

		$1/\alpha$			
$N = 1{,}000$		1	2	3	4
ρ	0.5	0.017	0.007	0.008	0.008
	1.0	0.011	0.011	0.012	0.012

(IV) RMSE: $\hat{\boldsymbol{\beta}}_L$

		$1/\alpha$			
$N = 1{,}000$		1	2	3	4
ρ	0.5	0.014	0.006	0.007	0.008
	1.0	0.010	0.007	0.007	0.008

(V) 95% Cover: $\hat{\boldsymbol{\beta}}^M$

		$1/\alpha$			
$N = 1{,}000$		1	2	3	4
ρ	0.5	0.917	0.804	0.756	0.732
	1.0	0.880	0.589	0.508	0.493

(VI) 95% Cover: $\hat{\boldsymbol{\beta}}_L$

		$1/\alpha$			
$N = 1{,}000$		1	2	3	4
ρ	0.5	0.956	0.948	0.954	0.946
	1.0	0.957	0.953	0.956	0.947

Note: Results are displayed for the second coordinate $\boldsymbol{\beta}_{01}$ of $\boldsymbol{\beta}_0$ and for the sample size $N = 1{,}000$. While exact stayers are absent (i.e., $\pi_0 = 0.0$), the values of $1/\alpha \in \{1, 2, 3, 4\}$ and $\rho \in \{0.5, 1.0\}$ are varied across sets of simulations. The order $L = 2$ of local polynomials is used for the estimation of $\boldsymbol{\delta}$ and $\boldsymbol{\beta}_L$. Displayed statistics are the bias, RMSE, and 95% coverage (i.e., the coverage frequency of the true parameter value by the estimated 95% confidence interval).

(RMSE) of the unified estimator $\hat{\boldsymbol{\beta}}_L$ is no worse than that of the conventional estimator $\hat{\boldsymbol{\beta}}^M$ in general especially as the sample size becomes larger. Third, confidence intervals associated with the conventional estimator $\hat{\boldsymbol{\beta}}^M$ suffer from severe undercoverage even in the absence of exact stayers, while confidence intervals associated with the unified estimator $\hat{\boldsymbol{\beta}}_L$ almost perfectly achieves the coverage probability at the desired nominal probability. We ran a lot more simulations whose results are not presented here for the purpose of conserving space, but the observed patterns summarized above remain to hold under all of the alternative simulation settings.

5.3.2. *Discussions on the Simulation Results.* The simulation results presented above demonstrate a positive feature of $\hat{\boldsymbol{\beta}}_L$. We now provide a discussion behind these positive outcomes. The gains come in two steps.

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

TABLE 2. Simulation results for the unified estimator $\hat{\boldsymbol{\beta}}_L$ based on 2500 Monte Carlo iterations

Unified estimator with large bandwidths and $L = 1$ (left) and $L = 3$ (right)

(I) Bias: $\hat{\boldsymbol{\beta}}_1$

		$1/\alpha$			
$N = 1{,}000$		1	2	3	4
ρ	0.5	0.000	0.000	0.000	0.000
	1.0	0.000	0.000	0.000	0.000

(II) Bias: $\hat{\boldsymbol{\beta}}_3$

		$1/\alpha$			
$N = 1{,}000$		1	2	3	4
ρ	0.5	0.000	0.000	0.000	0.000
	1.0	0.000	0.000	0.000	0.001

(III) RMSE: $\hat{\boldsymbol{\beta}}_1$

		$1/\alpha$			
$N = 1{,}000$		1	2	3	4
ρ	0.5	0.005	0.006	0.007	0.008
	1.0	0.006	0.006	0.007	0.008

(IV) RMSE: $\hat{\boldsymbol{\beta}}_3$

		$1/\alpha$			
$N = 1{,}000$		1	2	3	4
ρ	0.5	0.006	0.007	0.008	0.009
	1.0	0.008	0.007	0.009	0.009

(V) 95% Cover: $\hat{\boldsymbol{\beta}}_1$

		$1/\alpha$			
$N = 1{,}000$		1	2	3	4
ρ	0.5	0.954	0.962	0.947	0.946
	1.0	0.950	0.950	0.947	0.952

(VI) 95% Cover: $\hat{\boldsymbol{\beta}}_3$

		$1/\alpha$			
$N = 1{,}000$		1	2	3	4
ρ	0.5	0.951	0.947	0.947	0.937
	1.0	0.954	0.951	0.945	0.951

Note: Results are displayed for the second coordinate $\boldsymbol{\beta}_{01}$ of $\boldsymbol{\beta}_0$ and for the sample size $N = 1{,}000$. While exact stayers are absent (i.e., $\pi_0 = 0.0$), the values of $1/\alpha \in \{1, 2, 3, 4\}$ and $\rho \in \{0.5, 1.0\}$ are varied across sets of simulations. The orders $L = 1$ (left panels) and $L = 3$ (right panels) of local polynomials are used for the estimation of $\boldsymbol{\delta}$ and $\boldsymbol{\beta}_L$. Displayed statistics are the bias, RMSE, and 95% coverage (i.e., the coverage frequency of the true parameter value by the estimated 95% confidence interval).

First, the gain in terms of the bias comes from the smoothness assumption as mentioned in Sections 3.1.3 and 4. Specifically, we can use a polynomial of degree L to approximate the $\mathbf{m}$ function under the assumption of $(L + 1)$-times differentiability. If the polynomial is flexible enough, then it approximates $\mathbf{m}(u)$ well even for those values of $|u|$ farther away from zero. In other words, extrapolation to farther points becomes feasible. This approximation for larger values of $|u|$ in turn implies that we can choose larger values of the trimming threshold h_N as discussed in Section 3.1.3.

Second, given that a higher degree L allows us to choose large values of the threshold parameter h_N as argued in the first step above, we in turn obtain the gain in terms of variance as well as the bias. Intuitively, this follows because larger h_N implies that we can trim away more slow movers whose $|D|$ values are smaller. If we were not to trim them away, then the slow movers would incur the weak identification problem in that their small $|D|$ values induce slow convergence rates

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

in the estimation of their partial effects. The possibility for us to choose large values of h_N mitigates this weak identification problem and therefore improves the convergence rate of the estimator.

With this said, we also observe that the simulation results for $L = 1$, 2, and 3 are similar to each other. This phenomenon is imputed to the fact that the data-generating process we construct for analytic tractability implies a linear conditional expectation function. In practice, however, we do not know whether the true conditional expectation function is linear, and using a higher order L of approximation yields more robustness in general. The empirical application presented in Section 6 shows a case in which different degrees $L = 1$, 2, and 3 can yield remarkably different results due to possible nonlinearity.

5.3.3. *With Exact Stayers.* We next turn to the set of simulation results based on the data-generating processes with a mass of exact stayers, i.e., $\pi_0 > 0$. In this setting, the conventional estimator $\hat{\boldsymbol{\beta}}^M$ that uses the information of only the movers is biased, and hence $\hat{\boldsymbol{\beta}}^M$ is not a consistent estimator for $\boldsymbol{\beta}$. Our unified estimator $\hat{\boldsymbol{\beta}}_L$, which generalizes the mixture estimator of Graham and Powell (2012, Sect. 3.2), continues to be asymptotically unbiased and consistent. In this light, our discussions hereafter disregard comparisons with $\hat{\boldsymbol{\beta}}^M$, and instead focus on the performance of the unified estimator $\hat{\boldsymbol{\beta}}_L$.

Table 3 summarizes simulation results under data-generating designs with $\pi_0 = 0.1$ (left column) and $\pi_0 = 0.2$ (right column), allowing for 10% and 20%, respectively, of exact stayers. Displayed statistics are the bias (panels (I) and (II)), RMSE (panels (III) and (IV)), and 95% coverage (panels (V) and (VI)) based on the unified estimator $\hat{\boldsymbol{\beta}}_L$ with the degree $L = 2$ of local polynomials. The results are displayed across combinations, $\pi_0 \in \{0.1, 0.2\}$, $\alpha \in \{1, 2, 3, 4\}$, and $\rho \in \{0.5, 1.0\}$, of the data-generating parameters and sample size $N = 1,000$.

We make the following observations in Table 3. First, the estimator remains unbiased even in the non-atomic presence of exact stayers (panels (I) and (II)). The magnitude of the bias is invariant from the sample size. Second, as a consequence of the bias reduction, the confidence intervals still achieve coverage frequencies that are close to the nominal probability of 95% (panels (V) and (VI)). The coverage accuracy improves as the sample size increases. Third, the RMSE becomes larger as the proportion π_0 of exact stayers increases (panels (III) and (IV)). Specifically, for instance, observe that the numbers gradually increase as we move from Table 1(IV) in which $\pi_0 = 0.0$ to Table 3(III) in which $\pi_0 = 0.1$, and then to Table 3(IV) in which $\pi_0 = 0.2$. This result concerning the RMSE is natural because the mass of stayers fails to provide data variations and a larger proportion π_0 means more mass of such stayers exhibiting no data variations. Finally, we remark that we ran more simulations whose results are not presented here for the purpose of conserving space, but the observed patterns summarized above remain to hold under all of the alternative simulation settings.

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

TABLE 3. Simulation results for the unified estimator $\hat{\boldsymbol{\beta}}_L$ in the presence of exact stayers based on 2500 Monte Carlo iterations

Robust performance of the unified estimator in the presence of exact stayers

(I) Bias: $\pi_0 = 0.1$

		$1/\alpha$			
$N = 1,000$		1	2	3	4
ρ	0.5	0.000	0.000	0.000	0.000
	1.0	0.001	0.001	0.001	0.000

(II) Bias: $\pi_0 = 0.2$

		$1/\alpha$			
$N = 1,000$		1	2	3	4
ρ	0.5	0.000	0.000	0.000	0.000
	1.0	0.001	0.001	0.000	0.000

(III) RMSE: $\pi_0 = 0.1$

		$1/\alpha$			
$N = 1,000$		1	2	3	4
ρ	0.5	0.007	0.007	0.009	0.010
	1.0	0.007	0.008	0.009	0.010

(IV) RMSE: $\pi_0 = 0.2$

		$1/\alpha$			
$N = 1,000$		1	2	3	4
ρ	0.5	0.009	0.010	0.011	0.012
	1.0	0.010	0.010	0.011	0.011

(V) 95% Cover: $\pi_0 = 0.1$

		$1/\alpha$			
$N = 1,000$		1	2	3	4
ρ	0.5	0.945	0.950	0.946	0.944
	1.0	0.948	0.952	0.948	0.940

(IV) 95% Cover: $\pi_0 = 0.2$

		$1/\alpha$			
$N = 1,000$		1	2	3	4
ρ	0.5	0.931	0.936	0.944	0.942
	1.0	0.937	0.942	0.943	0.941

Note: Results are displayed for the second coordinate $\boldsymbol{\beta}_{01}$ of $\boldsymbol{\beta}_0$ and for the sample size $N = 1,000$. The values of $\pi \in \{0.1, 0.2\}$, $1/\alpha \in \{1, 2, 3, 4\}$, and $\rho \in \{0.5, 1.0\}$ are varied across sets of simulations. The order $L = 2$ of local polynomials are used for the estimation of $\boldsymbol{\delta}$ and $\boldsymbol{\beta}_L$. Displayed statistics are the bias, RMSE, and 95% coverage (i.e., the coverage frequency of the true parameter value by the estimated 95% confidence interval).

6. HETEROGENEOUS MARGINAL PROPENSITIES TO CONSUME

In this section, we apply our proposed econometric method and investigate the average of the heterogeneous marginal propensities to consume (MPC) using the U.S. PSID. The PSID has started to collect information via surveys about income since 1968, and it also introduced a comprehensive measure of consumption in 1999, hence allowing for the estimation of the MPC. While the literature on empirical analysis of the MPC is vast, we in particular feature the following list of recent studies that use the PSID: Cooper, Dynan, and Rhodenhiser (2019), Aguiar, Bils, and Boar (2020), and Fisher et al. (2020). These papers find heterogeneous MPC that vary with observables, such as wealth. On the other hand, none of these and other existing papers considers unobserved heterogeneity in the MPC to our knowledge. Treating correlated random coefficients as heterogeneous MPC, we estimate the average of them.

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

Table 4. Estimates of the marginal propensities to consume in year 2017 with PSID 2017–2019

				$\hat{\beta}_L$				
	OLS	FE	$\hat{\beta}^M$	$L=1$	$L=2$	$L=3$	$L=4$	$L=5$
Total	0.468	0.104	0.172	0.356	0.194	0.162	0.203	0.221
expenditures	(0.008)	(0.010)	(0.074)	(0.274)	(0.095)	(0.096)	(0.080)	(0.072)
Food	0.364	0.052	0.207	0.513	0.435	0.417	0.276	0.213
expenditures	(0.009)	(0.014)	(0.120)	(0.412)	(0.166)	(0.168)	(0.139)	(0.125)
Housing	0.379	0.082	0.153	0.401	0.213	0.220	0.201	0.216
expenditures	(0.008)	(0.009)	(0.087)	(0.317)	(0.109)	(0.111)	(0.091)	(0.081)

Note: OLS stands for the ordinary least squares estimator. FE stands for the fixed-effect estimator. $\hat{\beta}_L$ indicates the biased-corrected average correlated random coefficient for each $L \in \{1, \ldots, 5\}$. The numbers in parentheses indicate standard errors.

The PSID has been based on biennial surveys since 1997. For our analysis, we use the most recent pair of available years, 2017 and 2019, as of the time of writing this section. We set survey years, 2017 and 2019, as time periods, $t = 1$ and $t = 2$, respectively. For each period t, the right-hand-side variable X_t is defined by the logarithm of the total family income. Recall from Figure 1 that this setting generates many slow movers, and hence the conventional estimator $\hat{\beta}^M$ without bias correction may not perform well as shown through our simulation studies. The left-hand-side variable Y_t is defined by the logarithm of the total family expenditures, including food, housing, telephone/internet, transportation, vehicle repair, education, child care, health care, household repairs, household furnishing, clothing, and other recreation.

As standard benchmarks, we first estimate the MPC based on constant coefficient models using the ordinary least squares (OLS) and the fixed-effect (FE) estimators. The first two columns of Table 4 show estimates based on these two benchmark estimators. Focusing on the first row of the table in which Y_t is the logarithm of the total family expenditures defined above, observe that the OLS yields 47% of the MPC and the FE yields 10% of the MPC. This ordinal relation between these two values is expected, as the OLS more likely fails to account for the endogeneity than the FE. While values differ, the same qualitative patterns still result even if we define the left-hand-side variable Y_t differently, as shown in the second row of the table where Y_t is defined by the logarithm of the family expenditures on food and the third row of the tale where Y_t is defined by the logarithm of the family expenditures on housing.

Even though the FE indeed accounts for additive endogeneity, it still fails to account for non-additive endogeneity. To address potentially non-additive endo-

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

geneity, we next suppose that the underlying model is described by the correlated random coefficient model (1), where the random coefficient $b_t(A, U_t)$ is interpreted as the heterogeneous MPC. The third column in Table 4 shows estimates of the average $\boldsymbol{\beta} = E[b_1(A, U_1)]$ of the heterogeneous MPC based on the local constant estimator, $\hat{\boldsymbol{\beta}}^M$, of Graham and Powell (2012). As shown through our simulation studies, however, this estimator $\hat{\boldsymbol{\beta}}^M$ without bias correction generally entails large bias, and inference based on its estimates and standard errors tend to be misleading under the presence of many slow movers as is the case with the current data. The last five columns of Table 4 show estimates of the average $\boldsymbol{\eta} = E[b_1(A, U_1)]$ of the heterogeneous MPC by the bias-corrected estimator $\hat{\boldsymbol{\beta}}_L$ with the five orders $L \in \{1, \ldots, 5\}$ of bias correction. For the tuning parameter h_N, we propose to use the plug-in rule given in Section 5.2.

Observe that the bias-corrected estimator $\hat{\boldsymbol{\beta}}_L$ is imprecise when the order of bias correction is set to $L = 1$, as indicated by the large standard errors. As the order L of bias correction increases, however, the bias-corrected estimator $\hat{\boldsymbol{\beta}}_L$ allows for smaller bandwidths and hence gains higher precision as indicated by the decreasing standard errors. In the first row of Table 4 in which the left-hand-side variable Y_t is defined as the logarithm of the total family expenditures, for instance, the case of $L = 1$ fails to conclude a significant average MPC, whereas the cases with $L \in \{2, 4, 5\}$ allow to conclude a significant average MPC. This observation illustrates that a higher-order bias correction is preferred, provided that a researcher is willing to assume sufficient smoothness.

Across all the three definitions of Y_t shown in Table 4, the estimates with the highest degree $L = 5$ of bias correction precisely indicate 21%–22% of the average MPC, which is between the OLS estimates and the FE estimates. While the OLS overestimates the MPC due to its failure to account for endogeneity, the FE on the other hand seems to underestimate the MPC due to its failure to account for non-additive endogeneity. The latter observation implies a likely presence of non-additive heterogeneity.

To investigate the heterogeneity of $b_1(A, U_1)$ in more detail, we plot the estimated conditional mean function $u \mapsto E[b_1(A, U_1)|D = u]$ in Figure 3. In this figure, we can observe the positive partial correlation between the MPC $b_1(A, U_1)$ and the difference $D = X_2 - X_1$ in the logarithm of total family incomes between 2017 and 2019. This observation leads to two implications. First, it evidences the heterogeneity in the MPC, and it is therefore crucial to employ the estimators like $\hat{\boldsymbol{\beta}}_L$ which presume correlated random coefficients, rather than the OLS or the FE which presume homogeneous effects. Second, the positive relation implies a systematic asymmetry in heterogeneity in the sense that subpopulations with negative income growths (i.e., $D < 0$) are associated with smaller MPC and those with positive income growths (i.e., $D > 0$) are associated with larger MPC.

Finally, observe that the estimated conditional mean function $u \mapsto E[b_1(A, U_1)|D = u]$ displayed in Figure 3 appears to be curved rather than linear. This observation implies that a high order of approximation, such as $L \geq 2$,

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

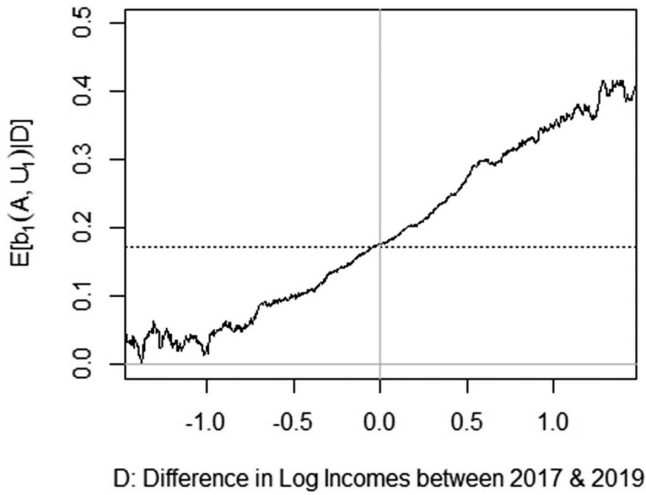


FIGURE 3. Estimated conditional mean function $u \mapsto E[b_1(A, U_1)|D = u]$ of heterogeneous MPC $b_1(A, U_1)$ given the difference $D = X_2 - X_1$ in the logarithm of total family incomes between 2017 and 2019.

is desired for bias correction. In other words, our local polynomial estimators ($\hat{\boldsymbol{\beta}}_L$ for $L \in \{2, \ldots, 5\}$) are preferred to the conventional local constant estimator $\hat{\boldsymbol{\beta}}^M$ or local linear estimator $\hat{\boldsymbol{\beta}}_1$.

7. ESTIMATION AND INFERENCE WITH $T > P$

Thus far, we have focused on the case with $T = p$. In this section, we modify the methodology in the previous sections to the case of $T > p$. We re-define D, $\mathbf{D}_{1:L}$, $\mathbf{m}$, $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\beta}}_L$, $\hat{\boldsymbol{\delta}}$, $\hat{\boldsymbol{\gamma}}$, $\hat{\mathbf{h}}$, $\hat{\boldsymbol{\zeta}}$ differently from the ones in the previous sections. In the current section, we only introduce an estimator and its estimated influence function. We present formal theoretical results to support this procedure in Appendix D.

The correlated random coefficient panel data model is the same as in the previous sections: $Y_t = \mathbf{X}_t' b_t(A, U_t)$ for $t = 1, \ldots, T$. The short-hand notations, such as $\mathbf{Y}$, $\mathbf{X}$, $\mathbf{W}$, $\boldsymbol{\beta}(\mathbf{X})$, $\boldsymbol{\beta}$, and $\boldsymbol{\delta}$, also remain the same as in the previous sections. Using a matrix $\mathbf{R}$ that extracts time effects as before, suppose that the parameter of interest is $\boldsymbol{\theta} = \boldsymbol{\beta} + \mathbf{R}\boldsymbol{\delta}$. The parameter $(\boldsymbol{\beta}', \boldsymbol{\delta}')'$ can be identified based on the following moment implications from Assumption 1:

$$E[(\mathbf{X}'\mathbf{X})^*\mathbf{X}'(\mathbf{Y} - \mathbf{W}\boldsymbol{\delta}) \mid \mathbf{X}] = D\boldsymbol{\beta}(\mathbf{X}) \text{ with } \boldsymbol{\beta} = E[\boldsymbol{\beta}(\mathbf{X})] \qquad \text{and} \qquad \textbf{(14)}$$

$$E[\mathbf{1}\{D > h_N\}\mathbf{W}\mathbf{M}_\mathbf{X}\mathbf{Y}] = E[\mathbf{1}\{D > h_N\}\mathbf{W}'\mathbf{M}_\mathbf{X}\mathbf{W}]\boldsymbol{\delta}, \qquad \textbf{(15)}$$

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

where $D = \det(\mathbf{X}'\mathbf{X})$, $\mathbf{M}_\mathbf{X} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and $(\mathbf{X}'\mathbf{X})^*$ denotes the adjoint of $\mathbf{X}'\mathbf{X}$. Equations (14) and (15) motivate estimators for $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ in $\boldsymbol{\theta} = \boldsymbol{\beta} + \mathbf{R}\boldsymbol{\delta}$.

Specifically, we can estimate the parameter $\boldsymbol{\theta} = \boldsymbol{\beta} + \mathbf{R}\boldsymbol{\delta}$ by $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\beta}}_L + \mathbf{R}\hat{\boldsymbol{\delta}}$, where

$$\hat{\boldsymbol{\beta}}_L = E_N[\mathbf{1}\{D > h_N\}D^{-1}(\mathbf{X}'\mathbf{X})^*\mathbf{X}'(\mathbf{Y} - \mathbf{W}\hat{\boldsymbol{\delta}})] + \hat{\boldsymbol{\gamma}}\hat{\mathbf{h}} \tag{16}$$

$$\hat{\boldsymbol{\delta}} = E_N[\mathbf{1}\{D > h_N\}\mathbf{W}'\mathbf{M}_\mathbf{X}\mathbf{W}]^{-1}E_N[\mathbf{1}\{D > h_N\}\mathbf{W}\mathbf{M}_\mathbf{X}\mathbf{Y}] \tag{17}$$

$$\hat{\boldsymbol{\gamma}} = E_N\left[(\mathbf{X}'\mathbf{X})^*\mathbf{X}'(\mathbf{Y} - \mathbf{W}\hat{\boldsymbol{\delta}})\mathbf{D}'_{1:L}\right]E_N\left[\mathbf{D}_{1:L}\mathbf{D}'_{1:L}\right]^{-1}$$

$$\hat{\mathbf{h}} = E_N[\mathbf{1}\{D \le h_N\}(\ 1 \quad \cdots \quad D^{L-1}\ )']$$

with the auxiliary notation $\mathbf{D}_{1:L} = \mathbf{1}\{D \le h_N\}(\ D \quad \cdots \quad D^L\ )'$. These formulas for $\hat{\boldsymbol{\beta}}_L$, $\hat{\boldsymbol{\delta}}$, $\hat{\boldsymbol{\gamma}}$, and $\hat{\mathbf{h}}$ parallel those presented in Section 3.1 for the just identified case with $T = p$, and hence they follow by analogous derivations to the baseline case. In particular, as a regularized sample counterpart of (14) with bias correction, the right-hand side of (16) consists of the trimmed empirical mean in the first term and the L-th-order local polynomial approximation of the trimming error in the second term. On the other hand, a direct sample counterpart of (15) yields (17).

The influence function for $\hat{\boldsymbol{\theta}}$ is estimated by

$$\hat{\boldsymbol{\zeta}} = \mathbf{1}\{D > h_N\}D^{-1}(\mathbf{X}'\mathbf{X})^*\mathbf{X}'(\mathbf{Y} - \mathbf{W}\hat{\boldsymbol{\delta}}) - E_N[\mathbf{1}\{D > h_N\}D^{-1}(\mathbf{X}'\mathbf{X})^*\mathbf{X}'(\mathbf{Y} - \mathbf{W}\hat{\boldsymbol{\delta}})] \tag{18}$$

$$+ ((\mathbf{X}'\mathbf{X})^*\mathbf{X}'(\mathbf{Y} - \mathbf{W}\hat{\boldsymbol{\delta}}) - \hat{\boldsymbol{\gamma}}\mathbf{D}_{1:L})\mathbf{D}'_{1:L}E_N\left[\mathbf{D}_{1:L}\mathbf{D}'_{1:L}\right]^{-1}\hat{\mathbf{h}} \tag{19}$$

$$+ \hat{\mathbf{Q}}\hat{\mathbf{V}}^{-1}\mathbf{1}\{D > h_N\}\mathbf{W}\mathbf{M}_\mathbf{X}(\mathbf{Y} - \mathbf{W}\hat{\boldsymbol{\delta}}), \tag{20}$$

where

$$\hat{\mathbf{V}} = E_N[\mathbf{1}\{D > h_N\}\mathbf{W}'\mathbf{M}_\mathbf{X}\mathbf{W}], \text{ and}$$

$$\hat{\mathbf{Q}} = \mathbf{R} - E_N\left[\left(\mathbf{1}\{D > h_N\}D^{-1} + \mathbf{D}'_{1:L}E_N\left[\mathbf{D}_{1:L}\mathbf{D}'_{1:L}\right]^{-1}\hat{\mathbf{h}}\right)(\mathbf{X}'\mathbf{X})^*\mathbf{X}'\mathbf{W}\right].$$

Similarly to the baseline case presented in Section 3.2, we can view the components of the influence function estimator intuitively as follows. Line (18) represents an estimate of the unit influence for the trimmed mean estimation. Line (19) represents an estimate of the unit influence for the estimation of the local polynomial coefficients $\boldsymbol{\gamma}$ by $\hat{\boldsymbol{\gamma}}$. Finally, Line (20) represents an estimate of the unit influence for the estimation of the mean time effects $\boldsymbol{\delta}$ by $\hat{\boldsymbol{\delta}}$. In total, Equations (18)–(20) capture the unit influence for all the estimated components of $\hat{\boldsymbol{\theta}}$.

With these notations, we provide a sufficient condition for the asymptotic normality

$$\sqrt{N}E_N\left[\hat{\boldsymbol{\zeta}}\hat{\boldsymbol{\zeta}}'\right]^{-1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \to_d \mathcal{N}\left(0, \mathbf{I}_p\right).$$

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

Theorem S.2 in Appendix D in the Supplementary Material provides a formal statement of this result. Consequently, a researcher may report the point estimate, $\hat{\boldsymbol{\theta}}$, for the APE $\boldsymbol{\theta}$ and its standard error, $\left[\hat{\boldsymbol{\xi}}\hat{\boldsymbol{\xi}}'\right]^{1/2}/\sqrt{N}$ under over-identified cases with $T > p$.

8. CONCLUSION

Panel data often contain units with no within-variations (stayers) and units with little within-variations (slow movers). Figure 1 demonstrates a case in point in the log of total family income in the U.S. PSID between 2017 and 2019, where there are a mass of stayers and many slow movers. The presence of stayers and many slow movers as in this figure can impose challenges in the analysis of random coefficient models. Extending some ideas from the existing literature, we propose a method of inference for the APEs in correlated random coefficient models robustly across various distributions of within-variations, including cases with a mass of stayers and many slow movers. The extent of this robustness encompasses all of the eight cases illustrated in Figure 2. In addition to the robustness property, our proposed method enjoys smaller biases and hence improved accuracy in inference. Simulation studies demonstrate theoretical claims about these preferred properties. In our simulation studies, our proposed 95% confidence interval achieves 93%–96% coverage frequencies, as opposed to 37%–93% coverage frequencies for the conventional method.

APPENDICES

A. PROOFS

This section collects the proofs of the main results together with auxiliary lemmas and their proofs.

A.1. Proof of Theorem 1

Proof. Note that, by Assumption 3, there is a constant $C < \infty$ such that $\|\mathbf{m}(u) - \sum_{l=0}^{L} \mathbf{m}^{(l)}(0)\frac{u^l}{l!}\| \leq C|u|^{L+1}$ for every u in a neighborhood of zero. Using $\mathbf{m}(u) = uE[\boldsymbol{\beta}(\mathbf{X}) \mid D = u]$, we can decompose $\mathbf{b}_L$ and $\boldsymbol{\beta}$ as

$$\mathbf{b}_L = E[\mathbf{X}^{-1}(\mathbf{Y} - \mathbf{W}\boldsymbol{\delta})\mathbf{1}\{|D| > h_N\}] + E[\boldsymbol{\beta}(\mathbf{X})\mathbf{1}\{D = 0\}] + E\left[\sum_{l=1}^{L} \mathbf{m}^{(l)}(0)\frac{D^{l-1}}{l!}\mathbf{1}\{0 < |D| \leq h_N\}\right],$$

and

$$\boldsymbol{\beta} = E[\mathbf{X}^{-1}(\mathbf{Y} - \mathbf{W}\boldsymbol{\delta})\mathbf{1}\{|D| > h_N\}] + E[\boldsymbol{\beta}(\mathbf{X})\mathbf{1}\{D = 0\}] + E\left[\frac{\mathbf{m}(D)}{D}\mathbf{1}\{0 < |D| \leq h_N\}\right].$$

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

Taking the difference and noting that ϕ is the conditional density function of D given $D \neq 0$ by Assumption 2, we have

$$\|\mathbf{b}_L - \boldsymbol{\beta}\| \leq \int_{-h_N}^{h_N} \left\| \frac{\mathbf{m}(u) - \mathbf{m}(0) - \sum_{l=1}^{L} \mathbf{m}^{(l)}(0) \frac{u^l}{l!}}{u} \right\| \phi(u) du \leq C \int_{-h_N}^{h_N} |u|^L \phi(u) du,$$

where the first inequality uses $\mathbf{m}(0) = 0$. □

A.2. Proof of Theorem 2

Before proceeding with the proof of Theorem 2, we first introduce short-hand notations to simplify the proof. When we derive the influence function representation, we control the remainder term as $o_p\left(\sqrt{\frac{\pi_h^2}{Nh_N^2 \pi_{h\backslash 0}}}\right)$ with $\pi_h = P(|D| \leq h_N)$ and $\pi_{h\backslash 0} = P(0 < |D| \leq h_N)$. For $\left(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\gamma}}\right)$, the design matrices, $E_N\left[\mathbf{D}_{0:L}\mathbf{D}_{0:L}'\right]$ and $E_N\left[\mathbf{D}_{1:L}\mathbf{D}_{1:L}'\right]$, are used. As in the standard local polynomial regression, the inverses of these design matrices are not bounded, so we normalize them to ensure that the inverses are bounded. The resulting matrices are

$$\hat{\boldsymbol{\Omega}} = \frac{1}{\pi_h} \boldsymbol{\Psi}_{0:L}^{-1} E_N\left[\mathbf{D}_{0:L}\mathbf{D}_{0:L}'\right] \boldsymbol{\Psi}_{0:L}^{-1} \qquad \text{and} \qquad \hat{\boldsymbol{\Pi}} = \frac{1}{\pi_{h\backslash 0}} \boldsymbol{\Psi}_{1:L}^{-1} E_N\left[\mathbf{D}_{1:L}\mathbf{D}_{1:L}'\right] \boldsymbol{\Psi}_{1:L}^{-1},$$

where $\boldsymbol{\Psi}_{0:L} = \text{diag}\begin{pmatrix} 1 & h_N & \cdots & h_N^L \end{pmatrix}$ and $\boldsymbol{\Psi}_{1:L} = \text{diag}\begin{pmatrix} h_N & \cdots & h_N^L \end{pmatrix}$. Their population counterparts are

$$\boldsymbol{\Omega} = \frac{1}{\pi_h} \boldsymbol{\Psi}_{0:L}^{-1} E\left[\mathbf{D}_{0:L}\mathbf{D}_{0:L}'\right] \boldsymbol{\Psi}_{0:L}^{-1} \qquad \text{and} \qquad \boldsymbol{\Pi} = \frac{1}{\pi_{h\backslash 0}} \boldsymbol{\Psi}_{1:L}^{-1} E\left[\mathbf{D}_{1:L}\mathbf{D}_{1:L}'\right] \boldsymbol{\Psi}_{1:L}^{-1}.$$

In the proof, we characterize the influence function for $\hat{\boldsymbol{\beta}}_L + \mathbf{R}\hat{\boldsymbol{\delta}}$ by

$$\boldsymbol{\zeta} = \boldsymbol{\xi}_1 + \frac{1}{\pi_{h\backslash 0}} \boldsymbol{\xi}_2 \boldsymbol{\Pi}^{-1} \boldsymbol{\Psi}_{1:L}^{-1} \mathbf{h} + \frac{1}{\pi_h} \mathbf{Q}\mathbf{V}^{-1} \boldsymbol{\xi}_3 \boldsymbol{\Omega}^{-1} \mathbf{e}_1,$$

where

$$\mathbf{V} = E\left[\left(\mathbf{D}_{0:L}' E\left[\mathbf{D}_{0:L}\mathbf{D}_{0:L}'\right]^{-1} \mathbf{e}_1\right) (\mathbf{X}^*\mathbf{W})' \mathbf{X}^*\mathbf{W}\right],$$

$$\mathbf{Q} = \mathbf{R} - E\left[\left(\mathbf{1}\{|D| > h_N\} D^{-1} + \mathbf{D}_{1:L}' E\left[\mathbf{D}_{1:L}\mathbf{D}_{1:L}'\right]^{-1} \mathbf{h}\right) \mathbf{X}^*\mathbf{W}\right],$$

$$\boldsymbol{\xi}_1 = \mathbf{1}\{|D| > h_N\} D^{-1} \mathbf{X}^* (\mathbf{Y} - \mathbf{W}\boldsymbol{\delta}) - E[\mathbf{1}\{|D| > h_N\} D^{-1} \mathbf{X}^* (\mathbf{Y} - \mathbf{W}\boldsymbol{\delta})],$$

$$\boldsymbol{\xi}_2 = (\mathbf{X}^* (\mathbf{Y} - \mathbf{W}\boldsymbol{\delta}) - \boldsymbol{\gamma}\mathbf{D}_{1:L}) \mathbf{D}_{1:L}' \boldsymbol{\Psi}_{1:L}^{-1}, \qquad \text{and}$$

$$\boldsymbol{\xi}_3 = (\mathbf{X}^*\mathbf{W})' \mathbf{X}^* (\mathbf{Y} - \mathbf{W}\boldsymbol{\delta}) \mathbf{D}_{0:L}' \boldsymbol{\Psi}_{0:L}^{-1}.$$

We also use the sample counterparts of $\boldsymbol{\xi}_1$, $\boldsymbol{\xi}_2$, and $\boldsymbol{\xi}_3$:

$$\hat{\boldsymbol{\xi}}_1 = \mathbf{1}\{|D| > h_N\} D^{-1} \mathbf{X}^* (\mathbf{Y} - \mathbf{W}\hat{\boldsymbol{\delta}}) - E_N[\mathbf{1}\{|D| > h_N\} D^{-1} \mathbf{X}^* (\mathbf{Y} - \mathbf{W}\hat{\boldsymbol{\delta}})],$$

$$\hat{\boldsymbol{\xi}}_2 = (\mathbf{X}^* (\mathbf{Y} - \mathbf{W}\hat{\boldsymbol{\delta}}) - \hat{\boldsymbol{\gamma}}\mathbf{D}_{1:L}) \mathbf{D}_{1:L}' \boldsymbol{\Psi}_{1:L}^{-1}, \qquad \text{and}$$

$$\hat{\boldsymbol{\xi}}_3 = (\mathbf{X}^*\mathbf{W})' \mathbf{X}^* (\mathbf{Y} - \mathbf{W}\hat{\boldsymbol{\delta}}) \mathbf{D}_{0:L}' \boldsymbol{\Psi}_{0:L}^{-1}.$$

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

In the proof, we use the operator norm (based on the Euclidean norms) as a norm for matrices. All the asymptotic results in the proof hold as $h_N \to 0$ and $N \to \infty$.

It is worthwhile discussing $(\pi_{h\backslash 0}, \pi_h)$. For any sample size, we have $\pi_{h\backslash 0} \leq \pi_h$. Also, since ϕ is bounded away from zero near zero, the term $\pi_{h\backslash 0}/h_N$ is also bounded away from zero.

With these notations introduced, we are now going to show Theorem 2. In the proof, we use the auxiliary lemmas collected in Appendix B.

Proof of Theorem 2. First, we are going to show

$$E_N\left[\hat{\boldsymbol{\xi}}\hat{\boldsymbol{\xi}}'\right]^{-1} Var(\boldsymbol{\zeta}) = \mathbf{I}_p + o_p(1). \tag{A.1}$$

By (13) and Lemmas B11 and B13, we have $E_N[\|\hat{\boldsymbol{\xi}} - \boldsymbol{\zeta}\|^2]^{1/2} = o_p\left(\sqrt{\frac{\pi_h^{3+\epsilon}}{h_N^2 \pi_{h\backslash 0}}}\right)$, $E_N[\|\boldsymbol{\zeta}\|^2]^{1/2} = O_p\left(\sqrt{\frac{\pi_h^{1-\epsilon}}{h_N^2}}\right)$, $(E_N - E)\left[\boldsymbol{\zeta}\boldsymbol{\zeta}'\right] = o_p\left(\frac{\pi_h^2}{h_N^2 \pi_{h\backslash 0}}\right)$, and $E[\boldsymbol{\zeta}] = o(1)$, where we can take a sufficiently small $\epsilon > 0$. Therefore,

$$\begin{aligned}\left\|E_N\left[\hat{\boldsymbol{\xi}}\hat{\boldsymbol{\xi}}'\right] - Var(\boldsymbol{\zeta})\right\| &\leq E_N[\|\hat{\boldsymbol{\xi}} - \boldsymbol{\zeta}\|^2] + 2E_N[\|\hat{\boldsymbol{\xi}} - \boldsymbol{\zeta}\|^2]^{1/2} E_N[\|\boldsymbol{\zeta}\|^2]^{1/2} \\ &\quad + \|(E_N - E)\left[\boldsymbol{\zeta}\boldsymbol{\zeta}'\right]\| + \|E[\boldsymbol{\zeta}]\|^2 \\ &= o_p\left(\frac{\pi_h^2}{h_N^2 \pi_{h\backslash 0}}\right).\end{aligned}$$

By Lemma B12, we have $Var(\boldsymbol{\zeta})^{-1} = O(\frac{h_N^2 \pi_{h\backslash 0}}{\pi_h^2})$, and so

$$\begin{aligned}E_N\left[\hat{\boldsymbol{\xi}}\hat{\boldsymbol{\xi}}'\right]^{-1} Var(\boldsymbol{\zeta}) - \mathbf{I}_p &= -\left(Var(\boldsymbol{\zeta}) + \left(E_N\left[\hat{\boldsymbol{\xi}}\hat{\boldsymbol{\xi}}'\right] - Var(\boldsymbol{\zeta})\right)\right)^{-1}\left(E_N\left[\hat{\boldsymbol{\xi}}\hat{\boldsymbol{\xi}}'\right] - Var(\boldsymbol{\zeta})\right) \\ &= o_p(1).\end{aligned}$$

Second, we are going to show

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} - (E_N - E)[\boldsymbol{\zeta}] = o_p\left(\sqrt{\frac{\pi_h^2}{N h_N^2 \pi_{h\backslash 0}}}\right). \tag{A.2}$$

By Lemmas B1, B2, B4–B7, B9, and B10 and Theorem 1,

$$\begin{aligned}&\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} - (E_N - E)[\boldsymbol{\zeta}]\| \\ &\leq \|\mathbf{b}_L - \boldsymbol{\beta} + E[\boldsymbol{\zeta}]\| \\ &\quad + \frac{1}{\pi_{h\backslash 0}}\left(\left\|\hat{\boldsymbol{\Pi}}^{-1} - \boldsymbol{\Pi}^{-1}\right\| \|\boldsymbol{\Psi}_{1:L}^{-1}\hat{\mathbf{h}}\| + \|\boldsymbol{\Pi}^{-1}\| \|\boldsymbol{\Psi}_{1:L}^{-1}\left(\hat{\mathbf{h}} - \mathbf{h}\right)\|\right) \|E_N\left[\boldsymbol{\xi}_2\right]\| + \|\boldsymbol{\gamma}\| \left\|\hat{\mathbf{h}} - \mathbf{h}\right\|\end{aligned}$$

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

$$+\frac{1}{\pi_h}(\|\hat{\mathbf{Q}}\|(\|\hat{\mathbf{V}}^{-1}\|\|(\hat{\boldsymbol{\Omega}}^{-1}-\boldsymbol{\Omega}^{-1})\mathbf{e}_1\|+\|\hat{\mathbf{V}}^{-1}-\mathbf{V}^{-1}\|\|\boldsymbol{\Omega}^{-1}\mathbf{e}_1\|)$$

$$+\|\hat{\mathbf{Q}}-\mathbf{Q}\|\|\mathbf{V}^{-1}\|\|\boldsymbol{\Omega}^{-1}\mathbf{e}_1\|)\|E_N[\boldsymbol{\xi}_3]\|$$

$$=O\left(h_N^L\right)+O_p\left(\sqrt{\frac{\pi_h^2}{Nh_N^2}}+\sqrt{\frac{\pi_h^2}{N^2h_N^4\pi_{h\backslash 0}}}\right).$$

By (13), we have (A.2).

Third, we are going to show

$$E[\|Var(\boldsymbol{\zeta})^{-1/2}\boldsymbol{\zeta}\|^4]=o(N). \tag{A.3}$$

By Lemmas B11 and B12, $E[\|Var(\boldsymbol{\zeta})^{-1/2}\boldsymbol{\zeta}\|^4]=O\left(\frac{1}{\pi_h^2}\right)$. By (13), we obtain (A.3).

Now, we show the theorem. By Lemmas B11 and B12, and (A.1) and (A.2),

$$\begin{aligned}\sqrt{N}E_N\left[\hat{\boldsymbol{\xi}}\hat{\boldsymbol{\xi}}'\right]^{-1/2}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})&=\sqrt{N}Var(\boldsymbol{\zeta})^{-1/2}(E_N-E)[\boldsymbol{\zeta}]\\&\quad+\sqrt{N}E_N\left[\hat{\boldsymbol{\xi}}\hat{\boldsymbol{\xi}}'\right]^{-1/2}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}-(E_N-E)[\boldsymbol{\zeta}]\right)\\&\quad+\sqrt{N}\left(E_N\left[\hat{\boldsymbol{\xi}}\hat{\boldsymbol{\xi}}'\right]^{-1/2}-Var(\boldsymbol{\zeta})^{-1/2}\right)(E_N-E)[\boldsymbol{\zeta}]\\&=\sqrt{N}Var(\boldsymbol{\zeta})^{-1/2}(E_N-E)[\boldsymbol{\zeta}]+o_p(1).\end{aligned}$$

By (A.3), Assumption 4(i), and Lyapunov's central limit theorem, $\sqrt{N}E_N\left[\hat{\boldsymbol{\xi}}\hat{\boldsymbol{\xi}}'\right]^{-1/2}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\to_d\mathcal{N}\left(0,\mathbf{I}_p\right)$. □

B. AUXILIARY LEMMAS FOR THE PROOF OF THEOREM 2

We introduce and prove the auxiliary results, which we use in the proof of Theorem 2. In all the lemmas to be stated in this section, we impose the assumptions invoked in Theorem 2 as well as $T=p$.

LEMMA B1. $\boldsymbol{\Psi}_{1:L}^{-1}\mathbf{h}=O(\frac{\pi_h}{h_N})$, $\left\|\frac{h_N}{\pi_h}\boldsymbol{\Psi}_{1:L}^{-1}\mathbf{h}\right\|$ *is bounded away from zero and* $\boldsymbol{\Psi}_{1:L}^{-1}(\hat{\mathbf{h}}-\mathbf{h})=O_p\left(\sqrt{\frac{\pi_h}{Nh_N^2}}\right)$.

Proof. The statement follows from

$$\frac{h_N}{\pi_h}\boldsymbol{\Psi}_{1:L}^{-1}\mathbf{h}=\frac{1}{\pi_h}E\left[\mathbf{1}\{|D|\le h_N\}\begin{pmatrix}1&\cdots&\frac{D^{L-1}}{h_N^{L-1}}\end{pmatrix}'\right]=O(1),$$

$$\mathbf{e}_1'\left(\frac{h_N}{\pi_h}\boldsymbol{\Psi}_{1:L}^{-1}\mathbf{h}\right)=\frac{1}{\pi_h}E[\mathbf{1}\{|D|\le h_N\}]=1,$$

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

and

$$E\left[\left\|\boldsymbol{\Psi}_{1:L}^{-1}(\hat{\mathbf{h}}-\mathbf{h})\right\|^2\right] = E\left[\left\|\frac{1}{h_N}(E_N-E)\left[\mathbf{1}\{|D|\le h_N\}\begin{pmatrix}1 & \cdots & \frac{D^{L-1}}{h_N^{L-1}}\end{pmatrix}'\right]\right\|^2\right]$$
$$= \frac{1}{Nh_N^2}E\left[\left\|\mathbf{1}\{|D|\le h_N\}\begin{pmatrix}1 & \cdots & \frac{D^{L-1}}{h_N^{L-1}}\end{pmatrix}'\right\|^2\right]$$
$$= O\left(\frac{\pi_h}{Nh_N^2}\right)$$

under Assumption 4(i). □

LEMMA B2. *(i) For every integer $l \ge 1$,*

$$\frac{1}{h_N^l \pi_{h\backslash 0}} E[D^l \mathbf{1}\{|D|\le h_N\}] = \frac{\alpha_1}{l+\alpha_1}\frac{P(0<D\le h_N)}{\pi_{h\backslash 0}} + (-1)^l \frac{\alpha_2}{l+\alpha_2}\frac{P(-h_N\le D<0)}{\pi_{h\backslash 0}} + o(1)$$

and

$$\frac{1}{h_N^l \pi_{h\backslash 0}} E[|D|^l \mathbf{1}\{|D|\le h_N\}] = \frac{\alpha_1}{l+\alpha_1}\frac{P(0<D\le h_N)}{\pi_{h\backslash 0}} + \frac{\alpha_2}{l+\alpha_2}\frac{P(-h_N\le D<0)}{\pi_{h\backslash 0}} + o(1).$$

(ii) For every positive integer l, $E\left[\left\|\mathbf{D}'_{0:L}\boldsymbol{\Psi}_{0:L}^{-1}\right\|^l\right] = O(\pi_h)$ and $E\left[\left\|\mathbf{D}'_{1:L}\boldsymbol{\Psi}_{1:L}^{-1}\right\|^l\right] = O(\pi_{h\backslash 0})$. (iii) For every pair of positive integers l_1 and l_2, $E\left[|D|^{l_1}\left\|\mathbf{D}'_{0:L}\boldsymbol{\Psi}_{0:L}^{-1}\right\|^{l_2}\right] = O(\pi_{h\backslash 0}h_N^{l_1})$ and $E\left[|D|^{l_1}\left\|\mathbf{D}'_{1:L}\boldsymbol{\Psi}_{1:L}^{-1}\right\|^{l_2}\right] = O(\pi_{h\backslash 0}h_N^{l_1})$. (iv) For every random variable Z in Assumption 4(ii) and for every positive integer l,

$$E\left[\|Z\|\|\mathbf{D}'_{0:L}\boldsymbol{\Psi}_{0:L}^{-1}\|^l\right] = O(\pi_h) \text{ and } E\left[\|Z\|\|\mathbf{D}'_{1:L}\boldsymbol{\Psi}_{1:L}^{-1}\|^l\right] = O(\pi_{h\backslash 0}).$$

Proof. First, we are going to show the statement in (i). Under Assumption 2(iv), Karamata's theorem (Resnick, 2007, Thm. 2.1 and Exerc. 2.5) implies

$$\frac{1}{h_N^l \pi_{h\backslash 0}} E\left[\mathbf{1}\{|D|\le h_N\}D^l\right] = \frac{E\left[\mathbf{1}\{|D|^{-1}\ge h_N^{-1}\}(|D|^{-1})^{-l} \mid D>0\right]}{(h_N^{-1})^{-l}P(|D|^{-1}\ge h_N^{-1}\mid D>0)}\frac{P(0<D\le h_N)}{\pi_{h\backslash 0}}$$
$$+(-1)^l\frac{E\left[\mathbf{1}\{|D|^{-1}\ge h_N^{-1}\}(|D|^{-1})^{-l} \mid D<0\right]}{(h_N^{-1})^{-l}P(|D|^{-1}\ge h_N^{-1}\mid D<0)}$$
$$\frac{P(-h_N\le D<0)}{\pi_{h\backslash 0}}$$
$$= \frac{\alpha_1}{l+\alpha_1}\frac{P(0<D\le h_N)}{\pi_{h\backslash 0}}$$
$$+(-1)^l\frac{\alpha_2}{l+\alpha_2}\frac{P(-h_N\le D<0)}{\pi_{h\backslash 0}} + o(1).$$

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

Similarly,

$$\frac{1}{h_N^l \pi_{h\backslash 0}} E\left[\mathbf{1}\{|D| \le h_N\}|D|^l\right] = \frac{\alpha_1}{l+\alpha_1}\frac{P(0<D\le h_N)}{\pi_{h\backslash 0}} + \frac{\alpha_2}{l+\alpha_2}\frac{P(-h_N\le D<0)}{\pi_{h\backslash 0}} + o(1).$$

Therefore, the statement in (i) holds.

Second, we are going to show the statement in (ii). Note that

$$\left\|\mathbf{D}'_{0:L}\boldsymbol{\Psi}_{0:L}^{-1}\right\|^l = \mathbf{1}\{|D|\le h_N\}\left\|\begin{pmatrix} 1 & \cdots & \frac{D^L}{h_N^L}\end{pmatrix}'\right\|^l \le (L+1)^{l/2}\mathbf{1}\{|D|\le h_N\}$$

and

$$\left\|\mathbf{D}'_{1:L}\boldsymbol{\Psi}_{1:L}^{-1}\right\|^l = \mathbf{1}\{0<|D|\le h_N\}\left\|\begin{pmatrix} \frac{D}{h_N} & \cdots & \frac{D^L}{h_N^L}\end{pmatrix}'\right\|^l \le L^{l/2}\mathbf{1}\{0<|D|\le h_N\}.$$

Therefore, $E\left[\left\|\mathbf{D}'_{0:L}\boldsymbol{\Psi}_{0:L}^{-1}\right\|^l\right] \le (L+1)^{l/2}P(|D|\le h_N) = O(\pi_h)$ and $E\left[\left\|\mathbf{D}'_{1:L}\boldsymbol{\Psi}_{1:L}^{-1}\right\|^l\right] \le L^{l/2}P(0<|D|\le h_N) = O(\pi_{h\backslash 0})$.

Third, we are going to show the statement in (iii). Note that

$$\begin{aligned}|D|^{l_1}\left\|\mathbf{D}'_{0:L}\boldsymbol{\Psi}_{0:L}^{-1}\right\|^{l_2} &= |D|^{l_1}\mathbf{1}\{0<|D|\le h_N\}\left\|\begin{pmatrix} 1 & \cdots & \frac{D^L}{h_N^L}\end{pmatrix}'\right\|^{l_2}\\ &\le (L+1)^{l_2/2}h_N^{l_1}\mathbf{1}\{0<|D|\le h_N\}\end{aligned}$$

and

$$\begin{aligned}|D|^{l_1}\left\|\mathbf{D}'_{1:L}\boldsymbol{\Psi}_{1:L}^{-1}\right\|^{l_2} &= |D|^{l_1}\mathbf{1}\{0<|D|\le h_N\}\left\|\begin{pmatrix} \frac{D}{h_N} & \cdots & \frac{D^L}{h_N^L}\end{pmatrix}'\right\|^{l_2}\\ &\le L^{l_2/2}h_N^{l_1}\mathbf{1}\{0<|D|\le h_N\}.\end{aligned}$$

Therefore,

$$E\left[|D|^{l_1}\left\|\mathbf{D}'_{0:L}\boldsymbol{\Psi}_{0:L}^{-1}\right\|^{l_2}\right] \le (L+1)^{l_2/2}h_N^{l_1}E[\mathbf{1}\{0<|D|\le h_N\}] = O(\pi_{h\backslash 0}h_N^{l_1})$$

and

$$E\left[|D|^{l_1}\left\|\mathbf{D}'_{1:L}\boldsymbol{\Psi}_{1:L}^{-1}\right\|^{l_2}\right] \le L^{l_2/2}h_N^{l_1}E[\mathbf{1}\{0<|D|\le h_N\}] = O(\pi_{h\backslash 0}h_N^{l_1}).$$

Fourth, we are going to show the statement in (iv). There is a positive constant c such that $\sup_{u\in[-c,c]}E[\|Z\|\mid D=u]<\infty$. By the statement in (ii),

$$E\left[\|Z\|\|\mathbf{D}'_{0:L}\boldsymbol{\Psi}_{0:L}^{-1}\|^l\right] \le E\left[\|\mathbf{D}'_{0:L}\boldsymbol{\Psi}_{0:L}^{-1}\|^l\right]\sup_{u\in[-h_N,h_N]}E[\|Z\|\mid D=u] = O(\pi_h), \text{ and}$$

$$E\left[\|Z\|\|\mathbf{D}'_{1:L}\boldsymbol{\Psi}_{1:L}^{-1}\|^l\right] \le E\left[\|\mathbf{D}'_{1:L}\boldsymbol{\Psi}_{1:L}^{-1}\|^l\right]\sup_{u\in[-h_N,h_N]}E[\|Z\|\mid D=u] = O(\pi_{h\backslash 0}). \qquad \square$$

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

LEMMA B3. (i) *For every integer l with $l \geq \max\{\alpha_1, \alpha_2\}$,*

$$\frac{h_N^l}{\pi_{h\backslash 0}} E\Big[\mathbf{1}\{h_N < |D|\}D^{-l}\Big] = \frac{\alpha_1}{l-\alpha_1}\frac{P(0<D\leq h_N)}{\pi_{h\backslash 0}} + (-1)^l \frac{\alpha_2}{l-\alpha_2}\frac{P(-h_N<D<0)}{\pi_{h\backslash 0}} + o(1).$$

(ii) *For every random variable Z in Assumption 4(ii) and for every positive integer l,*

$$E\Big[\mathbf{1}\{|D|>h_N\}|D|^{-1}\|Z\|\Big] = O\left(\frac{\pi_{h\backslash 0}^{1-\epsilon}}{h_N}\right) \text{ for every number } \epsilon \in (0,1),$$

$$E\Big[\mathbf{1}\{|D|>h_N\}D^{-2}\|Z\|\Big] = O\left(\frac{\pi_{h\backslash 0}}{h_N^2}\right), \text{ and}$$

$$E\Big[\mathbf{1}\{|D|>h_N\}D^{-4}\|Z\|\Big] = O\left(\frac{\pi_{h\backslash 0}}{h_N^4}\right).$$

Proof. Under Assumption 2(iv), Karamata's theorem (Resnick, 2007, Thm. 2.1 and Exerc. 2.5) implies that

$$\begin{aligned}\frac{h_N^l}{\pi_{h\backslash 0}} E\Big[\mathbf{1}\{|D|>h_N\}D^{-l}\Big] =& \frac{E\Big[\mathbf{1}\{|D|^{-1}\leq h_N^{-1}\}(|D|^{-1})^l \mid D>0\Big]}{(h_N^{-1})^l P(|D|^{-1}\geq h_N^{-1}\mid D>0)}\frac{P(0<D\leq h_N)}{\pi_{h\backslash 0}} \\ &+(-1)^l \frac{E\Big[\mathbf{1}\{|D|^{-1}\leq h_N^{-1}\}(|D|^{-1})^l \mid D<0\Big]}{(h_N^{-1})^l P(|D|^{-1}\geq h_N^{-1}\mid D<0)} \\ &\frac{P(-h_N<D<0)}{\pi_{h\backslash 0}} \\ =& \frac{\alpha_1}{l-\alpha_1}\frac{P(0<D\leq h_N)}{\pi_{h\backslash 0}} + (-1)^l \frac{\alpha_2}{l-\alpha_2} \\ &\frac{P(-h_N<D<0)}{\pi_{h\backslash 0}} + o(1),\end{aligned}$$

which proves the statement in (i). We are going to show the statement in (ii). There is a positive constant c such that $\sup_{u\in[-c,c]} E[\|Z\| \mid D=u] < \infty$. By the triangle inequality and Hölder's inequality,

$$\begin{aligned}\Big\|E\Big[\mathbf{1}\{|D|>h_N\}|D|^{-1}\|Z\|\Big]\Big\| \leq& \Big\|E\Big[\mathbf{1}\{|D|>c\}|D|^{-1}\|Z\|\Big]\Big\| \\ &+\Big\|E\Big[\mathbf{1}\{h_N<|D|\leq c\}|D|^{-1}\|Z\|\Big]\Big\| \\ \leq& c^{-1}E[\|Z\|] + E\Big[\mathbf{1}\{h_N<|D|\}|D|^{-1}\Big]\sup_{u\in[-c,c]} \\ &E[\|Z\| \mid D=u]\end{aligned}$$

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

$$\leq c^{-1}E[\|Z\|]+E\left[\mathbf{1}\{h_N<|D|\}|D|^{-1/1-\epsilon}\right]^{1-\epsilon}\sup_{u\in[-c,c]}E[\|Z\|\mid D=u],$$

$$\left\|E\left[\mathbf{1}\{|D|>h_N\}|D|^{-2}\|Z\|\right]\right\| \leq \left\|E\left[\mathbf{1}\{|D|>c\}|D|^{-2}\|Z\|\right]\right\| + \left\|E\left[\mathbf{1}\{h_N<|D|\}|D|^{-2}\|Z\|\right]\right\|$$
$$\leq c^{-2}E[\|Z\|]+E\left[\mathbf{1}\{h_N<|D|\}|D|^{-2}\right]\sup_{u\in[-c,c]}E[\|Z\|\mid D=u],$$

and

$$\left\|E\left[\mathbf{1}\{|D|>h_N\}|D|^{-4}\|Z\|\right]\right\| \leq \left\|E\left[\mathbf{1}\{|D|>c\}|D|^{-4}\|Z\|\right]\right\| + \left\|E\left[\mathbf{1}\{h_N<|D|\}|D|^{-4}\|Z\|\right]\right\|$$
$$\leq c^{-4}E[\|Z\|]+E\left[\mathbf{1}\{h_N<|D|\}|D|^{-4}\right]\sup_{u\in[-c,c]}E[\|Z\|\mid D=u].$$

By the statement in (i), the statement in (ii) holds. □

LEMMA B4. *The minimum eigenvalue of* $\mathbf{\Pi}$ *is bounded away from zero, the maximum eigenvalue of* $\mathbf{\Pi}$ *is bounded, and* $\hat{\mathbf{\Pi}}^{-1}-\mathbf{\Pi}^{-1}=O_p\left(\frac{1}{\sqrt{N\pi_{h\backslash 0}}}\right)$.

Proof. First, we are going to show the first two statements. Let $\mathbf{c}$ be any L-dimensional column vector with $\|\mathbf{c}\|=1$. Define $\tilde{\mathbf{c}}=\text{diag}(1,-1,1,-1,\ldots)\mathbf{c}$. Note that $\|\tilde{\mathbf{c}}\|=1$. Define

$$\mathbf{C}_1=\begin{pmatrix} \frac{\alpha_1}{\alpha_1+2} & \cdots & \frac{\alpha_1}{\alpha_1+L+1} \\ \vdots & \ddots & \vdots \\ \frac{\alpha_1}{\alpha_1+L+1} & \cdots & \frac{\alpha_1}{\alpha_1+2L} \end{pmatrix} \quad \text{and} \quad \mathbf{C}_2=\begin{pmatrix} \frac{\alpha_2}{\alpha_2+2} & \cdots & \frac{\alpha_2}{\alpha_2+L+1} \\ \vdots & \ddots & \vdots \\ \frac{\alpha_2}{\alpha_2+L+1} & \cdots & \frac{\alpha_2}{\alpha_2+2L} \end{pmatrix}.$$

Note that $\mathbf{C}_1$ and $\mathbf{C}_2$ are Cauchy matrices and positive definite. By Lemma B2,

$$\mathbf{c}'\mathbf{\Pi}\mathbf{c}=\frac{P(0<D\leq h_N)}{\pi_{h\backslash 0}}\mathbf{c}'\mathbf{C}_1\mathbf{c}+\frac{P(-h_N\leq D<0)}{\pi_{h\backslash 0}}\tilde{\mathbf{c}}'\mathbf{C}_2\tilde{\mathbf{c}}+o(1),$$

which implies that $\mathbf{c}'\mathbf{\Pi}\mathbf{c}$ is bounded away from zero. Therefore, the minimum eigenvalue of $\mathbf{\Pi}$ is bounded away from zero and the maximum eigenvalue of $\mathbf{\Pi}$ is bounded as $h_N\to 0$.

Second, we are going to show $\hat{\mathbf{\Pi}}-\mathbf{\Pi}=O_p\left(\frac{1}{\sqrt{N\pi_{h\backslash 0}}}\right)$. Since $\hat{\mathbf{\Pi}}-\mathbf{\Pi}=\frac{1}{\pi_{h\backslash 0}}(E_N-E)\left[\mathbf{\Psi}_{1:L}^{-1}\mathbf{D}_{1:L}\mathbf{D}_{1:L}'\mathbf{\Psi}_{1:L}^{-1}\right]$, by Assumption 4(i) and Lemma B2,

$$E\left[\left\|\hat{\mathbf{\Pi}}-\mathbf{\Pi}\right\|^2\right]^{1/2}=\frac{1}{\sqrt{N\pi_{h\backslash 0}^2}}E\left[\left\|\mathbf{D}_{1:L}'\mathbf{\Psi}_{1:L}^{-1}\right\|^4\right]^{1/2}=O\left(\frac{1}{\sqrt{N\pi_{h\backslash 0}}}\right).$$

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

Third, we are going to show $\hat{\mathbf{\Pi}}^{-1} - \mathbf{\Pi}^{-1} = O_p\left(\frac{1}{\sqrt{N\pi_{h\backslash 0}}}\right)$. Note that $\hat{\mathbf{\Pi}}^{-1} - \mathbf{\Pi}^{-1} = -(\mathbf{\Pi} + (\hat{\mathbf{\Pi}} - \mathbf{\Pi}))^{-1}(\hat{\mathbf{\Pi}} - \mathbf{\Pi})\mathbf{\Pi}^{-1}$. Since the minimum eigenvalue of $\mathbf{\Pi}$ is bounded away from zero and $\hat{\mathbf{\Pi}} - \mathbf{\Pi} = o_p(1)$, the matrix $(\mathbf{\Pi} + (\hat{\mathbf{\Pi}} - \mathbf{\Pi}))^{-1}$ is $O_p(1)$. Since $\hat{\mathbf{\Pi}} - \mathbf{\Pi} = \frac{1}{\pi_{h\backslash 0}}(E_N - E)\left[\mathbf{\Psi}_{1:L}^{-1}\mathbf{D}_{1:L}\mathbf{D}_{1:L}'\mathbf{\Psi}_{1:L}^{-1}\right] = O_p\left(\frac{1}{\sqrt{N\pi_{h\backslash 0}}}\right)$ by Lemma B2, we have $\hat{\mathbf{\Pi}}^{-1} - \mathbf{\Pi}^{-1} = O_p\left(\frac{1}{\sqrt{N\pi_{h\backslash 0}}}\right)$. □

LEMMA B5. $\mathbf{\Omega}^{-1}\mathbf{e}_1 = O(1)$ *and* $\left(\hat{\mathbf{\Omega}}^{-1} - \mathbf{\Omega}^{-1}\right)\mathbf{e}_1 = O_p\left(\sqrt{\frac{\pi_h}{N\pi_{h\backslash 0}^2}}\right)$.

Proof. First, the minimum eigenvalue of $\bar{\mathbf{C}}_N$ is positive and bounded away from zero for sufficiently large N, where

$$\bar{\mathbf{C}}_N = \begin{bmatrix} 1 & \frac{1}{\pi_{h\backslash 0}}E[\mathbf{D}_{1:L}'\mathbf{\Psi}_{1:L}^{-1}] \\ \frac{1}{\pi_{h\backslash 0}}E[\mathbf{D}_{1:L}'\mathbf{\Psi}_{1:L}^{-1}]' & \mathbf{\Pi} \end{bmatrix}.$$

Let $\mathbf{c}$ be any $(L+1)$-dimensional column vector with $\|\mathbf{c}\| = 1$. Define $\tilde{\mathbf{c}} = \mathrm{diag}(1, -1, 1, -1, \ldots)\mathbf{c}$. Note that $\|\tilde{\mathbf{c}}\| = 1$. Define

$$\tilde{\mathbf{C}}_1 = \begin{pmatrix} \frac{\alpha_1}{\alpha_1} & \cdots & \frac{\alpha_1}{\alpha_1+L} \\ \vdots & \ddots & \vdots \\ \frac{\alpha_1}{\alpha_1+L} & \cdots & \frac{\alpha_1}{\alpha_1+2L} \end{pmatrix} \text{ and } \tilde{\mathbf{C}}_2 = \begin{pmatrix} \frac{\alpha_2}{\alpha_2} & \cdots & \frac{\alpha_2}{\alpha_2+L} \\ \vdots & \ddots & \vdots \\ \frac{\alpha_2}{\alpha_2+L} & \cdots & \frac{\alpha_2}{\alpha_2+2L} \end{pmatrix}.$$

Note that $\tilde{\mathbf{C}}_1$ and $\tilde{\mathbf{C}}_2$ are Cauchy matrices and positive definite. By Lemma B2,

$$\mathbf{c}'\bar{\mathbf{C}}_N\mathbf{c} = \frac{P(0 < D \le h_N)}{\pi_{h\backslash 0}}\mathbf{c}'\tilde{\mathbf{C}}_1\mathbf{c} + \frac{P(-h_N \le D < 0)}{\pi_{h\backslash 0}}\tilde{\mathbf{c}}'\tilde{\mathbf{C}}_2\tilde{\mathbf{c}} + o(1),$$

which implies that $\mathbf{c}'\bar{\mathbf{C}}_N\mathbf{c}$ is bounded away from zero. Therefore, the minimum eigenvalue of $\bar{\mathbf{C}}_N$ is bounded away from zero.

Second, we are going to show $\mathbf{\Omega}^{-1}\mathbf{e}_1 = O(1)$. Since

$$\mathbf{\Omega} = \begin{bmatrix} 1 & \frac{1}{\pi_h}E[\mathbf{D}_{1:L}'\mathbf{\Psi}_{1:L}^{-1}] \\ \frac{1}{\pi_h}E[\mathbf{D}_{1:L}'\mathbf{\Psi}_{1:L}^{-1}]' & \frac{\pi_{h\backslash 0}}{\pi_h}\mathbf{\Pi} \end{bmatrix},$$

the inverse matrix formula for block matrices implies

$$\begin{aligned} \mathbf{\Omega}^{-1}\mathbf{e}_1 &= \begin{bmatrix} \left(1 - \frac{1}{\pi_h\pi_{h\backslash 0}}E[\mathbf{D}_{1:L}'\mathbf{\Psi}_{1:L}^{-1}]\mathbf{\Pi}^{-1}E[\mathbf{D}_{1:L}'\mathbf{\Psi}_{1:L}^{-1}]'\right)^{-1} \\ -\frac{1}{\pi_{h\backslash 0}}\mathbf{\Pi}^{-1}E[\mathbf{D}_{1:L}'\mathbf{\Psi}_{1:L}^{-1}]'\left(1 - \frac{1}{\pi_h\pi_{h\backslash 0}}E[\mathbf{D}_{1:L}'\mathbf{\Psi}_{1:L}^{-1}]\mathbf{\Pi}^{-1}E[\mathbf{D}_{1:L}'\mathbf{\Psi}_{1:L}^{-1}]'\right)^{-1} \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ -\frac{1}{\pi_{h\backslash 0}}\mathbf{\Pi}^{-1}E[\mathbf{D}_{1:L}'\mathbf{\Psi}_{1:L}^{-1}] \end{bmatrix}\left(1 - \frac{1}{\pi_h\pi_{h\backslash 0}}E[\mathbf{D}_{1:L}'\mathbf{\Psi}_{1:L}^{-1}]\mathbf{\Pi}^{-1}E[\mathbf{D}_{1:L}'\mathbf{\Psi}_{1:L}^{-1}]'\right)^{-1} \end{aligned} \quad \textbf{(B.2)}$$

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

as long as $\mathbf{\Pi}$ is invertible and $1-\frac{1}{\pi_h\pi_{h\backslash 0}}E[\mathbf{D}'_{1:L}\mathbf{\Psi}^{-1}_{1:L}]\mathbf{\Pi}^{-1}E[\mathbf{D}'_{1:L}\mathbf{\Psi}^{-1}_{1:L}]'\neq 0$. It suffices to show that $1-\frac{1}{\pi_h\pi_{h\backslash 0}}E[\mathbf{D}'_{1:L}\mathbf{\Psi}^{-1}_{1:L}]\mathbf{\Pi}^{-1}E[\mathbf{D}'_{1:L}\mathbf{\Psi}^{-1}_{1:L}]'$ is bounded away from zero for sufficiently large N. By the determinant formula for block matrices, we have

$$1-\frac{1}{\pi^2_{h\backslash 0}}E[\mathbf{D}'_{1:L}\mathbf{\Psi}^{-1}_{1:L}]\mathbf{\Pi}^{-1}E[\mathbf{D}'_{1:L}\mathbf{\Psi}^{-1}_{1:L}]'=\frac{\det\left(\bar{\mathbf{C}}_N\right)}{\det(\mathbf{\Pi})}.$$

Since the minimum eigenvalue of $\bar{\mathbf{C}}_N$ is bounded away from zero, $1-\frac{1}{\pi^2_{h\backslash 0}}E[\mathbf{D}'_{1:L}\mathbf{\Psi}^{-1}_{1:L}]\mathbf{\Pi}^{-1}E[\mathbf{D}'_{1:L}\mathbf{\Psi}^{-1}_{1:L}]'$ is positive and bounded away from zero. Since

$$1-\frac{1}{\pi_h\pi_{h\backslash 0}}E[\mathbf{D}'_{1:L}\mathbf{\Psi}^{-1}_{1:L}]\mathbf{\Pi}^{-1}E[\mathbf{D}'_{1:L}\mathbf{\Psi}^{-1}_{1:L}]'\geq 1-\frac{1}{\pi^2_{h\backslash 0}}E[\mathbf{D}'_{1:L}\mathbf{\Psi}^{-1}_{1:L}]\mathbf{\Pi}^{-1}E[\mathbf{D}'_{1:L}\mathbf{\Psi}^{-1}_{1:L}]',$$

it follows that $1-\frac{1}{\pi_h\pi_{h\backslash 0}}E[\mathbf{D}'_{1:L}\mathbf{\Psi}^{-1}_{1:L}]\mathbf{\Pi}^{-1}E[\mathbf{D}'_{1:L}\mathbf{\Psi}^{-1}_{1:L}]'$ is bounded away from zero.

Third, we are going to show that the minimum eigenvalue of $\frac{\pi_h}{\pi_{h\backslash 0}}\mathbf{\Omega}$ is bounded away from zero. Note that

$$\frac{\pi_h}{\pi_{h\backslash 0}}\mathbf{\Omega}=\bar{\mathbf{C}}_N+\frac{\pi_h-\pi_{h\backslash 0}}{\pi_{h\backslash 0}}\mathbf{e}_1\mathbf{e}'_1.$$

The minimum eigenvalue of $\bar{\mathbf{C}}_N$ is bounded away from zero, and $\frac{\pi_h-\pi_{h\backslash 0}}{\pi_{h\backslash 0}}\mathbf{e}_1\mathbf{e}'_1$ is positive semidefinite. Therefore, the minimum eigenvalue of $\frac{\pi_h}{\pi_{h\backslash 0}}\mathbf{\Omega}$ is bounded away from zero.

Fourth, we are going to show that $\frac{\pi_h}{\pi_{h\backslash 0}}(\hat{\mathbf{\Omega}}-\mathbf{\Omega})=O_p\left(\sqrt{\frac{\pi_h}{N\pi^2_{h\backslash 0}}}\right)$. Since $\frac{\pi_h}{\pi_{h\backslash 0}}(\hat{\mathbf{\Omega}}-\mathbf{\Omega})=\frac{1}{\pi_{h\backslash 0}}(E_N-E)\left[\mathbf{\Psi}^{-1}_{0:L}\mathbf{D}_{0:L}\mathbf{D}'_{0:L}\mathbf{\Psi}^{-1}_{0:L}\right]$, by Assumption 4(i) and Lemma B2,

$$E\left[\left\|\frac{\pi_h}{\pi_{h\backslash 0}}(\hat{\mathbf{\Omega}}-\mathbf{\Omega})\right\|^2\right]^{1/2}=\frac{1}{\sqrt{N\pi^2_{h\backslash 0}}}E\left[\left\|\mathbf{D}'_{0:L}\mathbf{\Psi}^{-1}_{0:L}\right\|^4\right]^{1/2}=O\left(\sqrt{\frac{\pi_h}{N\pi^2_{h\backslash 0}}}\right).$$

Fifth, we are going to show $\left(\hat{\mathbf{\Omega}}^{-1}-\mathbf{\Omega}^{-1}\right)\mathbf{e}_1=O_p\left(\sqrt{\frac{\pi_h}{N\pi^2_{h\backslash 0}}}\right)$. Note that

$$\left(\hat{\mathbf{\Omega}}^{-1}-\mathbf{\Omega}^{-1}\right)\mathbf{e}_1=-\left(\frac{\pi_h}{\pi_{h\backslash 0}}\mathbf{\Omega}+\frac{\pi_h}{\pi_{h\backslash 0}}(\hat{\mathbf{\Omega}}-\mathbf{\Omega})\right)^{-1}\frac{\pi_h}{\pi_{h\backslash 0}}\left(\hat{\mathbf{\Omega}}-\mathbf{\Omega}\right)\mathbf{\Omega}^{-1}\mathbf{e}_1.$$

Since the minimum eigenvalue of $\frac{\pi_h}{\pi_{h\backslash 0}}\mathbf{\Omega}$ is bounded away from zero and $\frac{\pi_h}{\pi_{h\backslash 0}}(\hat{\mathbf{\Omega}}-\mathbf{\Omega})=o_p(1)$, the matrix $\left(\frac{\pi_h}{\pi_{h\backslash 0}}\mathbf{\Omega}+\frac{\pi_h}{\pi_{h\backslash 0}}(\hat{\mathbf{\Omega}}-\mathbf{\Omega})\right)^{-1}$ is $O_p(1)$. Since $\frac{\pi_h}{\pi_{h\backslash 0}}\left(\hat{\mathbf{\Omega}}-\mathbf{\Omega}\right)=O_p\left(\sqrt{\frac{\pi_h}{N\pi^2_{h\backslash 0}}}\right)$ and $\mathbf{\Omega}^{-1}\mathbf{e}_1=O(1)$, we have $\left(\hat{\mathbf{\Omega}}^{-1}-\mathbf{\Omega}^{-1}\right)\mathbf{e}_1=O_p\left(\sqrt{\frac{\pi_h}{N\pi^2_{h\backslash 0}}}\right)$. □

LEMMA B6. $\mathbf{V}^{-1}=O(1)$ *and* $\hat{\mathbf{V}}^{-1}-\mathbf{V}^{-1}=O_p\left(\sqrt{\frac{\pi_h}{N\pi^2_{h\backslash 0}}}\right)$.

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

Proof. First, we are going to show $\mathbf{V}^{-1} = O(1)$. It suffices to investigate the minimum eigenvalue of $\mathbf{V}$. Let $\mathbf{c}$ be any $(T-1)p$-dimensional column vector with $\|c\| = 1$. Define $m_1(u) = E\left[\|\mathbf{X}^*\mathbf{W}\mathbf{c}\|^2 \mid D = u\right]$. Assumption 3(ii) implies that there is a constant C such that $|m_1(u) - m_1(0)| \leq C|u|$ for every $u \in [-h_N, h_N]$ when h_N is sufficiently small. Since $\mathbf{c}'\mathbf{V}\mathbf{c} - m_1(0) = \mathbf{e}_1'\mathbf{\Omega}^{-1}\frac{1}{\pi_h}E\left[\mathbf{\Psi}_{0:L}^{-1}\mathbf{D}_{0:L}(m_1(D) - m_1(0))\right]$, we have

$$\left|\mathbf{c}'\mathbf{V}\mathbf{c} - m_1(0)\right| \leq \|\mathbf{\Omega}^{-1}\mathbf{e}_1\|\frac{1}{\pi_h}E\left[|D|\|\mathbf{D}_{0:L}'\mathbf{\Psi}_{0:L}^{-1}\|\right] = o(1)$$

by Lemmas B2 and B5. Since $m_1(0) > 0$ by Assumption 4(iii), it follows that $\mathbf{c}'\mathbf{V}\mathbf{c}$ is bounded away from zero.

Next, we are going to show $\hat{\mathbf{V}}^{-1} - \mathbf{V}^{-1} = O_p\left(\sqrt{\frac{\pi_h}{N\pi_{h\backslash 0}^2}}\right)$. Since $\mathbf{V}^{-1} = O(1)$, it suffices to show $\hat{\mathbf{V}} - \mathbf{V} = O_p\left(\sqrt{\frac{\pi_h}{N\pi_{h\backslash 0}^2}}\right)$. Let $\mathbf{c}$ be any $(T-1)p$-dimensional column vector with $\|c\| = 1$. Then,

$$\begin{aligned}(\hat{\mathbf{V}} - \mathbf{V})\mathbf{c} = &\frac{1}{\pi_h}E\left[(\mathbf{X}^*\mathbf{W})'\mathbf{X}^*\mathbf{W}\mathbf{c}\mathbf{D}_{0:L}'\mathbf{\Psi}_{0:L}^{-1}\right]\left(\hat{\mathbf{\Omega}}^{-1} - \mathbf{\Omega}^{-1}\right)\mathbf{e}_1 \\ &+ \frac{1}{\pi_h}(E_N - E)\left[(\mathbf{X}^*\mathbf{W})'\mathbf{X}^*\mathbf{W}\mathbf{c}\mathbf{D}_{0:L}'\mathbf{\Psi}_{0:L}^{-1}\right]\hat{\mathbf{\Omega}}^{-1}\mathbf{e}_1.\end{aligned}$$

Since Assumption 4 and Lemma B2 imply

$$\|E\left[(\mathbf{X}^*\mathbf{W})'\mathbf{X}^*\mathbf{W}\mathbf{c}\mathbf{D}_{0:L}'\mathbf{\Psi}_{0:L}^{-1}\right]\| \leq E\left[\|\mathbf{X}^*\mathbf{W}\|^2\|\mathbf{D}_{0:L}'\mathbf{\Psi}_{0:L}^{-1}\|\right] = O(\pi_h)$$

and

$$\begin{aligned}E\left[\left\|(E_N - E)\left[(\mathbf{X}^*\mathbf{W})'\mathbf{X}^*\mathbf{W}\mathbf{c}\mathbf{D}_{0:L}'\mathbf{\Psi}_{0:L}^{-1}\right]\right\|^2\right]^{1/2} &\leq \frac{1}{\sqrt{N}}E\left[\|\mathbf{X}^*\mathbf{W}\|^4\left\|\mathbf{D}_{0:L}'\mathbf{\Psi}_{0:L}^{-1}\right\|^2\right]^{1/2} \\ &= O\left(\sqrt{\frac{\pi_h}{N}}\right),\end{aligned}$$

together with Lemma B5, we have $(\hat{\mathbf{V}} - \mathbf{V})\mathbf{c} = O_p\left(\sqrt{\frac{\pi_h}{N\pi_{h\backslash 0}^2}}\right)$. □

LEMMA B7. $\mathbf{Q} = O(\frac{\pi_h^{1-\epsilon}}{h_N})$ *for every positive number* $\epsilon \in (0,1)$, *and* $\hat{\mathbf{Q}} - \mathbf{Q} = O_p\left(\sqrt{\frac{\pi_h^2}{Nh_N^2\pi_{h\backslash 0}}}\right)$.

Proof. Let $\mathbf{c}$ be any $(T-1)p$-dimensional column vector with $\|\mathbf{c}\| = 1$. The first statement follows from Lemmas B1 and B3 and

$$\mathbf{Q}\mathbf{c} = \mathbf{R}\mathbf{c} - E\left[\mathbf{1}\{|D| > h_N\}D^{-1}\mathbf{X}^*\mathbf{W}\right]\mathbf{c} - \frac{1}{\pi_{h\backslash 0}}E\left[\mathbf{X}^*\mathbf{W}\mathbf{c}\mathbf{D}_{1:L}'\mathbf{\Psi}_{1:L}^{-1}\right]\mathbf{\Pi}^{-1}\mathbf{\Psi}_{1:L}^{-1}\mathbf{h}.$$

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

We are going to show the second statement in the rest of the proof. Assumption 4 and Lemma B2 imply

$$\|E\left[\mathbf{X}^*\mathbf{W}\mathbf{c}\mathbf{D}'_{1:L}\boldsymbol{\Psi}_{1:L}^{-1}\right]\| \leq E\left[\|\mathbf{X}^*\mathbf{W}\mathbf{c}\|\|\mathbf{D}'_{1:L}\boldsymbol{\Psi}_{1:L}^{-1}\|\right] = O\left(\pi_{h\backslash 0}\right),$$

$$\begin{aligned} E\left[\left\|(E_N - E)\left[\mathbf{X}^*\mathbf{W}\mathbf{c}\mathbf{D}'_{1:L}\boldsymbol{\Psi}_{1:L}^{-1}\right]\right\|^2\right]^{1/2} &\leq \frac{1}{\sqrt{N}} E\left[\left\|\mathbf{X}^*\mathbf{W}\mathbf{c}\right\|^2 \left\|\mathbf{D}'_{1:L}\boldsymbol{\Psi}_{1:L}^{-1}\right\|^2\right]^{1/2} \\ &= O\left(\sqrt{\frac{\pi_{h\backslash 0}}{N}}\right), \end{aligned}$$

and

$$\begin{aligned} E\left[\left\|(E_N - E)\left[\mathbf{X}^*\mathbf{W}\mathbf{1}\{|D| > h_N\}D^{-1}\right]\right\|^2\right]^{1/2} &\leq \frac{1}{\sqrt{N}} E\left[\left\|\mathbf{X}^*\mathbf{W}\right\|^2 \mathbf{1}\{|D| > h_N\}D^{-2}\right]^{1/2} \\ &= O\left(\sqrt{\frac{\pi_{h\backslash 0}}{Nh_N^2}}\right). \end{aligned}$$

Since

$$\begin{aligned} (\hat{\mathbf{Q}} - \mathbf{Q})\mathbf{c} = &\frac{1}{\pi_{h\backslash 0}} E\left[\mathbf{X}^*\mathbf{W}\mathbf{c}\mathbf{D}'_{1:L}\boldsymbol{\Psi}_{1:L}^{-1}\right](\hat{\boldsymbol{\Pi}}^{-1} - \boldsymbol{\Pi}^{-1})\boldsymbol{\Psi}_{1:L}^{-1}\mathbf{h} \\ &+ \frac{1}{\pi_{h\backslash 0}}(E_N - E)\left[\mathbf{X}^*\mathbf{W}\mathbf{c}\mathbf{D}'_{1:L}\boldsymbol{\Psi}_{1:L}^{-1}\right]\hat{\boldsymbol{\Pi}}^{-1}\boldsymbol{\Psi}_{1:L}^{-1}\mathbf{h} \\ &+ (E_N - E)\left[\mathbf{X}^*\mathbf{W}\mathbf{1}\{|D| > h_N\}D^{-1}\right]\mathbf{c}, \end{aligned}$$

together with Lemmas B1 and B4, we have $(\hat{\mathbf{Q}} - \mathbf{Q})\mathbf{c} = O_p\left(\sqrt{\frac{\pi_h^2}{Nh_N^2\pi_{h\backslash 0}}}\right)$. □

LEMMA B8. *(i)* $E[\boldsymbol{\xi}_1] = 0$. *(ii)* $E[\|\boldsymbol{\xi}_1\|^2] = O\left(\frac{\pi_{h\backslash 0}}{h_N^2}\right)$ *and* $E[\|\boldsymbol{\xi}_1\|^4] = O\left(\frac{\pi_{h\backslash 0}}{h_N^4}\right)$. *(iii)* $(E_N - E)[\boldsymbol{\xi}_1] = O_p\left(\sqrt{\frac{\pi_{h\backslash 0}}{Nh_N^2}}\right)$.

Proof. The statement in (i) follows from the definition of $\boldsymbol{\xi}_1$. The statement in (ii) follows from Assumption 4(ii) and Lemma B3. The statement in (iii) follows from Assumption 4(i) and the statement in (ii). □

LEMMA B9. *(i)* $E[\boldsymbol{\xi}_2] = O(h_N^{L+1}\pi_{h\backslash 0})$. *(ii)* $E[\|\boldsymbol{\xi}_2\|^2] = O\left(\pi_{h\backslash 0}\right)$ *and* $E[\|\boldsymbol{\xi}_2\|^4] = O\left(\pi_{h\backslash 0}\right)$. *(iii)* $(E_N - E)[\boldsymbol{\xi}_2] = O\left(\sqrt{\frac{\pi_{h\backslash 0}}{N}}\right)$.

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

Proof. By Assumption 3(i) and $\mathbf{m}(0) = 0$, there is a constant C such that $\|\mathbf{m}(u) - \sum_{l=1}^{L} \mathbf{m}^{(l)}(0)\frac{u^l}{l!}\| \leq C|u|^{L+1}$ for every $u \in [-h_N, h_N]$ when h_N is sufficiently small. Then,

$$\begin{aligned}\left\|E\left[\boldsymbol{\xi}_2\right]\right\| &= \left\|E\left[\left(\mathbf{m}(D) - \sum_{l=1}^{L}\mathbf{m}^{(l)}(0)\frac{D^l}{l!}\right)\mathbf{D}'_{1:L}\boldsymbol{\Psi}_{1:L}^{-1}\right]\right\| \\ &\leq E\left[\left\|\mathbf{m}(D) - \sum_{l=1}^{L}\mathbf{m}^{(l)}(0)\frac{D^l}{l!}\right\|\left\|\mathbf{D}'_{1:L}\boldsymbol{\Psi}_{1:L}^{-1}\right\|\right] \\ &\leq CE\left[D^{L+1}\left\|\mathbf{D}'_{1:L}\boldsymbol{\Psi}_{1:L}^{-1}\right\|\right] \\ &\leq O(\pi_{h\setminus 0}h_N^{L+1}).\end{aligned}$$

The statement in (ii) follows from Assumption 4(ii) and Lemma B2. The statement in (iii) follows from Assumption 4(i) and the statement in (ii). □

LEMMA B10. *(i)* $E\left[\boldsymbol{\xi}_3\right] = O\left(h_N\pi_{h\setminus 0}\right)$ *and* $E\left[\boldsymbol{\xi}_3\boldsymbol{\Omega}^{-1}\mathbf{e}_1\right] = O\left(h_N^{L+1}\pi_{h\setminus 0}\right)$. *(ii)* $E[\|\boldsymbol{\xi}_3\|^2] = O(\pi_h)$ *and* $E[\|\boldsymbol{\xi}_3\|^4] = O(\pi_h)$. *(iii)* $(E_N - E)\left[\boldsymbol{\xi}_3\right] = O_p\left(\sqrt{\frac{\pi_h}{N}}\right)$.

Proof. Define $m_2(u) = E\left[(\mathbf{X}^*\mathbf{W})'\mathbf{X}^*(\mathbf{Y} - \mathbf{W}\boldsymbol{\delta}) \mid D = u\right]$. Note that $m_2(0) = E[(\mathbf{X}^*\mathbf{W})'D\boldsymbol{\beta}(\mathbf{X}) \mid D = 0] = 0$ by (2), and Assumption 3(i) implies that there is a constant C such that $\left\|m_2(u) - \sum_{l=0}^{L} m_2^{(l)}(0)\frac{u^l}{l!}\right\| \leq C|u|^{L+1}$ for every $u \in [-h_N, h_N]$ when h_N is sufficiently small. Then,

$$\begin{aligned}\left\|E\left[\boldsymbol{\xi}_3\right]\right\| &= \left\|E\left[(m_2(D) - m_2(0))\mathbf{D}'_{0:L}\boldsymbol{\Psi}_{0:L}^{-1}\right]\right\| \\ &\leq \left\|E\left[\left(\sum_{l=1}^{L} m_2^{(l)}(0)\frac{D^l}{l!}\right)\mathbf{D}'_{0:L}\boldsymbol{\Psi}_{0:L}^{-1}\right]\right\| + C\left\|E\left[|D|^{L+1}\mathbf{D}'_{0:L}\boldsymbol{\Psi}_{0:L}^{-1}\right]\right\|\end{aligned}$$

and

$$\begin{aligned}\left\|E\left[\boldsymbol{\xi}_3\right]\boldsymbol{\Omega}^{-1}\mathbf{e}_1\right\| \leq{}& \left\|E\left[\left(m_2(D) - \sum_{l=0}^{L} m_2^{(l)}(0)\frac{D^l}{l!}\right)\mathbf{D}'_{0:L}\boldsymbol{\Psi}_{0:L}^{-1}\right]\boldsymbol{\Omega}^{-1}\mathbf{e}_1\right\| \\ &+ C\left\|E\left[|D|^{L+1}\mathbf{D}'_{0:L}\boldsymbol{\Psi}_{0:L}^{-1}\right]\boldsymbol{\Omega}^{-1}\mathbf{e}_1\right\|.\end{aligned}$$

By Lemmas B2 and B5, the statement in (i) holds. The statement in (ii) follows from Assumption 4(ii) and Lemma B2. The statement in (iii) follows from Assumption 4(i) and the statement in (ii). □

LEMMA B11. *The following statements hold for every positive number* $\epsilon \in (0, 1)$. *(i)* $E[\boldsymbol{\zeta}] = O\left(h_N^L\right)$. *(ii)* $E[\|\boldsymbol{\zeta}\|^2] = O\left(\frac{\pi_h^{1-\epsilon}}{h_N^2}\right)$ *and* $E[\|\boldsymbol{\zeta}\|^4] = O\left(\frac{\pi_h^{1-\epsilon}}{h_N^4}\right)$. *(iii)* $(E_N - E)[\boldsymbol{\zeta}] = O_p\left(\sqrt{\frac{\pi_h^{1-\epsilon}}{Nh_N^2}}\right)$ *and* $(E_N - E)\left[\boldsymbol{\zeta}\boldsymbol{\zeta}'\right] = O_p\left(\sqrt{\frac{\pi_h^{1-\epsilon}}{Nh_N^4}}\right)$.

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

Proof. Note that

$$E[\boldsymbol{\zeta}] = \frac{1}{\pi_{h\backslash 0}} E[\boldsymbol{\xi}_2]\boldsymbol{\Pi}^{-1}\boldsymbol{\Psi}_{1:L}^{-1}\mathbf{h} + \frac{1}{\pi_h}\mathbf{Q}\mathbf{V}^{-1}E[\boldsymbol{\xi}_3]\boldsymbol{\Omega}^{-1}\mathbf{e}_1.$$

By Lemmas B1, B4, B6, and B7, $E[\boldsymbol{\zeta}] = O\left(h_N^L\right)$, which shows the statement in (i). Since

$$\begin{aligned} E[\|\boldsymbol{\zeta}\|^l]^{1/l} \leq\ & E[\|\boldsymbol{\xi}_1\|^l]^{1/l} + \frac{1}{\pi_{h\backslash 0}} E[\|\boldsymbol{\xi}_2\|^l]^{1/l}\|\boldsymbol{\Pi}^{-1}\|\|\boldsymbol{\Psi}_{1:L}^{-1}\mathbf{h}\| \\ & + \frac{1}{\pi_h}\|\mathbf{Q}\|\|\mathbf{V}^{-1}\|E[\|\boldsymbol{\xi}_3\|^l]^{1/l}\|\boldsymbol{\Omega}^{-1}\mathbf{e}_1\| \end{aligned}$$

for $l = 2,4$, by Lemmas B7–10, we have $E[\|\boldsymbol{\zeta}\|^l] = O\left(\frac{\pi_h^{1-\epsilon}}{h_N^l}\right)$ for $l = 2,4$, which shows the statement in (ii). The statement in (iii) follows from Assumption 4(i) and the statement in (ii). □

LEMMA B12. *The minimum eigenvalue of* $\frac{h_N^2\pi_{h\backslash 0}}{\pi_h^2} Var(\boldsymbol{\zeta})$ *is bounded away from zero.*

Proof. Let $\mathbf{c}$ be any p-dimensional column vector with $\|\mathbf{c}\| = 1$. By Assumptions 3(ii) and 4(iii), there are positive constants c_1 and C such that $\mathbf{c}'Var\left(\mathbf{X}^*(\mathbf{Y} - \mathbf{W}\boldsymbol{\delta}) \mid D = u\right)\mathbf{c} \geq C$ for every $u \in [-c_1, c_1]$. Without loss of generality, assume $h_N \leq c_1$ for the rest of this proof. Note that $\boldsymbol{\zeta} = \boldsymbol{\xi}_1 + (\frac{1}{\pi_{h\backslash 0}}\boldsymbol{\xi}_2\boldsymbol{\Pi}^{-1}\boldsymbol{\Psi}_{1:L}^{-1}\mathbf{h} + \frac{1}{\pi_h}\mathbf{Q}\mathbf{V}^{-1}\boldsymbol{\xi}_3\boldsymbol{\Omega}^{-1}\mathbf{e}_1)$ and that $\boldsymbol{\xi}_1$ and $(\frac{1}{\pi_{h\backslash 0}}\boldsymbol{\xi}_2\boldsymbol{\Pi}^{-1}\boldsymbol{\Psi}_{1:L}^{-1}\mathbf{h} + \frac{1}{\pi_h}\mathbf{Q}\mathbf{V}^{-1}\boldsymbol{\xi}_3\boldsymbol{\Omega}^{-1}\mathbf{e}_1)$ are independent because they use disjoint sets of observations. Therefore,

$$Var(\mathbf{c}'\boldsymbol{\zeta}) \geq \max\left\{Var(\mathbf{c}'\boldsymbol{\xi}_1), Var\left(\mathbf{c}'\left(\frac{1}{\pi_{h\backslash 0}}\boldsymbol{\xi}_2\boldsymbol{\Pi}^{-1}\boldsymbol{\Psi}_{1:L}^{-1}\mathbf{h} + \frac{1}{\pi_h}\mathbf{Q}\mathbf{V}^{-1}\boldsymbol{\xi}_3\boldsymbol{\Omega}^{-1}\mathbf{e}_1\right)\right)\right\}.$$

We want to show that $\frac{h_N^2\pi_{h\backslash 0}}{\pi_h^2} Var(\mathbf{c}'\boldsymbol{\xi}_1)$ is bounded away from zero. We divide the proof into two cases: $P(D = 0) = 0$ and $P(D = 0) > 0$. Consider the case when $P(D = 0) = 0$. We are going to show that $\frac{h_N^2}{\pi_{h\backslash 0}} Var(\mathbf{c}'\boldsymbol{\xi}_1)$ is bounded away from zero. By the conditional variance identity,

$$Var(\mathbf{c}'\boldsymbol{\xi}_1) \geq E\left[Var\left(\mathbf{c}'\boldsymbol{\xi}_1 \mid D\right)\right].$$

Note that

$$\begin{aligned} E\left[Var\left(\mathbf{c}'\boldsymbol{\xi}_1 \mid D\right)\right] &= E\left[\mathbf{1}\{|D| > h_N\}D^{-2}\mathbf{c}'Var\left(\mathbf{X}^*(\mathbf{Y} - \mathbf{W}\boldsymbol{\delta}) \mid D\right)\mathbf{c}\right] \\ &\geq E\left[\mathbf{1}\{h_N < |D| \leq c_1\}D^{-2}\mathbf{c}'Var\left(\mathbf{X}^*(\mathbf{Y} - \mathbf{W}\boldsymbol{\delta}) \mid D\right)\mathbf{c}\right] \\ &\geq CE\left[\mathbf{1}\{h_N < |D| \leq c_1\}D^{-2}\right]. \end{aligned}$$

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

Lemma B3 provides the limit of $\frac{h_N^2}{\pi_{h\backslash 0}} E\left[\mathbf{1}\{h_N < |D| \le c_1\} D^{-2}\right]$, so

$$\frac{h_N^2}{\pi_{h\backslash 0}} Var(\mathbf{c}'\boldsymbol{\xi}_1) \ge C\left(\frac{\alpha_1}{2-\alpha_1}\frac{P(0<D\le h_N)}{\pi_{h\backslash 0}} + \frac{\alpha_2}{2-\alpha_2}\frac{P(-h_N<D<0)}{\pi_{h\backslash 0}}\right) + o(1),$$

which implies that $\frac{h_N^2}{\pi_{h\backslash 0}} Var(\mathbf{c}'\boldsymbol{\xi}_1)$ is bounded away from zero.

Now consider the case when $P(D=0) > 0$. Since $\boldsymbol{\xi}_1$ and $(\frac{1}{\pi_{h\backslash 0}}\boldsymbol{\xi}_2\boldsymbol{\Pi}^{-1}\boldsymbol{\Psi}_{1:L}^{-1}\mathbf{h} + \frac{1}{\pi_h}\mathbf{Q}\mathbf{V}^{-1}\boldsymbol{\xi}_3\boldsymbol{\Omega}^{-1}\mathbf{e}_1)$ are independent, we need to show that $h_N^2\pi_{h\backslash 0} Var\left(\mathbf{c}'\left(\frac{1}{\pi_{h\backslash 0}}\boldsymbol{\xi}_2\boldsymbol{\Pi}^{-1}\boldsymbol{\Psi}_{1:L}^{-1}\mathbf{h} + \frac{1}{\pi_h}\mathbf{Q}\mathbf{V}^{-1}\boldsymbol{\xi}_3\boldsymbol{\Omega}^{-1}\mathbf{e}_1\right)\right)$ is bounded away from zero. Note that, by Lemmas B7, B9, and B10,

$$\begin{aligned}
&\left\| h_N^2\pi_{h\backslash 0} Var\left(\mathbf{c}'\left(\frac{1}{\pi_{h\backslash 0}}\boldsymbol{\xi}_2\boldsymbol{\Pi}^{-1}\boldsymbol{\Psi}_{1:L}^{-1}\mathbf{h} + \frac{1}{\pi_h}\mathbf{Q}\mathbf{V}^{-1}\boldsymbol{\xi}_3\boldsymbol{\Omega}^{-1}\mathbf{e}_1\right)\right)\right. \\
&\quad \left. - \frac{h_N^2}{\pi_{h\backslash 0}} Var\left(\mathbf{c}'\mathbf{h}'\boldsymbol{\Psi}_{1:L}^{-1}\boldsymbol{\Pi}^{-1}\boldsymbol{\xi}_2\right)\right\| \\
&\le 2\frac{h_N^2}{\pi_h}\|\mathbf{Q}\| E\left[\|\boldsymbol{\xi}_2\|^2\right]^{1/2} E\left[\|\boldsymbol{\xi}_3\|^2\right]^{1/2} \|\boldsymbol{\Pi}^{-1}\|\|\boldsymbol{\Psi}_{1:L}^{-1}\mathbf{h}\|\|\mathbf{V}^{-1}\|\|\boldsymbol{\Omega}^{-1}\mathbf{e}_1\| \\
&\quad + \frac{h_N^2\pi_{h\backslash 0}}{\pi_h^2}\|\mathbf{Q}\|^2 E\left[\|\boldsymbol{\xi}_3\|^2\right]\|\mathbf{V}^{-1}\|^2\|\boldsymbol{\Omega}^{-1}\mathbf{e}_1\|^2 \\
&= O(\pi_{h\backslash 0}^{1/2}),
\end{aligned}$$

where we use the fact that $1/\pi_h = O(1)$ when $P(D=0) > 0$. Therefore, it suffices to show that $\frac{h_N^2}{\pi_{h\backslash 0}} Var\left(\mathbf{c}'\mathbf{h}'\boldsymbol{\Psi}_{1:L}^{-1}\boldsymbol{\Pi}^{-1}\boldsymbol{\xi}_2\right)$ is bounded away from zero. By the conditional variance identity,

$$Var\left(\mathbf{c}'\mathbf{h}'\boldsymbol{\Psi}_{1:L}^{-1}\boldsymbol{\Pi}^{-1}\boldsymbol{\xi}_2\right) \ge E[Var((\boldsymbol{\xi}_2\boldsymbol{\Pi}^{-1}\boldsymbol{\Psi}_{1:L}^{-1}\mathbf{h})'\mathbf{c} \mid D)].$$

Since

$$\begin{aligned}
Var((\boldsymbol{\xi}_2\boldsymbol{\Pi}^{-1}\boldsymbol{\Psi}_{1:L}^{-1}\mathbf{h})'\mathbf{c} \mid D) &= \left(\boldsymbol{\Psi}_{1:L}^{-1}\mathbf{h}\right)'\boldsymbol{\Pi}^{-1}\boldsymbol{\Psi}_{1:L}^{-1}\mathbf{D}_{1:L}\mathbf{c}' Var(\mathbf{X}^*(\mathbf{Y}-\mathbf{W}\boldsymbol{\delta}) \mid D) \\
&\quad \mathbf{c}\mathbf{D}_{1:L}'\boldsymbol{\Psi}_{1:L}^{-1}\boldsymbol{\Pi}^{-1}\boldsymbol{\Psi}_{1:L}^{-1}\mathbf{h} \\
&= (\mathbf{c}' Var(\mathbf{X}^*(\mathbf{Y}-\mathbf{W}\boldsymbol{\delta}) \mid D)\mathbf{c})\left(\boldsymbol{\Psi}_{1:L}^{-1}\mathbf{h}\right)'\boldsymbol{\Pi}^{-1}\boldsymbol{\Psi}_{1:L}^{-1} \\
&\quad \mathbf{D}_{1:L}\mathbf{D}_{1:L}'\boldsymbol{\Psi}_{1:L}^{-1}\boldsymbol{\Pi}^{-1}\boldsymbol{\Psi}_{1:L}^{-1}\mathbf{h} \\
&\ge C\left(\boldsymbol{\Psi}_{1:L}^{-1}\mathbf{h}\right)'\boldsymbol{\Pi}^{-1}\boldsymbol{\Psi}_{1:L}^{-1}\mathbf{D}_{1:L}\mathbf{D}_{1:L}'\boldsymbol{\Psi}_{1:L}^{-1}\boldsymbol{\Pi}^{-1}\boldsymbol{\Psi}_{1:L}^{-1}\mathbf{h},
\end{aligned}$$

we have

$$\begin{aligned}
Var\left(\mathbf{c}'\mathbf{h}'\boldsymbol{\Psi}_{1:L}^{-1}\boldsymbol{\Pi}^{-1}\boldsymbol{\xi}_2\right) &\ge E[Var((\boldsymbol{\xi}_2\boldsymbol{\Pi}^{-1}\boldsymbol{\Psi}_{1:L}^{-1}\mathbf{h})'\mathbf{c} \mid D)] \\
&\ge C\left(\boldsymbol{\Psi}_{1:L}^{-1}\mathbf{h}\right)'\boldsymbol{\Pi}^{-1}E\left[\boldsymbol{\Psi}_{1:L}^{-1}\mathbf{D}_{1:L}\mathbf{D}_{1:L}'\boldsymbol{\Psi}_{1:L}^{-1}\right]\boldsymbol{\Pi}^{-1}\boldsymbol{\Psi}_{1:L}^{-1}\mathbf{h}
\end{aligned}$$

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

$$=\pi_{h\backslash 0} C\left(\boldsymbol{\Psi}_{1:L}^{-1}\mathbf{h}\right)' \boldsymbol{\Pi}^{-1}\boldsymbol{\Psi}_{1:L}^{-1}\mathbf{h}.$$

By Lemmas B1 and B4, $\frac{h_N^2}{\pi_{h\backslash 0}} Var\left(\mathbf{c}'\mathbf{h}'\boldsymbol{\Psi}_{1:L}^{-1}\boldsymbol{\Pi}^{-1}\boldsymbol{\xi}_2\right)$ is bounded away from zero. □

LEMMA B13. $E_N[\|\hat{\boldsymbol{\zeta}}-\boldsymbol{\zeta}\|^2]^{1/2} = o(1) + O_p\left(\sqrt{\frac{\pi_h^{2(1-\epsilon)}}{Nh_N^2\pi_{h\backslash 0}^2}}\right)$ *for every positive number* $\epsilon \in (0,1)$.

Proof. By Lemmas B1–B7 and B9–B11, we have

$$\begin{aligned}
\|\hat{\boldsymbol{\delta}}-\boldsymbol{\delta}\| \leq& \frac{1}{\pi_h}\|\hat{\mathbf{V}}^{-1}\|\|(E_N-E)\left[\boldsymbol{\xi}_3\right]\|\|\hat{\boldsymbol{\Omega}}^{-1}\mathbf{e}_1\| \\
&+\frac{1}{\pi_h}\|\hat{\mathbf{V}}^{-1}\|\|E\left[\boldsymbol{\xi}_3\right]\|\|(\hat{\boldsymbol{\Omega}}^{-1}-\boldsymbol{\Omega}^{-1})\mathbf{e}_1\| \\
&+\frac{1}{\pi_h}\|\hat{\mathbf{V}}^{-1}\|\|E\left[\boldsymbol{\xi}_3\right]\boldsymbol{\Omega}^{-1}\mathbf{e}_1\| \\
=& O\left(h_N^{L+1}\right)+O_p\left(\sqrt{\frac{1}{N\pi_h}}\right), \\
\|(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})\boldsymbol{\Psi}_{1:L}\| \leq& \frac{1}{\pi_{h\backslash 0}}\|E\left[\boldsymbol{\xi}_2\right]\|\|\hat{\boldsymbol{\Pi}}^{-1}\|+\frac{1}{\pi_{h\backslash 0}}\|(E_N-E)\left[\boldsymbol{\xi}_2\right]\|\|\hat{\boldsymbol{\Pi}}^{-1}\| \\
&+\frac{1}{\pi_{h\backslash 0}}E_N\left[\|(\mathbf{X}^*\mathbf{W}\|\|\mathbf{D}'_{1:L}\boldsymbol{\Psi}_{1:L}^{-1}\|\right]\|\hat{\boldsymbol{\delta}}-\boldsymbol{\delta}\|\|\hat{\boldsymbol{\Pi}}^{-1}\| \\
=& O\left(h_N^{L+1}\right)+O_p\left(\sqrt{\frac{1}{N\pi_{h\backslash 0}}}\right), \\
E_N[\|\hat{\boldsymbol{\xi}}_1-\boldsymbol{\xi}_1\|^2]^{1/2} \leq& E_N[\mathbf{1}\{|D|>h_N\}D^{-2}\|\mathbf{X}^*\mathbf{W}\|^2]^{1/2}\|\hat{\boldsymbol{\delta}}-\boldsymbol{\delta}\|+\|(E_N-E)[\boldsymbol{\xi}_1]\| \\
=& O\left(\sqrt{h_N^{2L}\pi_{h\backslash 0}}\right)+O_p\left(\sqrt{\frac{\pi_{h\backslash 0}}{Nh_N^2\pi_h}}\right), \\
E_N[\|\hat{\boldsymbol{\xi}}_2-\boldsymbol{\xi}_2\|^2]^{1/2} \leq& E_N[\|\mathbf{X}^*\mathbf{W}\|^2\|\mathbf{D}'_{1:L}\boldsymbol{\Psi}_{1:L}^{-1}\|^2]^{1/2}\|\hat{\boldsymbol{\delta}}-\boldsymbol{\delta}\| \\
&+\|(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})\boldsymbol{\Psi}_{1:L}\|E_N[\|\mathbf{D}'_{1:L}\boldsymbol{\Psi}_{1:L}^{-1}\|^4]^{1/2} \\
=& O\left(\sqrt{h_N^{2(L+1)}\pi_{h\backslash 0}}\right)+O_p\left(\sqrt{\frac{1}{N}}\right), \\
E_N[\|\hat{\boldsymbol{\xi}}_3-\boldsymbol{\xi}_3\|^2]^{1/2} \leq& E_N[\|(\mathbf{X}^*\mathbf{W})'\mathbf{X}^*\mathbf{W}\|^2\|\mathbf{D}'_{0:L}\boldsymbol{\Psi}_{0:L}^{-1}\|^2]^{1/2}\|\hat{\boldsymbol{\delta}}-\boldsymbol{\delta}\| \\
=& O\left(\sqrt{h_N^{2(L+1)}\pi_h}\right)+O_p\left(\sqrt{\frac{1}{N}}\right), \text{ and} \\
E_N[\|\hat{\boldsymbol{\zeta}}-\boldsymbol{\zeta}\|^2]^{1/2} =& E_N[\|\hat{\boldsymbol{\xi}}_1-\boldsymbol{\xi}_1\|^2]^{1/2}+\frac{1}{\pi_{h\backslash 0}}E_N[\|\hat{\boldsymbol{\xi}}_2-\boldsymbol{\xi}_2\|^2]^{1/2}\|\hat{\boldsymbol{\Pi}}^{-1}\|\|\boldsymbol{\Psi}_{1:L}^{-1}\hat{\mathbf{h}}\| \\
&+\frac{1}{\pi_h}\|\hat{\mathbf{Q}}\|\|\hat{\mathbf{V}}^{-1}\|E_N[\|\hat{\boldsymbol{\xi}}_3-\boldsymbol{\xi}_3\|^2]^{1/2}\|\hat{\boldsymbol{\Omega}}^{-1}\mathbf{e}_1\|
\end{aligned}$$

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

$$
\begin{aligned}
&+\frac{1}{\pi_{h\backslash 0}} E_N[\|\boldsymbol{\xi}_2\|^2]^{1/2} \\
&\left(\left\|\hat{\boldsymbol{\Pi}}^{-1}-\boldsymbol{\Pi}^{-1}\right\| \|\boldsymbol{\Psi}_{1:L}^{-1}\hat{\mathbf{h}}\| + \|\boldsymbol{\Pi}^{-1}\| \|\boldsymbol{\Psi}_{1:L}^{-1}\left(\hat{\mathbf{h}}-\mathbf{h}\right)\|\right) \\
&+\frac{1}{\pi_h}\|\hat{\mathbf{Q}}\|\left(\|\hat{\mathbf{V}}^{-1}\|\|(\hat{\boldsymbol{\Omega}}^{-1}-\boldsymbol{\Omega}^{-1})\mathbf{e}_1\| + \left\|\hat{\mathbf{V}}^{-1}-\mathbf{V}^{-1}\right\| \|\boldsymbol{\Omega}^{-1}\mathbf{e}_1\|\right) \\
&E_N[\|\boldsymbol{\xi}_3\|^2]^{1/2} \\
&+\frac{1}{\pi_h}\left\|\hat{\mathbf{Q}}-\mathbf{Q}\right\| \|\mathbf{V}^{-1}\|\|\boldsymbol{\Omega}^{-1}\mathbf{e}_1\| E_N[\|\boldsymbol{\xi}_3\|^2]^{1/2} \\
&= o(1) + O_p\left(\sqrt{\frac{\pi_h^{2(1-\epsilon)}}{Nh_N^2\pi_{h\backslash 0}^2}}\right).
\end{aligned}
$$

Therefore, the statement of the lemma follows. □

SUPPLEMENTARY MATERIAL

Sasaki, Y. and T. Ura (July 12, 2025): Supplement to "Slow Movers in Panel Data," *Econometric Theory* Supplementary Material. To view, please visit: https://doi.org/10.1017/S0266466625100157.

REFERENCES

Aguiar, M. A., Bils, M., & Boar, C. (2020). Who are the hand-to-mouth? National Bureau of Economic Research Working Paper 26643.

Altonji, J. G., & Matzkin, R. L. (2005). Cross section and panel data estimators for nonseparable models with endogenous regressors. *Econometrica*, 73, 1053–1102.

Arellano, M., & Bonhomme, S. (2012). Identifying distributional characteristics in random coefficients panel data models. *The Review of Economic Studies*, 79, 987–1020.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82, 2295–2326.

Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of Econometrics*, 18, 5–46.

Chamberlain, G. (1992). Efficiency bounds for semiparametric regression. *Econometrica*, 60, 567–596.

Chaudhuri, S., & Hill, J. B. (2016). Heavy tail robust estimation and inference for average treatment effects. Working paper.

Chernozhukov, V., Fernández-Val, I., Hahn, J., & Newey, W. (2013). Average and quantile effects in nonseparable panel models. *Econometrica*, 81, 535–580.

Chernozhukov, V., Fernández-Val, I., Hoderlein, S., Holzmann, H., & Newey, W. (2015). Nonparametric identification in panels using quantiles. *Journal of Econometrics*, 188, 378–392.

Cooper, D., Dynan, K. E., & Rhodenhiser, H. (2019). Measuring household wealth in the panel study of income dynamics: The role of retirement assets. FRB of Boston Working Paper No. 19-6.

Fernández-Val, I., & Lee, J. (2013). Panel data models with nonadditive unobserved heterogeneity: Estimation and inference. *Quantitative Economics*, 4, 453–481.

Fisher, J. D., Johnson, D. S., Smeeding, T. M., & Thompson, J. P. (2020). Estimating the marginal propensity to consume using the distributions of income, consumption, and wealth. *Journal of Macroeconomics*, 65, 103218.

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press

Graham, B. S., Hahn, J., Poirier, A., & Powell, J. L. (2018). A quantile correlated random coefficients panel data model. *Journal of Econometrics*, 206, 305–335.

Graham, B. S., & Powell, J. L. (2012). Identification and estimation of average partial effects in "irregular" correlated random coefficient panel data models. *Econometrica*, 80, 2105–2152.

Hoderlein, S., & White, H. (2012). Nonparametric identification in nonseparable panel data models with generalized fixed effects. *Journal of Econometrics*, 168, 300–314.

Holland, S., Mansur, E., Verdier, V., & Yates, A. (2023). Estimating marginal emissions on electrical grids: A naturally regularized panel data regression. Working Paper.

Hsiao, C. (2014). *Analysis of panel data, econometric society monographs*. (3rd ed.) Cambridge University Press.

Hsiao, C., & Pesaran, M. H. (2008). Random coefficient models. In L. Mátyás, & P. Sevestre (Eds.), *The econometrics of panel data: Fundamentals and recent developments in theory and practice*. Springer.

Hsiao, C., Pesaran, M. H., & Tahmiscioglu, A. K. (1999). Bayes estimation of short-run coefficients in dynamic panel data models. In C. Hsiao, M. H. Pesaran, K. Lahiri, & L. F. Lee (Eds.), *Analysis of panels and limited dependent variable models* (pp. 268–296). Cambridge University Press.

Laage, L. (2024). A correlated random coefficient panel model with time-varying endogeneity. *Journal of Econometrics*, 242(2), 105804.

Murtazashvili, I., & Wooldridge, J. M. (2008). Fixed effects instrumental variables estimation in correlated random coefficient panel data models. *Journal of Econometrics*, 142, 539–552.

Pesaran, M. H., & Smith, R. (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics*, 68, 79–113.

Resnick, S. I. (2007). *Heavy-tail phenomena: Probabilistic and statistical modeling*. Springer Science & Business Media.

Robinson, P. M. (1988). Root-N-consistent semiparametric regression. *Econometrica*, 56, 931–954.

Sasaki, Y., & Ura, T. (2022). Estimation and inference for moments of ratios with robustness against large trimming bias. *Econometric Theory*, 38, 66–112.

Swamy, P. A. (1970). Efficient inference in a random coefficient regression model. *Econometrica*, 38, 311–323.

Verdier, V. (2020). Average treatment effects for stayers with correlated random coefficient models of panel data. *Journal of Applied Econometrics*, 35, 917–939.

Wooldridge, J. M. (2005). Fixed-effects and related estimators for correlated random-coefficient and treatment-effect panel data models. *Review of Economics and Statistics*, 87, 385–390.

https://doi.org/10.1017/S0266466625100157 Published online by Cambridge University Press