*Econometric Theory*, **0**, 2024, 1–62.
doi:10.1017/S0266466623000403

# INFERENCE IN PARTIALLY IDENTIFIED PANEL DATA MODELS WITH INTERACTIVE FIXED EFFECTS

SHENGJIE HONG
*Renmin University of China*

LIANGJUN SU
*Tsinghua University*

YAQI WANG
*Central University of Finance and Economics*

In this paper, we develop methods for statistical inferences in a partially identified nonparametric panel data model with endogeneity and interactive fixed effects. Under some normalization rules, we can concentrate out the large-dimensional parameter vector of factor loadings and specify a set of conditional moment restrictions that are involved with only the finite-dimensional factor parameters along with the infinite-dimensional nonparametric component. For a conjectured restriction on the parameter, we consider testing the null hypothesis that the restriction is satisfied by at least one element in the identified set and propose a test statistic based on a novel martingale difference divergence measure for the distance between a conditional expectation object and zero. We derive a tight asymptotic distributional upper bound for the resultant test statistic under the null and show that it is divergent at rate-$N$ under the global alternative. To obtain the critical values for our test, we propose a version of multiplier bootstrap and establish its asymptotic validity. Simulations demonstrate the finite sample properties of our inference procedure. We apply our method to study Engel curves for major nondurable expenditures in China by using a panel dataset from the China Family Panel Studies.

## 1. INTRODUCTION

Recently there has been growing interest in panel data models with interactive fixed effects (IFEs). Under a linear specification of the regression relationship, these models have been extensively studied in the literature (see Coakley, Fuertes, and

The authors thank Iván Fernández-Val and two anonymous referees for their constructive comments. They also thank Xiaohong Chen, Jack Porter, Andres Santos, and Yu Zhu for very helpful comments and suggestions. Hong, Su, and Wang thank the National Natural Science Foundation of China (NSFC) for financial support under the Grant numbers 72373175, 72133002, and 72273164, respectively. Address correspondence to Liangjun Su, School of Economics and Management, Tsinghua University, Beijing, China; e-mail: sulj@sem.tsinghua.edu.cn. Address correspondence to Liangjun Su, Tsinghua University; e-mail: sulj@sem.tsinghua.edu.cn

 

Smith, 2002; Phillips and Sul, 2003, 2007; Pesaran, 2006; Kapetanios and Pesaran, 2007; Pesaran and Tosetti, 2007; Greenaway-McGrevy, Han, and Sul, 2008; Bai (2009); Moon and Weidner, 2015, 2017; Lu and Su, 2016, among others). More recently, in an effort to relax the linear specification, much attention has been turned to the study of nonparametric panel data models with interactive effects (see, e.g., Su and Jin, 2012; Su, Jin, and Zhang, 2015; Freyberger, 2018; Su and Zhang, 2018; Dong, Gao, and Peng, 2020 for an overview). In particular, Freyberger (2018) studies a very general nonparametric and nonseparable panel model with IFEs. Nevertheless, all of these papers restrict the covariates to be either strictly or weakly exogenous and assume that the model parameters are point-identified.

In this paper, we consider the following nonparametric panel data regression model:

$$y_{it} = g^0(x_{it}) + \lambda_i^{0\prime} F_t^0 + u_{it}, \tag{1.1}$$

where $i = 1,\ldots,N,\ t = 1,\ldots,T,\ x_{it}$ is a $d_x \times 1$ vector of general regressors with support $\mathcal{X}$, $y_{it}$ is a scalar output variable with support $\mathcal{Y}$, $F_t^0$ and $\lambda_i^0$ are $R \times 1$ vectors of unobserved factors and factor loadings, respectively, $u_{it}$ is a zero mean error term, and the functional form of $g^0(\cdot)$ is unknown. We allow $x_{it}$ and $u_{it}$ to be correlated, and are interested in the inference of $g^0(\cdot)$ by assuming the presence of a $d_z \times 1$ vector of weakly exogenous instruments $z_{it}$ with support $\mathcal{Z}$, such that

$$\mathbb{E}(u_{it}|z_{i1},\ldots,z_{it}) = 0 \text{ almost surely (a.s.).} \tag{1.2}$$

Throughout the paper, we assume that $T$ and $R$ are fixed with $T \geq R+1,$ and the asymptotic theory is established by passing $N$ to infinity.

When $g^0(x_{it})$ is linear in $x_{it}$ so that $g^0(x_{it}) = \beta^{0\prime} x_{it}$ for some $\beta^0 \in \mathbb{R}^{d_x}$, $x_{it}$ is strictly exogenous, and $R = 1$, Ahn, Lee, and Schmidt (2001) follow the lead of Holz-Eakin, Newey, and Rosen (1988) to study the asymptotic properties of the GMM estimator of $\beta^0$ based on the quasi-differencing of the equation in (1.1). Ahn, Lee, and Schmidt (2013, ALS hereafter) consider the GMM estimation of $\beta$ when $x_{it}$ is either strictly or weakly exogenous and $R \geq 1$. Su and Jin (2012) study the asymptotic properties of the sieve estimator of $g^0$ in (1.1) when $x_{it}$ is strictly exogenous. Su and Zhang (2018) consider the sieve estimation of $g^0$ when $x_{it}$ is weakly exogenous. Freyberger (2018) considers the point-identification and estimation of a model that is more general than that in (1.1), but still restricts $x_{it}$ to be either strictly or weakly exogenous.

The nonparametric IV (NPIV) model for cross-sectional data (i.e., $y_i = g^0(x_i) + u_i$ with $\mathbb{E}(u_i|z_i) = 0$ a.s.) has been widely studied in the literature. NPIV is encompassed by our model as a special case where $T = 1$ and $R = 0$. Point-identification of $g^0(\cdot)$ in NPIV relies heavily on a completeness assumption regarding the joint distribution of $x_i$ and $z_i$, as formalized by Newey and Powell (2003). Nevertheless, Santos (2012) shows that $g^0(\cdot)$ in the NPIV model is only partially identified in general. The partial identification nature still exists in Model (1.1) without imposing strong ad hoc assumptions when we treat the IFEs as fixed parameters. Therefore, we aim at developing an effective inference method for $g^0(\cdot)$ in the potential absence of point-identification in this paper.

For a conjectured restriction on $g(\cdot)$, we develop a consistent procedure for testing the hypothesis that the restriction is satisfied by at least one element in the identified set of $g(\cdot)$.[1] A broad group of restrictions can be tested in this way, making the procedure applicable to various inference tasks, including testing model specification and constructing a confidence set for $g^0(\cdot)$ at any given point. We derive a tight asymptotic distributional upper bound for our test statistic under the null and show that it is divergent at rate-$N$ under the global alternative based on the $U$-process theory. To make asymptotically valid inferences, we propose a version of multiplier bootstrap and justify its use to obtain conservative bootstrap critical values.We conduct Monte Carlo simulations to demonstrate the finite sample properties of our inference procedure.

Our test statistic is based on a novel martingale difference divergence (MDD) measure for the distance between a conditional expectation object and zero. This way of constructing a statistic for testing conditional moment specification is rather different from the widely adopted method, dating back to Bierens (1982), that involves transforming conditional moments into infinitely many unconditional ones via a family of instrument functions and then constructing Kolmogorov Smirnov (KS) or Cramér-von Mises (CvM) type statistic over the instrument function family. As shown in the paper, under partial identification, our MDD-based statistic has one main advantage over Bierens-type statistics: Computing our MDD-based statistic is relatively simple regardless of the dimension of the conditioning variables (i.e., $\underline{z}_{it} \equiv (z'_{i1}, \ldots, z'_{it})'$) so that it does not suffer from high computational cost even if the dimension of the conditioning variables is moderately large.

The main technical challenges of our analysis arise largely because our test statistic is associated with a second-order $U$-process that is asymptotically degenerate under the null and nondegenerate under the alternative. First, our test statistic can be written as a minimizer of an MDD-based process indexed by $\theta = (\phi', g)'$, where $\phi$ is a finite-dimensional vector associated with the unobserved factor $F = (F_1, \ldots, F_T)'$ and $g$ is the infinite-dimensional parameter of interest. The MDD-based process is a third-order $U$-process that can be decomposed into the summation of a bias term, a second-order canonical $U$-process and a third-order canonical $U$ -process via standard Hoeffding decompositions. To study these canonical $U$ -process components, we find helpful insights from de la Peña and Giné (1999) who state some weak convergence results for canonical $U$-processes with kernel functions belonging to the VC-subgraph class. Such results are not directly applicable to our setting because the kernel functions of our $U$-processes do not belong to the VC-subgraph class due to the presence of the

[1] As shown in the paper, the proposed procedure actually allows for testing restrictions on the joint parameter $\theta = (\phi', g(\cdot))'$, where $\phi$ is a $(T-R)R$-dimensional vector of unrestricted parameters from $F = (F_1, F_2, \ldots, F_T)'$ to be formally introduced in Section 2.1.

infinite-dimensional parameter $g(\cdot)$. Fortunately, we can verify some primitive conditions in Arcones and Giné (1993) to show that the second-order canonical $U$-process in our Hoeffding decomposition converges weakly to a Gaussian chaos process and the third-order term asymptotically vanishes. To the best of our knowledge, such results are the first ones for degenerate $U$-processes indexed by a non-VC-subgraph class, which complements the literature in both econometrics and statistics.

Second, to derive the asymptotic null distribution for our statistic, we borrow ideas from the growing literature on nonparametric partial identification (see Santos, 2012; Andrews and Shi, 2014; Hong, 2017; Chernozhukov, Newey, and Santos, 2023, among others). Our study complements this literature and is closely related to Santos (2012) and Hong (2017). The main difference is that Santos (2012) and Hong (2017) establish their limiting distributions based on standard empirical process theory while we establish our asymptotic results based on the $U$-process theory. As a first step, we manage to show that any minimizer $\hat{\theta}_N$ of our MDD-based process lies in the $o_p(N^{-1/4})$-neighborhood of the identified set under the $L^2$-norm. Then we show that our test statistic converges to the minimum of a well-defined (noncentered) Gaussian chaos process except for a usual drift term that also appears in Santos (2012) and Hong (2017). Santos (2012) employs an additional sieve approximation to mimic the corresponding drifting term in his bootstrap procedure so that the resulting test is exact. In comparison, Hong (2017) sets the corresponding drifting term to zero in his bootstrap procedure, which saves computational cost, but leads to potentially conservative tests. The characterization of the drifting term is extremely hard in our setup. So we follow the lead of Hong (2017) when implementing a multiplier bootstrap procedure, which makes our test potentially conservative too. To show the asymptotic validity of the bootstrap procedure, an essential step is the study of the unconditional central limit theorem (CLT) for the underlying $U$-process of our bootstrap statistic, which is analogous to the unconditional multiplier CLT for empirical processes studied in van der Vaart and Wellner (1996) and Kororok (2008). It extends the unconditional multiplier CLT for degenerate second-order $U$-statistics in Leucht and Neumann (2013) to degenerate second-order $U$-processes.

As an empirical illustration, we apply our method to study Engel curves for four major nondurable expenditures in China by using a panel dataset from the China Family Panel Studies (CFPS). One of our interesting findings is that, even with a nonparametric specification on $g(\cdot)$, the model is not sufficient to adequately describe the Engel curve for food consumption among urban households in China when setting the number $R$ of factors to be 0 or 1. Our test fails to reject the log-linear specification for the Engel curves among rural households when setting $R = 1$. Our empirical study suggests a difference in the degree of heterogeneity in consumption patterns between the urban population and rural population in China. It also suggests that, even a nonparametric specification on $g(\cdot)$, as general as it is, might still be insufficient to compensate for an inadequate handling of heterogeneity to make the corresponding Engel curve a correctly specified one.

These results provide some new insights to the huge literature on empirical studies of Engel curves.

The rest of the paper is organized as follows: In Section 2, we introduce the model, the moment conditions, and the hypotheses. In Section 3, we construct the MDD-based test statistic, derive its asymptotic behavior, and propose a consistent multiplier bootstrap procedure to obtain the $p$ -values. In Section 4, we study the finite sample performance of our inference procedure by Monte Carlo simulations. In Section 5, we apply our method to study Chinese households' Engel curves. Final remarks are contained in Section 6. The proofs of all theorems and lemmas are delegated to the Appendix. Additional materials are provided in the Appendixes B and C of the Supplementary Material.

NOTATION. For a vector or matrix $A$, we denote its transpose as $A'$ and its Frobenius norm as $|A|$ $(\equiv [\mathrm{tr}\left(AA'\right)]^{1/2})$, where $\equiv$ signifies a definitional relationship. We use $\|\cdot\|$ to denote generic (pseudo) norm. For example, for $\theta = \left(\phi', g\right)'$, where $\phi$ is a finite-dimensional vector to be specified later on and $g$ is the infinite-dimensional parameter, we define $\|\theta\| \equiv |\phi| + \|g\|$ to denote a generic (pseudo) norm for $\theta = \left(\phi', g\right)'$, and one popular choice for $\|\cdot\|$ is the $L^2$ norm, yielding $\|\theta\|_{L^2} = |\phi| + \|g\|_{L^2}$. The true value of $\theta = \left(\phi', g\right)'$ is denoted as $\theta^0 = \left(\phi^{0\prime}, g^0\right)'$. The operator $\xrightarrow{p}$, $\xrightarrow{\mathcal{L}}$, and $\Longrightarrow$ denote convergence in probability, convergence in law and weak convergence in the sense of Chapter 1.3 in van der Vaart and Wellner (1996), respectively.

## 2. THE MODEL AND HYPOTHESES

Let $X_i = (x_{i1}, \ldots, x_{iT})'$, $Y_i = (y_{i1}, \ldots, y_{iT})'$, $Z_i = (z_{i1}, \ldots, z_{iT})'$, $\mathbf{g}^0(X_i) = \left(g^0(x_{i1}), \ldots, g^0(x_{iT})\right)'$, $F^0 = \left(F_1^0, \ldots, F_T^0\right)'$, and $U_i = (u_{i1}, \ldots, u_{iT})'$. We can rewrite Model (1.1) with Condition (1.2) in vector form:

$$Y_i = \mathbf{g}^0(X_i) + F^0\lambda_i^0 + U_i \text{ with } \mathbb{E}\left(u_{it}|\underline{z}_{it}\right) = 0 \text{ a.s.}, \tag{2.1}$$

where $\underline{z}_{it} \equiv (z_{i1}', \ldots, z_{it}')'$.

### 2.1. The Moment Condition

To proceed, we show that Model (2.1) is equivalent to a number of conditional moment equations. Note that $F^0\lambda_i^0 = F^0D^{-1}D\lambda_i^0 = F^*\lambda_i^*$ for any nonsingular matrix $D$, where $F^* = F^0D^{-1}$ and $\lambda_i^* = D\lambda_i^0$. So as a first step, to rule out such trivial non-identification, we make the normalization assumption that the $T \times R$ matrix $F$ takes a form similar to ALS and Freyberger (2018), as follows:

$$F = \begin{pmatrix} \Phi \\ -I_R \end{pmatrix}, \tag{2.2}$$

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

where $\Phi$ is a $(T-R)\times R$ matrix of unrestricted parameters, and (2.2) imposes $R^2$ restrictions by requiring the last $R$ rows of $F$ to be $-I_R$. Let $\phi = \text{vec}\left(\Phi'\right) \equiv (\phi_1', \ldots, \phi_{T-R}')'$, where $\phi_t$ denotes the $t$th column of $\Phi'$ for $t = 1, \ldots, T-R$.

Then we define the $T\times(T-R)$ matrix:

$$H(\phi) \equiv \begin{pmatrix} I_{T-R} \\ \Phi' \end{pmatrix} \equiv \left[H_1(\phi_1), \ldots, H_{T-R}(\phi_{T-R})\right]. \tag{2.3}$$

Note that

$$H\left(\phi^0\right)' F^0 = \left(I_{T-R}, \Phi^0\right)\begin{pmatrix} \Phi^0 \\ -I_R \end{pmatrix} = \mathbf{0}_{(T-R)\times R}, \tag{2.4}$$

where $\phi^0 = \text{vec}\left(\Phi^{0\prime}\right) = (\phi_1^{0\prime}, \ldots, \phi_{T-R}^{0\prime})'$ denotes the true value of $\phi$. Consequently, premultiplying both sides of (2.1) by $H\left(\phi^0\right)'$ helps to eliminate the incidental parameters $\left\{\lambda_i^0\right\}$ from the equation:

$$H\left(\phi^0\right)' Y_i = H\left(\phi^0\right)' \mathbf{g}^0(X_i) + H\left(\phi^0\right)' U_i, \tag{2.5}$$

where

$$H(\phi)' U_i = \begin{pmatrix} H_1(\phi_1)' U_i \\ H_2(\phi_2)' U_i \\ \vdots \\ H_{T-R}(\phi_{T-R})' U_i \end{pmatrix} = \begin{pmatrix} u_{i1} + \phi_1' \dot{U}_i \\ u_{i2} + \phi_2' \dot{U}_i \\ \vdots \\ u_{i,T-R} + \phi_{T-R}' \dot{U}_i \end{pmatrix}$$

and $\dot{U}_i = \left(u_{i,T-R+1}, u_{i,T-R+2}, \ldots, u_{i,T}\right)'$. Let

$$\begin{aligned} \mathbf{m}(Y_i, \phi, \mathbf{g}(X_i)) \equiv\ & H(\phi)' [Y_i - \mathbf{g}(X_i)] \\ =\ & \begin{pmatrix} H_1(\phi_1)' [Y_i - \mathbf{g}(X_i)] \\ H_2(\phi_2)' [Y_i - \mathbf{g}(X_i)] \\ \vdots \\ H_{T-R}\left(\phi_{T-R}\right)' [Y_i - \mathbf{g}(X_i)] \end{pmatrix} \equiv \begin{pmatrix} m_1(Y_i, \phi_1, \mathbf{g}(X_i)) \\ m_2(Y_i, \phi_2, \mathbf{g}(X_i)) \\ \vdots \\ m_{T-R}(Y_i, \phi_{T-R}, \mathbf{g}(X_i)) \end{pmatrix}. \end{aligned} \tag{2.6}$$

Then under the condition that the instrument $z_{it}$ is weakly exogenous, we can easily see that

$$\mathbb{E}\left[m_s(Y_i, \phi_s^0, \mathbf{g}^0(X_i)) | \underline{z}_{is}\right] = 0 \text{ a.s. for } s = 1, \ldots, T-R. \tag{2.7}$$

When $\phi^0$ and $g^0$ are point-identified, various methods have been proposed to study the estimation of $\phi$ and $g$ in the above model (see Ai and Chen, 2003; Chen and Pouzo, 2012, among others).

## 2.2. The Parameter Space $\Theta$

The parameter space $\Theta$ for $\theta = (\phi', g)'$ is specified as $\Theta = \mathbf{\Phi} \times \mathcal{G}$ as in Hong (2017), where $\mathbf{\Phi}$ is a compact subset of $\mathbb{R}^{(T-R)R}$ and $\mathcal{G}$ is a bounded subset of the following

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

Sobolev space:

$$\mathcal{W}^s(\mathcal{X}) \equiv \left\{ g : \mathcal{X} \to \mathbb{R} \mid \ g \text{ is } d\text{-times differentiable and } \|g\|_s \leq \infty \right\}$$

with $\|\cdot\|_s$ being a commonly used norm for weighted Sobolev spaces, defined as

$$\|g\|_s^2 \equiv \sum_{\langle\lambda\rangle \leq d} \int_{\mathcal{X}} \left| D^\lambda g(x) \right|^2 \left(1 + x'x\right)^{\zeta_0} dx,$$

where $\lambda \in \mathbb{N}_+^{d_x}$, $\langle\lambda\rangle \equiv \sum_{j=1}^{d_x} \lambda_j$, $D^\lambda g(x) \equiv \partial^{\langle\lambda\rangle} g(x) / \prod_{j=1}^{d_x} \partial x_j^{\lambda_j}$, $d \in \mathbb{N}_+$ measures the degree of smoothness, and $\zeta_0 \geq 0$. Define another norm $\|\cdot\|_c$ as follows:

$$\|g\|_c \equiv \max_{\langle\lambda\rangle \leq \frac{d}{2}} \left[ \sup_{x \in \mathcal{X}} \left| D^\lambda g(x) \right| \left(1 + x'x\right)^{\zeta/2} \right]$$

with $\zeta = 0$ for bounded $\mathcal{X}$, and $\left(\frac{d_x}{2} \lfloor \frac{d}{2} \rfloor\right) / \left(\lfloor \frac{d}{2} \rfloor - \frac{d_x}{2}\right) < \zeta < \zeta_0$ for unbounded $\mathcal{X}$. Here, $\lfloor a \rfloor$ represents the largest integer that is not larger than $a$.

To be precise, we specify the parameter space $\Theta = \boldsymbol{\Phi} \times \mathcal{G}$ as follows.

**Assumption 2.1.** (i) $\boldsymbol{\Phi} \subset \mathbb{R}^{(T-R)R}$ is compact; (ii) $\mathcal{G} = \{g \in \mathcal{W}^s(\mathcal{X}) : \|g\|_s \leq C_g\}$ for some $C_g < \infty$ and $d \geq d_x + 2$. $\zeta_0 = 0$ for bounded $\mathcal{X}$, and $\zeta_0 > \left(\frac{d_x}{2} \cdot \lfloor \frac{d}{2} \rfloor\right) / \left(\lfloor \frac{d}{2} \rfloor - \frac{d_x}{2}\right)$ for unbounded $\mathcal{X}$; (iii) $\nu < 1$, with $\nu$ defined as follows:

$$\nu \equiv \begin{cases} (d - \lfloor d/2 \rfloor + \zeta)\, d_x / \{\zeta (d - \lfloor d/2 \rfloor)\}, & \text{if } \mathcal{X} \text{ is unbounded,} \\ d_x / (d - \lfloor d/2 \rfloor), & \text{if } \mathcal{X} \text{ is bounded;} \end{cases} \qquad \textbf{(2.8)}$$

(iv) $\mathcal{X}$ satisfies a uniform cone condition.

Assumption 2.1(i) is standard and Assumption 2.1(ii)-(iii) parallels Assumption 2.1(i) in Santos (2012). Assumption 2.1(ii) specifies $\mathcal{G}$ to be a bounded ball under $\|\cdot\|_s$ with radius $C_g$ in $\mathcal{W}^s(\mathcal{X})$. Such a specification enjoys two benefits. First, as Santos (2012) notes, $\mathcal{G}$ is compact under the norm $\|\cdot\|_c$. Consequently, $\Theta = \boldsymbol{\Phi} \times \mathcal{G}$ is compact under $\|\cdot\|_c$ defined on $\mathbb{R}^{(T-R)R} \times \mathcal{W}^s(\mathcal{X})$ as

$$\|\theta\|_c \equiv |\phi| + \|g\|_c \qquad \textbf{(2.9)}$$

for $\theta = \left(\phi', g\right)'$. See Lemma A.3 in the Appendix. Second, $\mathcal{G}$ is also compact under $\|\cdot\|_{L^2}$, which can be verified by following the arguments of Freyberger and Masten (2019). As Hong (2017) remarks, the compactness of $\mathcal{G}$ under $\|\cdot\|_{L^2}$ makes the results in Schumaker (2007) applicable to developing primitive conditions for the required uniform rate (over $\Theta$) of sieve approximation errors (to be specified by Assumption 3.3 later in the paper). Assumption 2.1(iv) is the same as Assumption 2.1(ii) in Santos (2012) and Assumption 3.2(ii) in Hong (2017). It imposes a weak regularity condition on the shape of $\mathcal{X}$. As pointed out by Santos (2012), heuristically, Assumption 2.1(iv) is satisfied if there exists some small finite cone whose vertex can be placed in each point in the boundary of $\mathcal{X}$ in such a way that the cone is contained in $\mathcal{X}$. See Paragraph 4.8 in Adams and Fournier (2003) for more details.

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

## 2.3. Hypotheses and Notion of Test

Define

$$\Theta_I \equiv \{\theta = (\phi', g)' \in \mathbf{\Phi} \times \mathcal{G} : \mathbb{E}\left[m_s(Y_i, \phi_s, \mathbf{g}(X_i))|\underline{z}_{is}\right] = 0 \text{ a.s. for } s = 1, \ldots, T-R\}. \quad \textbf{(2.10)}$$

$\Theta_I$ is referred to as the identified set in the literature. We say that $\theta = (\phi', g)'$ is partially identified by (2.7) if $\Theta_I$ contains more than one element. The following lemma implies that there is no loss of information by considering $\Theta_I$ defined in (2.10) instead of (1.1) and (1.2), that is, the original model.

LEMMA 2.1 (No loss of information). *$\Theta_I$, the identified set defined by (2.10), is the same as the identified set characterized by (1.1) and (1.2). That is, $\Theta_I$ is equivalent to*

$$\left\{\theta = (\phi', g)' \in \Theta : \begin{array}{l} \text{For some } R\text{-dimensional vector } \lambda_i, \text{ it holds} \\ \mathbb{E}\left[y_{it} - g(x_{it}) - \lambda_i'\phi_t|\underline{z}_{it}\right] = 0 \text{ a.s. for } t = 1, \ldots, T-R \\ \mathbb{E}\left[y_{it} - g(x_{it}) - \lambda_i'\left(-\iota_{t-(T-R)}\right)|\underline{z}_{it}\right] = 0 \text{ a.s. for } t = T-R+1, \ldots, T \end{array}\right\}$$

*where $\iota_t$ represents the t'th column of the $R \times R$ identity matrix.*

For hypothesis testing on a conjectured restriction on $\theta$, in the generic form

$L(\theta) = l,$

we consider testing whether such a restriction is satisfied by at least one element of the identified set. Equivalently, defining the restricted set as $\Theta_R \equiv \{\theta \in \Theta : L(\theta) = l\}$, the null and alternative hypotheses under our consideration are

$\mathbb{H}_0 : \Theta_I \cap \Theta_R \neq \varnothing$ v.s. $\mathbb{H}_1 : \Theta_I \cap \Theta_R = \varnothing$,

where $\varnothing$ denotes an empty set. The notion of the above testing hypotheses is widely adopted under partial identification. When $\theta^0 = (\phi^{0\prime}, g^0)'$ is point-identified by (2.7), the above null hypothesis $\mathbb{H}_0$ simply tests whether $\theta^0$ satisfies the specified restriction in $\Theta_R : L(\theta^0) = l$.

We consider a group of restrictions that is otherwise identical to the one considered in Santos (2012) and Hong (2017), except for relaxing their requirement for restrictions to be linear, as follows.

**Assumption 2.2.** For $(\mathcal{L}, \|\cdot\|_{\mathcal{L}})$ a Banach space, $L : (\mathcal{G}, \|\cdot\|_c) \to (\mathcal{L}, \|\cdot\|_{\mathcal{L}})$ is a bounded operator.

As discussed in Santos (2012), Assumption 2.2, even when strengthened by requiring $L(\cdot)$ to be linear, encompasses a broad group of restrictions since one can flexibly choose the Banach space $(\mathcal{L}, \|\cdot\|_{\mathcal{L}})$. For example, we can test whether the value of the function $g^0$ at a point $x^0$ is given by a value $\gamma^0$ by setting $(\mathcal{L}, \|\cdot\|_{\mathcal{L}}) = (\mathbb{R}, |\cdot|)$, $\Theta_R = \left\{g \in \mathcal{W}^s(\mathcal{X}) : g\left(x^0\right) = \gamma^0\right\}$, $L(g) = g\left(x^0\right)$, and $l = \gamma^0$. For another example, we can test whether $g^0$ is an affine function by setting $(\mathcal{L}, \|\cdot\|_{\mathcal{L}}) =$

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

$\left(L^2(X), \|\cdot\|_{L^2}\right)$, $\Theta_R = \{g \in \mathcal{W}^c(\mathcal{X}) : g(x) = \beta_0 + \beta_1' x$ for some $\left(\beta_0, \beta_1'\right)' \in \mathbb{R}^{d_x+1}\}$, $L(g) = g - P_{\mathcal{A}}(g)$, and $l = 0$. Here, $L^2(X) = \{b : \mathcal{X} \to \mathbb{R} : \mathbb{E}[b(X)^2] < \infty\}$, $\|b\|_{L^2} = \{\mathbb{E}[b(X)^2]\}^{1/2}$, and $P_{\mathcal{A}}(g)$ denote the projection of $g \in L^2(X)$ onto $\mathcal{A} \equiv \{g \in W^c(\mathcal{X}) : g(x) = \beta_0 + \beta_1' x$ for some $\left(\beta_0, \beta_1'\right)' \in \mathbb{R}^{d_x+1}\}$. Moreover, since we rely on the nonparametric sieve method for the estimation of the nonparametric function $g(\cdot)$ and the derivative operator is a linear operator, our method works directly for testing hypotheses on some functionals of the function $g$ such as its first derivatives. See Example 2.6 in Santos (2012) on the restriction of the price elasticity of demand which is about the derivative of a nonparametric function. Since we do not restrict the functional $L$ to be linear here, higher-order derivatives can also be tested in principle. For additional examples of restrictions that satisfy Assumption 2.2, see Santos (2012). We also note that, although we do not exclude nonlinear restrictions, they are in general computationally costly to incorporate for constructing the corresponding test statistics (to be specified in Section 3.1). In contrast, linear restrictions are computationally easy to incorporate, as we discuss at the end of Section 3.1.

In the panel data model with IFEs, $g(\cdot)$, or its functional, is typically the parameter of interest, in which case $\phi$ can be regarded as a nuisance parameter. For this reason, we mainly consider hypotheses that impose restrictions on $g(\cdot)$ alone, in which case the restriction to be tested takes the special form $L(g) = l$. Then the restricted set becomes $\Theta_R = \{\theta = (\phi, g) \in \mathbf{\Phi} \times \mathcal{G} : L(g) = l\}$.

## 3. THE TESTING PROCEDURE

### 3.1. Test Statistics

A popular method to handle hypothesis testing for conditional moment models is to construct test statistics based on equivalent unconditional moments. This method dates back to Bierens (1982) and has been adopted in many papers on point-identification analysis (see, e.g., Stinchcombe and White, 1998; Dominguez and Lobato, 2004), and in more recent papers on partial identification analysis such as Santos (2012), Andrews and Shi (2013), and Hong (2017). To adopt this method in our study, it requires the choice of a family of generically revealing functions $(\varphi_1(t_1,\cdot), \varphi_2(t_2,\cdot), \ldots \varphi_{T-R}(t_{T-R},\cdot))$ indexed by $\mathbf{t} \equiv \left(t_1', t_2', \ldots, t_{T-R}'\right)' \in \prod_{s=1}^{T-R} \mathcal{T}_s \equiv \mathcal{T}$ that satisfies the following condition:

$$\mathbb{E}\left[m_s(Y_i, \phi_s, \mathbf{g}(X_i)) | \underline{z}_{is}\right] = 0 \text{ a.s. iff } \mathbb{E}\left[m_s(Y_i, \phi_s, \mathbf{g}(X_i)) \varphi_s(t_s, \underline{z}_{is})\right] = 0 \text{ for all } t_s \in \mathcal{T}_s,$$

where it is worth noting that the dimension of $t_s \in \mathcal{T}_s$ is typically equal or comparable to $d_{\underline{z}_s} = s \cdot d_z$, where recall that $d_z$ is the dimension of $z_{it}$. Then we can construct the following test statistic:

$$\bar{J}_N \equiv \min_{\theta \in \Theta_N \cap \Theta_R} \max_{\mathbf{t} \in \mathcal{T}} N \cdot |J_N(\theta, \mathbf{t})|^2, \tag{3.1}$$

where

$$J_N(\theta,\mathbf{t}) \equiv \begin{pmatrix} \frac{1}{N}\sum_{i=1}^{N} m_{i1}(\phi_1,g)\varphi_1(t_1,\underline{z}_{i1}) \\ \frac{1}{N}\sum_{i=1}^{N} m_{i2}(\phi_2,g)\varphi_2(t_2,\underline{z}_{i2}) \\ \vdots \\ \frac{1}{N}\sum_{i=1}^{N} m_{i,T-R}(\phi_{T-R},g)\varphi_{T-R}(t_{T-R},\underline{z}_{i,T-R}) \end{pmatrix} \equiv \begin{pmatrix} J_{N1}(\phi_1,g,t_1) \\ J_{N2}(\phi_2,g,t_2) \\ \vdots \\ J_{N,T-R}(\phi_{T-R},g,t_{T-R}) \end{pmatrix},$$

with $m_{is}(\phi_s,g) \equiv m_s(Y_i,\phi_s,\mathbf{g}(X_i))$ and $\Theta_N$ is an approximating sieve space for $\Theta$.

Hong (2017) shows under a general setting that statistics of the form (3.1) weakly converge to a certain functional of a Gaussian process under the null, and proposes a penalized bootstrap procedure for testing. These results are potentially applicable to the statistic $\bar{J}_N$ in our study. However, note that the dimension of $\underline{z}_{is}$ is given by $sd_z$, and the dimension of the index $\mathbf{t}$ is typically equal or comparable to $\left(\sum_{s=1}^{T-R} s\right) d_z$, which can get relatively large for moderate sizes of $T-R$ and $d_z$. As a result, the computation of $\bar{J}_N$ can be rather expensive. For this reason, we opt for a different test statistic based on the notion of MDD.

An MDD-based statistic for our study is motivated by two recent papers: Shao and Zhang (2014) and Su and Zheng (2017). In its primitive form, Shao and Zhang (2014) define for any real-valued variable $V$ and vector-valued variable $W$,

$$\mathrm{MDD}_o(V|W)^2 \equiv \int_{\mathbb{R}^{d_W}} \left[\mathrm{Cov}\left(V,\exp\left(\mathbf{i}s'W\right)\right)\right]^2 \cdot q(s)\,ds, \tag{3.2}$$

where $\mathbf{i} \equiv \sqrt{-1}$, $d_W$ is the dimension of $W$, and $q(\cdot)$ is a nonnegative weight function. As revealed by ( 3.2), $\mathrm{MDD}_o(V|W)^2$ is constructed as a weighted integration of $\left[\mathrm{Cov}\left(V,\exp\left(\mathbf{i}s'W\right)\right)\right]^2$ over $s \in \mathbb{R}^{d_W}$ and it is clearly motivated by a known statistic result that

$$\mathbb{E}(V|W) = \mathbb{E}(V) \text{ if and only if } \mathrm{Cov}\left(V,\exp\left(\mathbf{i}s'W\right)\right) = 0 \text{ for all } s \in \mathbb{R}^{d_W}. \tag{3.3}$$

When picking $q(s) = 1/[c|s|^{(1+d_Z)}]$ with $c \equiv \pi^{(1+d_Z)/2}/\Gamma(\frac{1+d_Z}{2})$ and $\Gamma(\cdot)$ being the complete gamma function $\Gamma(z) \equiv \int_0^\infty t^{(z-1)}\exp(-t)\,dt$, Shao and Zhang (2014) establish that

$$\mathrm{MDD}_o(V|W)^2 \equiv -\mathbb{E}\left\{[V-\mathbb{E}(V)]\left[V^\dagger - \mathbb{E}(V^\dagger)\right]\left|W-W^\dagger\right|\right\}, \tag{3.4}$$

where $(V^\dagger, W^\dagger)$ is an independent copy of $(V,W)$. Then under some suitable moment conditions ($\mathbb{E}(V^2) < \infty$ and $0 < \mathbb{E}(|W|^2) < \infty$), one has $\mathrm{MDD}_o(V|W)^2 \geq 0$ and

$$\mathrm{MDD}_o(V|W)^2 = 0 \text{ iff } \mathbb{E}(V|W) = \mathbb{E}(V) \quad \text{a.s.} \tag{3.5}$$

Based on the above properties, they propose a consistent test for conditional mean independence condition of (3.5).

Su and Zheng (2017) propose a modified version of MDD, defined as

$$\mathrm{MDD}(\varepsilon|W)^2 \equiv -\mathbb{E}\left[\varepsilon\varepsilon^\dagger\left|W-W^\dagger\right|\right] + 2\mathbb{E}\left[\varepsilon\left|W-W^\dagger\right|\right]\mathbb{E}\left[\varepsilon^\dagger\right], \tag{3.6}$$

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

where $\varepsilon$ is the error term of a nonlinear regression model, $W$ is the regressor, and $\left(\varepsilon^{\dagger}, W^{\dagger}\right)$ is an independent copy of $(\varepsilon, W)$. Su and Zheng (2017) show that

$$\mathrm{MDD}(\varepsilon|W)^2 = \mathrm{MDD}_o(\varepsilon|W)^2 + [\mathbb{E}(\varepsilon)]^2 \mathbb{E}\left[\left|W - W^{\dagger}\right|\right] \geq 0 \tag{3.7}$$

and

$$\mathrm{MDD}(\varepsilon|W)^2 = 0 \text{ iff } \mathbb{E}(\varepsilon|W) = 0 \quad \text{a.s.}, \tag{3.8}$$

where $\mathrm{MDD}_o(\varepsilon|W)^2$ refers to the one used in (3.4). Then they propose a novel and effective test for correct (parametric) specification of the regression function based on the above properties of MDD. Notably, in their setting, the regression function is correctly specified if and only if the conditional mean zero condition of (3.8) holds. Simulation results in Su and Zheng (2017) indicate that a test statistic based on the MDD significantly outperforms many popular specification tests in the literature, and that it performs well even when the dimension of $W$ is moderately large.

Following the insights from Su and Zheng (2017), it can be shown, under our setting of $\Theta$ and $\Theta_R$, that $\Theta_I \cap \Theta_R \neq \varnothing$ if and only if

$$\min_{\theta \in \Theta \cap \Theta_R} \left\{ \sum_{s=1}^{T-R} \mathrm{MDD}\left[m_s(Y, \phi_s, \mathbf{g}(X)) | \underline{z}_s\right]^2 \right\} = 0. \tag{3.9}$$

Then the construction of the test statistic for

$$\mathbb{H}_0 : \Theta_I \cap \Theta_R \neq \varnothing \quad \text{v.s.} \quad \mathbb{H}_1 : \Theta_I \cap \Theta_R = \varnothing \tag{3.10}$$

proceeds in two steps as follows:

I. Fix $\theta \in \Theta$ and derive a test statistic $S_N(\theta)$ for the null hypothesis $\mathbb{H}_0$ : $\mathbb{E}\left[m_s(Y, \phi_s, \mathbf{g}(X)) | \underline{z}_s\right] = 0$ a.s. for all $s = 1, \ldots, T-R$, or equivalently, $\mathbb{H}_0$ : $\sum_{s=1}^{T-R} \mathrm{MDD}\left[m_s(Y, \phi_s, \mathbf{g}(X)) | \underline{z}_s\right]^2 = 0$ for all $s = 1, \ldots, T-R$.

II. Let $\Theta_N$ be a sieve approximating space for $\Theta$. Then, following (3.9), test $\mathbb{H}_0$ : $\Theta_I \cap \Theta_R \neq \varnothing$ by using the statistic $\hat{S}_N = \min_{\theta \in \Theta_N \cap \Theta_R} S_N(\theta)$.

For Step I, we propose the following statistic:

$$S_N(\theta) \equiv \sum_{s=1}^{T-R} S_{Ns}(\theta), \tag{3.11}$$

where $S_{Ns}(\theta)$ is constructed in a way similar to Su and Zheng (2017):

$$S_{Ns}(\theta) = -\frac{1}{N} \sum_{1 \leq i \neq j \leq N} m_{is}(\theta) m_{js}(\theta) \kappa_{ij,s} + \frac{2}{N} \sum_{1 \leq i \neq j \leq N} m_{is}(\theta) \kappa_{ij,s} \frac{1}{N} \sum_{k=1}^{N} m_{ks}(\theta) \tag{3.12}$$

with $m_{is}(\theta) \equiv m_s(Y_i, \phi_s, \mathbf{g}(X_i))$ and $\kappa_{ij,s} = \left|\underline{z}_{is} - \underline{z}_{js}\right|$ for $s = 1, \ldots, T-R$.

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

For Step II, we define the test statistic accordingly as

$$\hat{S}_N = \min_{\theta \in \Theta_N \cap \Theta_R} S_N(\theta), \tag{3.13}$$

where $\Theta_N = \mathbf{\Phi} \times \mathcal{G}_N$ is an approximating space for $\Theta = \mathbf{\Phi} \times \mathcal{G}$, with

$$\mathcal{G}_N = \left\{ g \in \mathcal{G} : g(\cdot) = p^{k_N}(\cdot)' \beta \text{ for some } \beta \in \mathbb{R}^{k_N} \right\}$$

being an approximating space for $\mathcal{G}$ by using a $k_N$-vector of basis functions $p^{k_N}(\cdot) = \left(p_1(\cdot), \ldots, p_{k_N}(\cdot)\right)'$ defined on $\mathcal{X}$. Computation-wise, it is helpful to note that the minimization over $\theta = \left(\phi', g(\cdot)\right)' \in \Theta_N \cap \Theta_R$, involved in the computation of $\hat{S}_N$, is equivalent to a minimization over

$$(\phi', \beta')' \in \left\{ (\phi', \beta')' \in \mathbb{R}^{T-R} \times \mathbb{R}^{k_N} : p^{k_N}(\cdot)' \beta \in \mathcal{G}_N \text{ and } L\left( \left( \phi', p^{k_N}(\cdot)' \beta \right)' \right) = l \right\} \subset \mathbb{R}^{(T-R)+k_N},$$

where $L(\theta) = l$ is the restriction being tested. A case of major interest is where: (i) the tested restriction takes the form $L(g) = l$ (i.e., it only concerns $g(\cdot)$, and puts no restriction on $\phi$), and (ii) $L(\cdot)$ is linear. In this case, $L\left( \left( p^{k_N}(\cdot)' \beta \right)' \right) = l$ becomes a linear restriction on $\beta \in \mathbb{R}^{k_N}$: $\Upsilon_N' \beta = l$ with $\Upsilon_N \equiv \left( L\left(p^1(\cdot)\right), \ldots, L\left(p^{k_N}(\cdot)\right)\right)'$, which is easy to incorporate to the involved minimization.

## 3.2. Definitions and Notations

For $\theta^0 = \left(\phi^{0\prime}, g^0\right)' \in \Theta_I \cap \Theta_R$, let

$$\Pi_N \theta^0 \equiv \begin{pmatrix} \phi^0 \\ \Pi_{\mathcal{G}_N} g^0 \end{pmatrix} \tag{3.14}$$

be the projection of $\theta^0$ onto $\Theta_N \cap \Theta_R$.

Definition 3.1 (Weak pseudo-metric). *Let* $m_s(Y, X, \theta) \equiv m_s(Y, \phi_s, \mathbf{g}(X))$. *Define the following pseudo-metric* $\mathrm{d}_w(\cdot, \cdot)$ *on* $\Theta$*:*

$$\mathrm{d}_w(\theta_1, \theta_2) \equiv \left\{ \sum_{s=1}^{T-R} \mathrm{MDD}\left[ \left( m_s(Y, X, \theta_1) - m_s(Y, X, \theta_2) \right) | \underline{z}_s \right]^2 \right\}^{1/2}$$

*for any* $\theta_1$, $\theta_2 \in \Theta$.

Note that $\Theta_I$ forms an equivalence class under $\mathrm{d}_w(\cdot, \cdot)$, that is, for any $\theta_1^0 \in \Theta_I$ and $\theta_2^0 \in \Theta_I$, $\mathrm{d}_w\left(\theta_1^0, \theta_2^0\right) = 0$, which is made clear by Lemma A.1 in the Appendix. It also follows from Lemma A.1 that for any given $\theta \in \Theta$ and $\theta^0 \in \Theta_I$,

$$\mathrm{d}_w\left(\theta, \theta^0\right) = \left\{ \sum_{s=1}^{T-R} \mathrm{MDD}\left[ m_s(Y, X, \theta) | \underline{z}_s \right]^2 \right\}^{1/2}.$$

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

Also note that $\mathrm{d}_w(\cdot,\cdot)$ is weaker than the $L^2$-metric and satisfies a triangle-like inequality, as stated in the following lemma.

LEMMA 3.1. *Let Assumptions 2.1 hold. Suppose that* $\mathbb{E}[|Y|^2] < \infty$ *and* $\mathbb{E}|Z| < \infty$. *(i) There exists a finite constant* $c > 0$ *s.t.* $\mathrm{d}_w(\theta_1,\theta_2) \le c\,\|\theta_1 - \theta_2\|_{L^2}$ *for any* $\theta_1$, $\theta_2 \in \Theta$*; (ii) It holds that*

$$\mathrm{d}_w(\theta_1,\theta_2) \;\le\; 2\sqrt{2}\left[\mathrm{d}_w(\theta_1,\theta_3) + \mathrm{d}_w(\theta_2,\theta_3)\right]$$

*for any* $\theta_1$, $\theta_2$, $\theta_3 \in \Theta$.

For a given point $\theta \in \Theta$ and a given subset $A \subseteq \Theta$, denote by

$$\mathrm{d}_{\|\cdot\|_{L^2}}(\theta,A) \equiv \inf_{\tilde{\theta}\in A}\left\|\theta - \tilde{\theta}\right\|_{L^2} \text{ and } \mathrm{d}_w(\theta,A) \equiv \inf_{\tilde{\theta}\in A}\mathrm{d}_w(\theta,\tilde{\theta}),$$

that is, the distances between point $\theta$ and set $A$ under $\|\cdot\|_{L^2}$ and $\mathrm{d}_w(\cdot,\cdot)$, respectively. As formalized in Lemma A.6(i) in the Appendix, under mild conditions (Assumptions 2.1, 2.2, and 3.1–3.3(i) to be specified in the next subsection), any minimizer $\hat{\theta}_N$ of $S_N(\theta)$ over $\Theta_N$ would lie in a $o_p(1)$ -neighborhood of the identified set $\Theta_I$ under the $L^2$-norm. Given this consistency result, we can restrict our attention on a shrinking $L^2$ sieve neighborhood around $\Theta_I$, defined as

$$\Theta_{oN} \equiv \left\{\theta \in \Theta_N : \mathrm{d}_{\|\cdot\|_{L^2}}(\theta,\Theta_I) \le \varsigma_N\right\}$$

for some positive $\varsigma_N \downarrow 0$. And we define the following measure of local ill-posedness.

DEFINITION 3.2 (Sieve measure of local ill-posedness). *Define the following sieve measure of local ill-posedness:*

$$\varrho_N \equiv \sup_{\theta\in\Theta_{oN}:\theta\notin\Pi_N\Theta_I} \frac{\mathrm{d}_{\|\cdot\|_{L^2}}(\theta,\Pi_N\Theta_I)}{\mathrm{d}_w(\theta,\Pi_N\Theta_I)}.$$

$\varrho_N$ provides a local upper bound of relative distance between $\theta$ and $\Pi_N\Theta_I$ under $L^2$ to that under our pseudo metric $\mathrm{d}_w(\cdot,\cdot)$. We note that our definition of $\varrho_N$ is otherwise identical to the sieve measure of local ill-posedness defined by Hong (2017), except for that our pseudo metric (in the involved denominator of $\varrho_N$) differs from that of Hong (2017). Under point-identification where $\Theta_I$ is just the singleton $\{\theta^0\}$, $\varrho_N$ becomes $\varrho_N = \sup_{\theta\in\Theta_{oN}:\theta\notin\Pi_N\theta^0} \frac{\mathrm{d}_{\|\cdot\|_{L^2}}\left(\theta,\Pi_N\theta^0\right)}{\mathrm{d}_w\left(\theta,\Pi_N\theta^0\right)}$, the form of which is conventional for sieve measures of local ill-posedness defined in the point-identification literature (see, for example, Ai and Chen, 2003; Chen and Pouzo, 2012, among others).

In our analysis, we allow for moderate ill-posedness in the sense that $\varrho_N \uparrow \infty$ but at a slow rate. $\varrho_N$ provides a link between $d_w(\cdot,\cdot)$ and the $L^2$ distance. So once we establish the rate of convergence under $d_w(\cdot,\cdot)$, a certain rate of convergence

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

under the $L^2$ distance can be established via $\varrho_N$. Note that $\varrho_N$ is defined in a way similar to Hong (2017) with the main difference that our pseudo metric is different from that of Hong (2017).

### 3.3. Asymptotic Theory

In this section, we study the asymptotic properties of $S_N(\theta)$ and $\hat{S}_N$, defined in (3.11) and (3.13), respectively. To establish the asymptotic behavior of $S_N(\theta)$, we impose the following assumption.

**Assumption 3.1.** (i) $\{X_i, Y_i, Z_i\}_{i=1}^N$ are i.i.d. with support $\mathcal{X}^T \times \mathcal{Y}^T \times \mathcal{Z}^T$ such that all marginal and joint density functions of $X_i$, $Y_i$ and $Z_i$ are bounded.
(ii) $\mathbb{E}[(|Y_i|^2+1)(|Z_i|^2+1)] < \infty$.

Assumption 3.1(i) is commonly imposed for panel data analyses with individual fixed effects or IFEs. Note that it does not rule out dynamic panels as long as we treat the unobserved factors $F_t^0$'s as nonrandom.[2] In the case where $F_t^0$'s are random, the independence assumption can be replaced by conditional independence: the lagged dependent variables (e.g., $Y_{i,t-1}$) can be independent across $i$ given the minimal sigma-field generated by the common factors. Assumption 3.1(ii) specifies some moment conditions on $Y_i$ and $Z_i$.

Let $\xi_i = \left(Y_i', X_i', Z_i'\right)'$ and $\tilde{m}_{is}(\theta) = m_s(Y_i, X_i, \theta) - \mathbb{E}\left[m_s(Y_i, X_i, \theta)\,|\underline{z}_{is}\right]$. Define the second-order $U$-process indexed by $\theta$ as follows:

$$\mathbb{U}_{Ns}(\theta) = \binom{N}{2}^{-1} \sum_{1 \le i < j \le N} h_s\left(\xi_i, \xi_j; \theta\right),$$

where $h_s\left(\xi_i, \xi_j; \theta\right) = \tilde{m}_{is}(\theta)\,\tilde{m}_{js}(\theta)\,[\mathbb{E}_j(\kappa_{ij,s}) + \mathbb{E}_i(\kappa_{ij,s}) - \kappa_{ij,s}]$, and $\mathbb{E}_j(\kappa_{ij,s})$ denotes the expectation with respect to (w.r.t.) the variable $\underline{z}_{js}$ alone in $\kappa_{ij,s} = \left|\underline{z}_{is} - \underline{z}_{js}\right|$. Denote by $< \cdot, \cdot >$ the usual inner product on $L^2(P)$, for $P$ a generic probability measure (or a generic marginal one) on $\mathcal{X}^T \times \mathcal{Y}^T \times \mathcal{Z}^T$. For $f \in L^2\left(P^2\right) \equiv L^2(P \otimes P)$, define a Hilbert–Schmidt operator $H_f$ on $L^2(P)$ by $\left(H_f g\right)(\xi) = Pf(\xi, \cdot)\, g(\cdot)$. Also, define a process $\mathbb{C}$ on $\mathcal{F}$ by

$$\mathbb{C}(f) = \sum_{\alpha=1}^{\infty} < H_f w_\alpha, w_\alpha > \left(W_\alpha^2 - 1\right),$$

[2]For simplicity, we consider a simple parametric model $y_{it} = \rho^0 y_{i,t-1} + \lambda_i^{0\prime} F_t^0 + u_{it}$, where $\rho^0 \neq 0$ and $\left|\rho^0\right| < 1$, and we assume $y_{i0}$'s are observed. By the continuous backward substitutions, we have

$$y_{it} = \rho^t y_{i,0} + \sum_{j=0}^{t-1} \rho^j (\lambda_i^{0\prime} F_{t-j}^0 + u_{i,t-j}).$$

If $F_t^0$'s are nonrandom, the randomness of $y_{it}$ is mainly driven by that of $\lambda_i^0$, $y_{i0}$, and $\{u_{is}\}_{s=1}^t$. Then $y_{it}$'s are cross-sectionally independent provided $\lambda_i^0$, $y_{i0}$, and $\{u_{is}\}_{s=1}^t$ are independent across $i$, and so are $Y_i \equiv (y_{i1}, \ldots, y_{iT})'$. Similarly, if $\lambda_i^0$, $y_{i0}$, and $\{u_{is}\}_{s=1}^T$ share the same distributions across $i$, $Y_i$ would be identically distributed. So a dynamic model could meet the i.i.d. requirement in Assumption 3.1(i) where $x_{it} = y_{i,t-1}$.

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

where $\{w_\alpha\}$ denotes the eigenfunctions of the operator $H_f$, and $\{W_\alpha\}$ is a sequence of independent $N(0,1)$ random variables.

The following theorem studies the asymptotic properties of the process $\{S_N(\theta)\}$.

THEOREM 3.1. *Let Assumptions 2.1, 2.2, and 3.1 hold. Then:*

(i) *For each* $s=1,\ldots,T-R$, $S_{Ns}(\theta)=2\mathbb{E}\left[m_s^2(Y,\phi_s,\mathbf{g}(X))\left|\underline{z}_s-\underline{z}_s^\dagger\right|\right]+\mathbb{U}_{Ns}(\theta)+O_p(N^{-1/2})\Longrightarrow\mathbb{B}_s(\theta)+\mathbb{C}_s(\theta)$ *in* $L^\infty(\Theta_I)$, *where* $\mathbb{B}_s(\theta)=2\mathbb{E}[m_s^2(Y,\phi_s,\mathbf{g}(X))|\underline{z}_s-\underline{z}_s^\dagger|]$ *and* $\mathbb{C}_s(\theta)=\mathbb{C}(h_s(\cdot,\cdot;\theta))$ *is a Gaussian chaos process on* $L^\infty(\Theta_I)$.[3] $S_N(\theta)\Longrightarrow\sum_{s=1}^{T-R}[\mathbb{B}_s(\theta)+\mathbb{C}_s(\theta)]$ *on* $L^\infty(\Theta_I)$.

(ii) $\frac{1}{N}S_N(\theta)=\sum_{s=1}^{T-R}\mathrm{MDD}\left[m_s(Y,X,\theta)|\underline{z}_s\right]^2+O_p(N^{-1/2})$ *uniformly in* $\theta\in\Theta\backslash\Theta_I$.

Theorem 3.1(i) indicates that $S_{Ns}(\theta)$, after being recentered around $\mathbb{B}_s(\theta)$, is essentially a degenerate second-order $U$-process on $\Theta_I$ that converges weakly to a Gaussian chaos process $\{\mathbb{C}_s(\theta)\}$. Theorem 3.1(ii) indicates that for $\theta\in\Theta\backslash\Theta_I$, $S_N(\theta)$ is dominated by its deterministic component that is associated with the MDD measure.

To establish the asymptotic behavior of $\hat{S}_N$, we need to impose some further assumptions.

**Assumption 3.2.** The eigenvalues of $\mathbb{E}\left[p^{k_N}(x_t)p^{k_N'}(x_t)\right]$ for $t=1,\ldots,T$ are uniformly bounded and uniformly bounded away from zero.

**Assumption 3.3.** $\Theta_N$ is a closed subset of $\Theta$, and there exists $\Pi_N\theta\in\Theta_N$ for each $\theta\in\Theta$ such that: (i) $\delta_{s,N}\equiv\sup_{\theta\in\Theta\cap\Theta_R}\left\|\Pi_N\theta-\theta\right\|_{L^2}=o\left(N^{-1/2}\right)$; (ii) $\delta_{w,N}\equiv\sup_{\theta\in\Theta_I\cap\Theta_R}\mathrm{d}_w\left(\Pi_N\theta,\theta\right)=o\left(N^{-1/2}\right)$.

**Assumption 3.4.** $\varrho_N=O\left(N^{(1-\epsilon)/(4(2-\epsilon))}\right)$ for some arbitrarily small $\epsilon>0$.

Assumption 3.2 is identical to Assumption 3.3(i) in Santos (2012) and is commonly assumed in the literature on sieve estimation. Assumption 3.3(i) requires a uniform sieve approximation error rate under $\|\cdot\|_{L^2}$ over $\Theta\cap\Theta_R$. As discussed previously, the compactness of $\Theta$ under $\|\cdot\|_{L^2}$ makes the results in Schumaker (2007) applicable to developing a primitive condition for Assumption 3.3(i). Specifically, according to Theorem 6.25 in Schumaker (2007), $\sup_{\theta\in\Theta}\|\Pi_N\theta-\theta\|_{L^2}=O(k_N^{-(d-1)})$ for B-splines with simple knots. Therefore, Assumption 3.3(i) is satisfied by picking $k_N\to\infty$ fast enough such that $1/k_N=o\left(N^{-1/2(d-1)}\right)$. Since $\mathrm{d}_w\left(\Pi_N\theta,\theta\right)$ is controlled from above by $\|\Pi_N\theta-\theta\|_{L^2}$ as shown by Lemma 3.1, Assumption 3.3(ii) can be verified by using results for $\|\cdot\|_{L^2}$. Assumption 3.4 allows $\varrho_N\to\infty$, but restricts its divergence rate to be slow enough. Essentially,

[3]Let $\mathbb{D}(b)$ be a generic stochastic process indexed by $b\in\mathcal{B}$. $\mathbb{D}(b)$ is said to be a process on $L^\infty(\mathcal{B})$ if $\mathbb{D}(\cdot)$ (treated as a random function with domain $\mathcal{B}$) has almost sure bounded paths (i.e., realizations) on $\mathcal{B}$.

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

this requires $k_N$ to grow sufficiently slow. Similar assumptions are required in semi/nonparametric analyses for regularization of ill-posed problems (see, e.g., Blundell, Chen, and Kristensen, 2007; Chen and Pouzo, 2012; Hong, 2017). As acknowledged in Hong (2017), Assumption 3.4 is generally hard to verify because the nature of the dependence of $\varrho_N$ on $k_N$ has not been well studied. In the special case of $R = 0$ (i.e., no IFEs) and point-identification, Assumption 3.2 is sufficient to guarantee $\mathrm{d}_w\left(\theta, \theta^0\right) \asymp \left\|\theta - \theta^0\right\|_{L^2}$ asymptotically for any $\theta^0 \in \Theta_I$ and $\left\|\theta - \theta^0\right\|_{L^2} = o(1)$, which implies that $\varrho_N = O(1)$. Then Assumption 3.4 holds trivially. In the Appendix C of the Supplementary Material, we clarify this claim for the case where $R = 0$ and also provide some further discussions on the sufficient conditions for Assumption 3.4 when $R \geq 1$.

With the above additional assumptions, we can state the next main result in this paper.

THEOREM 3.2 (Convergence of $\hat{\theta}_N$). *Let Assumptions 2.1, 2.2, and 3.1–3.4 hold. For any $\hat{\theta}_N \in \underset{\theta \in \Theta_N \cap \Theta_R}{\operatorname{argmin}} S_N(\theta)$, when $\Theta_I \cap \Theta_R \neq \Phi$, it holds that*

$$\mathrm{d}_w\left(\hat{\theta}_N, \Theta_I \cap \Theta_R\right) = O_p(\min(N^{-1/4}, \varrho_N N^{-1/2})) = o_p(N^{-\frac{1}{2}+\frac{\epsilon}{4}} \varrho_N^{1-\epsilon}). \tag{3.15}$$

A close examination of the proof of Theorem 3.2 suggests that we can first show that $\mathrm{d}_w(\hat{\theta}_N, \Theta_I \cap \Theta_R) = O_p(N^{-1/4})$ under Assumptions 2.1, 2.2, and 3.1–3.3. Such a rate can be improved to $o_p(N^{-\frac{1}{2}+\frac{\epsilon}{4}} \varrho_N^{1-\epsilon})$ by using the link between $\mathrm{d}_w$ and $\mathrm{d}_{\|\cdot\|_{L^2}}$ through the sieve measure of ill-posedness and some iterative arguments. By Assumptions 3.3–3.4 and Lemma A.6 in the Appendix, we can show that $\mathrm{d}_{\|\cdot\|_{L^2}}(\hat{\theta}_N, \Theta_I \cap \Theta_R) = o_p(\varrho_N^{2-\epsilon} N^{-\frac{1}{2}+\frac{\epsilon}{4}}) = o_p(N^{-1/4})$, which will be used in the proof of the next main result.

THEOREM 3.3 (Asymptotic distribution under $\mathbb{H}_0$). *Let Assumption 2.1, 2.2, and 3.1–3.3 hold. Under $\mathbb{H}_0 : \Theta_I \cap \Theta_R \neq \varnothing$, we have*

$$\hat{S}_N \leq \inf_{\theta^0 \in \Theta_I \cap \Theta_R} S_N\left(\theta^0\right) + o_p(1) \xrightarrow{\mathcal{L}} \inf_{\theta \in \Theta_I \cap \Theta_R} \sum_{s=1}^{T-R} \left[\mathbb{B}_s(\theta) + \mathbb{C}_s(\theta)\right]. \tag{3.16}$$

Given the non-negativity of $\hat{S}_N$ by construction,[4] Theorem 3.3 establishes a stochastic upper bound, viz., $\inf_{\theta^0 \in \Theta_I \cap \Theta_R} S_N\left(\theta^0\right)$, of $\hat{S}_N$ under the null hypothesis, which weakly converges to a tight distribution.[5] Apparently, the asymptotic distribution in (3.16) is not asymptotically pivotal and we will provide a bootstrap

[4]To see why $\hat{S}_N$ is nonnegative, note that $S_{Ns}(\theta)$ can be viewed as $\mathrm{MDD}\left(m_s(Y, \phi_s, \mathbf{g}(X)) | \underline{Z}_s\right)^2$ for random variables $\{X, Y, Z\}$ that follow the empirical distribution generated by the i.i.d. sample $\{X_i, Y_i, Z_i\}_{i=1}^N$. It follows from the non-negativity of $\mathrm{MDD}(\cdot|\cdot)^2$ in general that $S_{Ns}(\theta)$ is nonnegative, which in turn implies $\hat{S}_N = \min_{\theta \in \Theta_N \cap \Theta_R} \sum_{s=1}^{T-R} S_{Ns}(\theta)$ (as specified in (3.11)–(3.13)) to be nonnegative.

[5]As shown in the Appendix, $\hat{S}_N = \inf_{\theta^0 \in \Theta_I \cap \Theta_R} S_N\left(\theta^0\right) - c_N + o_p(1)$, where $0 \leq c_N \leq \bar{c}_N$ is defined in (A.56).

method in the next subsection to obtain a potentially conservative bootstrap $p$-value for the purpose of inference.

THEOREM 3.4 (Asymptotic behavior under $\mathbb{H}_1$). *Let Assumptions 2.1, 2.2, and 3.1–3.3 hold. Under* $\mathbb{H}_1 : \Theta_I \cap \Theta_R = \varnothing$, *we have*

$$N^{-1}\hat{S}_N \xrightarrow{p} \min_{\theta\in\Theta\cap\Theta_R} \sum_{s=1}^{T-R} \mathrm{MDD}\left[m_s(Y,X,\theta)\,|\underline{z}_s\right]^2 > 0.$$

Theorem 3.4 studies the asymptotic behavior of $\hat{S}_N$ under the alternative. It indicates that $\hat{S}_N$ diverges to infinity in probability at rate$-N$, which gives the power of the MDD-based test.

Note that we do not provide the analysis on the asymptotic local power property of our test. In the case of point-identification, it is well known that an MDD-based test can detect the local alternatives converging to the null at the usual parametric rate despite the dimension of the conditioning variable as long as it is fixed. Nevertheless, like any other nonparametric nonsmoothing test, the MDD test still suffers from the notorious curse of dimensionality in the nonparametric literature. To appreciate this, we can focus on the general MMD measure considered in (3.6) in Section 3.1. Let $W = \left(W_1,\ldots,W_{d_W}\right)'$ and $W^\dagger = (W_1^\dagger,\ldots,W_{d_W}^\dagger)'$. Note that $\left|W-W^\dagger\right|$ enters $\mathrm{MDD}(\varepsilon|W)^2$ in ( 3.6) explicitly. It is important to understand its behavior as $d_W$ increases. For clarity, suppose the dependence among $W_1,\ldots,W_{d_W}$ is weak with sufficiently high-order moments so that

$$\frac{1}{d_W}\left|W-W^\dagger\right|^2 \equiv \frac{1}{d_W}\sum_{l=1}^{d_W}\left(W_l-W_l^\dagger\right)^2 \xrightarrow{p} \lim_{d_W\to\infty}\frac{1}{d_W}\sum_{l=1}^{d_W} E\left[\left(W_l-W_l^\dagger\right)^2\right] \equiv \mu_W.$$

Intuitively, when a law of large numbers applies to $\{W_l^2\}_{l=1}^{d_W}$ as $d_W\to\infty$, $\frac{1}{\sqrt{d_W}}$ $\left|W-W^\dagger\right|$ converges in probability to a constant $\sqrt{\mu_W}$. Then the information inside $\left|W-W^\dagger\right|$ vanishes asymptotically. In this case, we can approximate the $\mathrm{MDD}(\varepsilon|W)^2$ in (3.7) as follows:

$$\mathrm{MDD}(\varepsilon|W)^2 \approx -\sqrt{d_W\mu_W}\left\{\mathbb{E}\left[\varepsilon\varepsilon^\dagger\right]+2\mathbb{E}[\varepsilon]\,\mathbb{E}\left[\varepsilon^\dagger\right]\right\},$$

where the right-hand side becomes 0 provided $\mathbb{E}[\varepsilon]=0$ by noting the independence between $\varepsilon$ and $\varepsilon^\dagger$. Since it is possible to have $\mathbb{E}[\varepsilon]=0$ and $\Pr(\mathbb{E}[\varepsilon|W]=0)<1$, this means $\mathrm{MDD}(\varepsilon|W)^2$ will lose power to test deviations from $\mathbb{E}[\varepsilon|W]=0$ in such a situation. In other words, the large dimension $d_W$ of $W$ may have an adverse effect on the asymptotic power of a MDD-based test, which is surely the curse of dimensionality problem.

### 3.4. A Multiplier Bootstrap

The asymptotic distribution specified in Theorem 3.3, as an asymptotic distributional upper bound for $\hat{S}_N$ under the null hypothesis, is nonstandard and unfamiliar. Here, we propose a bootstrap procedure to obtain the bootstrap $p$-values despite its conservative nature. To ensure the consistency of the bootstrap procedure, we need to find appropriate ways: (i) to mimic the limiting law of the stochastic upper bound $\inf_{\theta^0\in\Theta_I\cap\Theta_R} S_N\left(\theta^0\right)$ for $\hat{S}_N$ that is associated with a Gaussian chaos process $\mathbb{C}(\theta) = (\mathbb{C}_1(\theta),\ldots,\mathbb{C}_{T-R}(\theta))'$ on $L^\infty(\Theta_I)$ under the null, and (ii) to ensure the bootstrap statistic is well behaved or divergent to infinity at a slower rate than $\hat{S}_N$ under the global alternative.

Let $\{v_i\}_{i=1}^N$ be an i.i.d. sequence that has mean zero and variance one and that is independent of the sample $\{(Y_i,X_i,Z_i)\}_{i=1}^N$. Two popular choices of distributions for $\{v_i\}_{i=1}^N$ are given by the standard normal distribution $(N(0,1))$ and the two-point distribution:

$$v_i = \begin{cases} -\left(\sqrt{5}-1\right)/2 & \text{with prob. } \left(\sqrt{5}+1\right)/\left(2\sqrt{5}\right), \\ \left(\sqrt{5}+1\right)/2 & \text{with prob. } \left(\sqrt{5}-1\right)/\left(2\sqrt{5}\right). \end{cases} \tag{3.17}$$

Let $m_{is}^*(\theta) \equiv m_s(Y_i,\phi_s,\mathbf{g}(X_i))\,v_i$. Motivated by the idea of multiplier bootstrap that is widely used for statistical tests involved with empirical processes or nondegenerate $U$-processes, we consider the following process:

$$S_N^*(\theta) \equiv \sum_{s=1}^{T-R} S_{Ns}^*(\theta), \tag{3.18}$$

where

$$S_{Ns}^*(\theta) \equiv -\frac{1}{N}\sum_{1\le i\ne j\le N} m_{is}^*(\theta)\,m_{js}^*(\theta)\,\kappa_{ij,s} + \frac{2}{N}\sum_{1\le i\ne j\le N} m_{is}^*(\theta)\,\kappa_{ij,s}\frac{1}{N}\sum_{k=1}^N m_{ks}^*(\theta). \tag{3.19}$$

Let $P^*$ and $\mathbb{E}^*$ denote, respectively, the probability law and expectation associated with $(Y_i,X_i,Z_i,v_i)$ in the bootstrap world. We make two remarks on the construction of $S_N^*(\theta)$. First, note that we perturb $m_s(Y_i,\phi_s,\mathbf{g}(X_i))$ through the multiplication by the random variable $v_i$ that ensures $\mathbb{E}^*\left[m_{is}^*(\theta)\right]=0$. This ensures that the dominant random component in the process $\left\{S_{Ns}^*(\theta)\right\}$ is given by a degenerate second-order $U$-process that converges to a Gaussian chaos process. More importantly, we can show that the limiting law of $\left\{S_{Ns}^*(\theta)\right\}$ coincides with that of $\{S_{Ns}(\theta)\}$ on $\Theta_I$. Second, $\left\{S_{Ns}^*(\theta)\right\}$ is also well behaved for $\theta\in\Theta\backslash\Theta_I$ (i.e., it is not divergent on $\Theta\backslash\Theta_I$), and if we were to define the bootstrap statistic as $\min_{\theta\in\Theta_N\cap\Theta_R} S_N^*(\theta)$, there is no way to ensure that the minimum is achieved at some value in $\Theta_I\cap\Theta_R$ asymptotically. In order to obtain the same limiting law for the bootstrap test statistic as the stochastic upper bound $\inf_{\theta^0\in\Theta_I\cap\Theta_R} S_N\left(\theta^0\right)$ under the null, we must ensure that the minimum is achieved in the bootstrap world for some $\theta\in\Theta_I\cap\Theta_R$

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

when the null hypothesis holds true. Fortunately, this can be achieved by adding a suitable penalty term to the bootstrap minimization objective function, yielding the following bootstrap test statistic:

$$\hat{S}_N^* = \min_{\theta \in \Theta_N \cap \Theta_R} \left[ S_N^*(\theta) + \mu_N \frac{S_N(\theta)}{N} \right], \tag{3.20}$$

where $P_N(\theta) \equiv \frac{1}{N} S_N(\theta)$ is a penalty term that ensures the minimum is achieved asymptotically for $\theta \in \Theta_I \cap \Theta_R$ under the null, and $\mu_N$ is a tuning parameter that diverges to infinity at a suitable rate (see Assumption 3.5). As a result, $\hat{S}_N^*$ shares the same limiting distribution as $\inf_{\theta^0 \in \Theta_I \cap \Theta_R} S_N(\theta^0)$ under the null. This ensures the first goal mentioned previously.

To ensure the good power properties of the bootstrap test, we require that $\hat{S}_N^*$ be well behaved under the alternative. When $\mu_N$ diverges to infinity at a rate slower than $N/\log(\log(N))$, we will show that $\mu_N^{-1}\hat{S}_N^* \xrightarrow{p} \min_{\theta \in \Theta \cap \Theta_R} \sum_{s=1}^{T-R} \mathbb{E}\left[\left\{\text{MDD}\left[m_s(Y,X,\theta)|\underline{z}_s\right]\right\}^2\right]$ under $\mathbb{H}_1 : \Theta_I \cap \Theta_R = \varnothing$. That is, $\hat{S}_N^*$ diverges to infinity at rate $\mu_N$, which is slower than the rate $N$ at which $\hat{S}_N$ diverges to infinity under the alternative. This implies that $\hat{S}_N \gg \hat{S}_N^*$ with probability approaching one (w.p.a.1) under the alternative, ensuring the second aforementioned goal. As we show in the simulation study, setting $\mu = N^{1/4}$ or $N^{1/3}$ provides reasonably good size control and satisfactory power performance at a moderate sample size ($N = 500$). This suggests that, in practice, one may want to pick $\mu_N$ to be noticeably slower than $N/\log(\log(N))$.

To proceed, we add the following assumption on $\{v_i\}_{i=1}^N$ and the tuning parameter $\mu_N$.

**Assumption 3.5.** (i) $\{v_i\}_{i=1}^N$ is i.i.d. with mean zero and variance one, and is independent of $\{(Y_i, X_i, Z_i)\}_{i=1}^N$.

(ii) As $N \to \infty$, $\mu_N \to \infty$ and $\mu_N = o(N/\log(\log(N)))$.

The following theorem states the asymptotic properties of $\hat{S}_N^*$ when the null hypothesis holds true or is violated.

THEOREM 3.5 (Consistency of the multiplier bootstrap). *Let Assumption 2.1, 2.2, and 3.1–3.5 hold. If $\mathbb{H}_0 : \Theta_I \cap \Theta_R \neq \varnothing$ holds true, then*

$$\hat{S}_N^* \xrightarrow{\mathcal{L}} \inf_{\theta \in \Theta_I \cap \Theta_R} \sum_{s=1}^{T-R} \left[\mathbb{B}_s(\theta) + \mathbb{C}_s(\theta)\right]. \tag{3.21}$$

*And if $\mathbb{H}_1 : \Theta_I \cap \Theta_R = \varnothing$ holds true, then*

$$\mu_N^{-1}\hat{S}_N^* \xrightarrow{p} \min_{\theta \in \Theta \cap \Theta_R} \sum_{s=1}^{T-R} \text{MDD}\left[m_s(Y,X,\theta)|\underline{z}_s\right]^2. \tag{3.22}$$

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

Theorem 3.5 shows that, with a properly chosen sequence of $\{\mu_N\}$, $\hat{S}_N^*$ converges weakly to the same asymptotic distribution as the stochastic upper bound of the original test statistic $\hat{S}_N$ does under the null hypothesis. This ensures the asymptotic level of the multiplier-bootstrap-based test despite its conservative nature. Now, we explain the intuition. Since $S_N^*(\theta)$ is properly centered on the entire $\Theta$, it can be shown that $S_N^*(\theta) \Longrightarrow \sum_{s=1}^{T-R}[\mathbb{B}_s(\theta) + \mathbb{C}_s(\theta)]$ on $L^\infty(\Theta)$. With the help of the weighted penalty term $\mu_N \frac{S_N(\theta)}{N}$ in (3.20), which asymptotically ensures the involved minimization to be focused on $\Theta_I \cap \Theta_R$, we have the result in (3.21). It is helpful to note that a counterpart of the term $c_N$ in $\hat{S}_N$ (shown in Theorem 3.3) does not appear in $\hat{S}_N^*$, for the following reasons: (i) Asymptotically, $c_N$ accounts for the local deviations of $S_N(\theta)$ from $\sum_{s=1}^{T-R}[\mathbb{B}_s(\theta) + \mathbb{C}_s(\theta)]$ on $L^\infty(\Theta_I)$ for $\theta \notin \Theta_I \cap \Theta_R$ (but still within a shrinking neighborhood of $\Theta_I \cap \Theta_R$). (ii) The deviations described in (i) occur due to the fact that $S_N(\theta)$ is not properly centered on $\Theta \backslash \Theta_I$. (iii) Unlike $S_N(\theta)$, $S_N^*(\theta)$ is properly centered on not only $\Theta_I$ but also $\Theta \backslash \Theta_I$, and there is no asymptotic deviation of $S_N^*(\theta)$ from $\sum_{s=1}^{T-R}[\mathbb{B}_s(\theta) + \mathbb{C}_s(\theta)]$ for $\theta \notin \Theta_I \cap \Theta_R$. Consequently, a counterpart of $c_N$ does not appear in $\hat{S}_N^*$. This explains why, under the null, $\hat{S}_N^*$ asymptotically mimics the distributional upper bound, rather than the exact distribution, of $\hat{S}_N$.

Theorem 3.5 also shows that, under any fixed alternative, $\hat{S}_N^*$ diverges to infinity at rate-$\mu_N$, which is slower than rate-$N$ at which $\hat{S}_N$ diverges to infinity.[6] Therefore, the proposed bootstrap procedure has asymptotic power one against any fixed alternative.

An essential step in the proof of Theorem 3.5 is the study of the unconditional central limit theorem (CLT) of $S_N^*(\theta)$ by applying the results of Arcones and Giné (1993), which is analogous to the unconditional multiplier CLT for empirical processes studied in van der Vaart and Wellner (1996, pp. 177–181) and Kororok (2008, pp. 181–183). It extends the unconditional multiplier CLT for degenerate second-order $U$-statistics in Leucht and Neumann (2013) to degenerate second-order $U$-processes. It is worth mentioning that bootstrap consistency is usually validated by the weak convergence of conditional laws given the data. Specifically, a bootstrap scheme can be considered consistent if an appropriate distance between the conditional distribution of bootstrap replicate of a statistic $S_{NT}$ given the data and the unconditional distribution of $S_{NT}$ is shown to converge to zero in probability. But as Bücher and Kojadinovic (2009) notice, the conditional distribution may not be easy to establish and it is useful to consider an equivalent formulation. In particular, they show that under minimal conditions, the aforementioned convergence of the conditional laws is actually equivalent to the unconditional weak convergence of $S_{NT}$ with two bootstrap replicates to independent copies of the same limit. A close examination of the proof of Theorem 3.5 suggests that $\hat{S}_N^*$ is asymptotically independent of $\hat{S}_N$ under the null and different $\hat{S}_N^*$'s based

[6]Recall that if $\mathbb{H}_1 : \Theta_I \cap \Theta_R = \emptyset$ holds true, then $\min_{\theta \in \Theta \cap \Theta_R} \sum_{s=1}^{T-R}\left[m_s(Y, X, \theta)\,|_{\underline{z}_s}\right]^2 > 0$, which has been established in Theorem 3.4.

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

on different independent sequences $\{v_i\}_{i=1}^N$ are also asymptotically independent of each other. Lemma 4.2 in Bücher and Kojadinovic (2009) further ensures the asymptotic validity to use the conditional quantiles of the empirical distribution of a sample of bootstrap replicates based on the equivalent unconditional formulation. This suggests that in practice, we can draw $\{v_i\}_{i=1}^N$ $B$ times independently from suitable distributions to construct $B$ bootstrap test statistics $\{\hat{S}_N^{*(b)}\}_{b=1}^B$. Then we can calculate the bootstrap $p$-value for our test statistic $\hat{S}_N$ as $p^* = \frac{1}{B}\sum_{b=1}^B \mathbf{1}\{\hat{S}_N \leq \hat{S}_N^{*(b)}\}$ with $\mathbf{1}\{\cdot\}$ being the usual indicator function, and reject the null hypothesis when $p^*$ is smaller than the prescribed level of significance.

## 4. MONTE CARLO SIMULATIONS

In this section, we conduct Monte Carlo simulations to evaluate the finite sample performance of our proposed inference method.

### 4.1. Design 1

We first consider testing for linearity. We test the null hypothesis

$$\mathbb{H}_{0,L} : \Theta_I \cap \Theta_{R_L} \neq \varnothing, \quad \Theta_{R_L} \equiv \left\{\theta \in \Theta : g(x) = a + bx \text{ for some } (a,b) \in \mathbb{R}^2\right\}. \tag{4.1}$$

We adopt a series of DGPs, indexed by $\gamma$, with $T = 3$ and $R = 2$ as follows:

$$y_{it} = x_{it} + \gamma x_{it}^2 + \lambda_i' F_t + u_{it}, \tag{4.2}$$

$$x_{it} = 0.25\lambda_i' F_t + 0.8 z_{it}^* + 0.8 u_{it} + \varepsilon_{it}^*, \tag{4.3}$$

$$z_{it} = z_{it}^*, \tag{4.4}$$

with

$$\begin{pmatrix} z_{it}^* \\ \varepsilon_{it}^* \\ u_{it} \end{pmatrix} \sim \mathrm{N}\left(\mathbf{0}, \begin{bmatrix} 1.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 0.5 \end{bmatrix}\right), \ F = \begin{pmatrix} 0.7 & 0.2 \\ -1 & 0 \\ 0 & -1 \end{pmatrix}, \text{ and } \lambda_i \sim \mathrm{N}\left(0, \begin{bmatrix} 0.5 & 0.0 \\ 0.0 & 0.5 \end{bmatrix}\right),$$

where $\lambda_i$ is independent of $\left(z_{it}^*, \varepsilon_{it}^*, u_{it}\right)'$.

For a given $\gamma$, we label the above DGP as DGP($\gamma$). Under DGP($\gamma$), the null in (4.1) is true or false, depending on the value of $\gamma$. The null is true under DGP(0), and is false under DGP($\gamma$) at any given $\gamma \neq 0$. And $\gamma$ can be viewed as a measure of how far the DGP is away from a linear specification. We conduct 500 bootstrap evaluations to calculate the bootstrap $p$-values. The Monte Carlo study consists of 500 replications. In Table 1, we report the simulated probabilities of rejecting the null in (4.1) under a nominal size of 0.05 for DGP($\gamma$) with $\gamma = 0, \pm 0.5,$ and $\pm 1$, and for $N = 500$ and $1{,}000$. Note that rows with $\gamma = 0$ show size, while all other rows show power. From Table 1, we have the following observations: (i) The choice $\mu_N = 0$ (i.e., no penalty in the bootstraps) leads to severe size distortions, as expected from our theory, and is considered to illustrate the extreme case of

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

**TABLE 1.** Size and power performance for Design 1, nominal size = 0.05.

| $\gamma$ | $\mu_N$ | $N=500$ | | $N=1,000$ | |
|---|---|---|---|---|---|
| | | 2-pt | Norm | 2-pt | Norm |
| 0 | 0 | 0.066 | 0.062 | 0.078 | 0.076 |
| 0 | $N^{1/4}$ | 0.042 | 0.040 | 0.052 | 0.048 |
| 0 | $N^{1/3}$ | 0.034 | 0.036 | 0.046 | 0.044 |
| 0 | $N^{1/2}$ | 0.024 | 0.028 | 0.034 | 0.036 |
| 0.5 | $N^{1/4}$ | 0.334 | 0.346 | 0.778 | 0.778 |
| 0.5 | $N^{1/3}$ | 0.308 | 0.312 | 0.762 | 0.760 |
| 0.5 | $N^{1/2}$ | 0.278 | 0.280 | 0.720 | 0.704 |
| −0.5 | $N^{1/4}$ | 0.436 | 0.476 | 0.834 | 0.826 |
| −0.5 | $N^{1/3}$ | 0.402 | 0.434 | 0.806 | 0.808 |
| −0.5 | $N^{1/2}$ | 0.366 | 0.406 | 0.784 | 0.790 |
| 1.0 | $N^{1/4}$ | 0.570 | 0.582 | 0.962 | 0.966 |
| 1.0 | $N^{1/3}$ | 0.518 | 0.506 | 0.938 | 0.940 |
| 1.0 | $N^{1/2}$ | 0.468 | 0.468 | 0.914 | 0.908 |
| −1.0 | $N^{1/4}$ | 0.604 | 0.612 | 0.968 | 0.966 |
| −1.0 | $N^{1/3}$ | 0.592 | 0.598 | 0.946 | 0.948 |
| −1.0 | $N^{1/2}$ | 0.552 | 0.562 | 0.898 | 0.908 |

*Note*: Rows with $\gamma = 0$ show size and all other rows show power; "2-pt" and "Norm" refer to two-point and standard normal distributions, respectively.

selecting too small a penalty weight. (ii) The size is not sensitive to the choice of disturbance distribution. (iii) The size is somewhat sensitive to the choice of $\mu_N$. Overall, choices of $\mu_N = N^{1/4}$ and $N^{1/3}$ provide good size control under the current design. (iv) The rejection probabilities increase noticeably as $\gamma$ deviates away from zero for all choices of $\mu_N$, and as the sample size increases. This indicates good power performance.

## 4.2. Design 2

Previously in Design 1, we allow the IFEs to correlate with the covariates $x_{it}$. Here, we modify Design 1 to allow the IFEs to correlate with the instruments $z_{it}$ as well, to which case our proposed method is applicable. Specifically, instead of (4.4), we now generate $z_{it} = z_{it}^* + 0.25\lambda_i' F_t$, while keeping everything else unchanged from the previous design. We still consider the null hypothesis specified in (4.1).

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

**TABLE 2.** Size and power performance for Design 2, nominal size = 0.05

| | | $N = 500$ | | $N = 1,000$ | |
|---|---|---|---|---|---|
| $\gamma$ | $\mu_N$ | 2-pt | Norm | 2-pt | Norm |
| 0 | 0 | 0.070 | 0.066 | 0.086 | 0.090 |
| 0 | $N^{1/4}$ | 0.042 | 0.042 | 0.050 | 0.054 |
| 0 | $N^{1/3}$ | 0.038 | 0.036 | 0.046 | 0.048 |
| 0 | $N^{1/2}$ | 0.026 | 0.024 | 0.038 | 0.036 |
| 0.5 | $N^{1/4}$ | 0.406 | 0.390 | 0.756 | 0.768 |
| 0.5 | $N^{1/3}$ | 0.368 | 0.354 | 0.738 | 0.752 |
| 0.5 | $N^{1/2}$ | 0.304 | 0.302 | 0.708 | 0.716 |
| −0.5 | $N^{1/4}$ | 0.494 | 0.498 | 0.794 | 0.806 |
| −0.5 | $N^{1/3}$ | 0.470 | 0.468 | 0.770 | 0.784 |
| −0.5 | $N^{1/2}$ | 0.412 | 0.406 | 0.738 | 0.752 |
| 1.0 | $N^{1/4}$ | 0.616 | 0.602 | 0.948 | 0.960 |
| 1.0 | $N^{1/3}$ | 0.592 | 0.570 | 0.908 | 0.922 |
| 1.0 | $N^{1/2}$ | 0.528 | 0.516 | 0.886 | 0.902 |
| −1.0 | $N^{1/4}$ | 0.628 | 0.620 | 0.956 | 0.958 |
| −1.0 | $N^{1/3}$ | 0.606 | 0.604 | 0.938 | 0.944 |
| −1.0 | $N^{1/2}$ | 0.564 | 0.566 | 0.902 | 0.896 |

*Note*: Rows with $\gamma = 0$ show size and all other rows show power; "2-pt" and "Norm" refer to two-point and standard normal distributions, respectively.

As in Design 1, we conduct 500 bootstrap evaluations to calculate the bootstrap $p$-values. The Monte Carlo study consists of 500 replications. In Table 2, we report the simulated probabilities of rejecting the null in (4.1) under a nominal size of 0.05, with $\gamma = 0, \pm 0.5$, and $\pm 1$, and for $N = 500$ and $1,000$ to show finite sample size ($\gamma = 0$) and power ($\gamma \neq 0$) performance. From these results, we obtain observations very similar to those from Table 1.

### 4.3. Design 3

Now, we consider testing for a quadratic specification. We test the null hypothesis

$$\mathbb{H}_{0,Q} : \Theta_I \cap \Theta_{R_L} \neq \varnothing, \quad \Theta_{R_Q} \equiv \left\{ \theta \in \Theta : g(x) = a + bx + cx^2 \text{ for some } (a,b,c) \in \mathbb{R}^3 \right\}. \tag{4.5}$$

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

**TABLE 3.** Size and power performance for Design 3, nominal size = 0.05

| | | $N = 500$ | | $N = 1,000$ | |
|---|---|---|---|---|---|
| $\gamma$ | $\mu_N$ | 2-pt | Norm | 2-pt | Norm |
| 0 | 0 | 0.054 | 0.048 | 0.060 | 0.058 |
| 0 | $N^{1/4}$ | 0.034 | 0.032 | 0.046 | 0.048 |
| 0 | $N^{1/3}$ | 0.028 | 0.028 | 0.042 | 0.042 |
| 0 | $N^{1/2}$ | 0.020 | 0.018 | 0.034 | 0.032 |
| 1 | $N^{1/4}$ | 0.276 | 0.286 | 0.656 | 0.658 |
| 1 | $N^{1/3}$ | 0.246 | 0.252 | 0.630 | 0.632 |
| 1 | $N^{1/2}$ | 0.188 | 0.206 | 0.596 | 0.590 |
| −1 | $N^{1/4}$ | 0.352 | 0.356 | 0.734 | 0.736 |
| −1 | $N^{1/3}$ | 0.324 | 0.326 | 0.712 | 0.718 |
| −1 | $N^{1/2}$ | 0.278 | 0.284 | 0.694 | 0.698 |
| 2 | $N^{1/4}$ | 0.494 | 0.502 | 0.902 | 0.904 |
| 2 | $N^{1/3}$ | 0.456 | 0.460 | 0.876 | 0.880 |
| 2 | $N^{1/2}$ | 0.420 | 0.426 | 0.846 | 0.854 |
| −2 | $N^{1/4}$ | 0.526 | 0.534 | 0.922 | 0.926 |
| −2 | $N^{1/3}$ | 0.508 | 0.512 | 0.896 | 0.904 |
| −2 | $N^{1/2}$ | 0.478 | 0.482 | 0.864 | 0.876 |

*Note*: Rows with $\gamma = 0$ show size and all other rows show power; "2-pt" and "Norm" refer to two-point and standard normal distributions, respectively.

The DGPs we adopt here are modified from those in Design 2. Specifically, we replace (4.2) by

$$y_{it} = x_{it} + x_{it}^2 + \gamma h(x_{it}) + \lambda_i' F_t + u_{it},$$

where $h(x) = \exp\left(-2x^2\right)/\sqrt{0.02\pi}$, while keeping everything else unchanged from Design 2. (Consequently, the IFEs are correlated with both $x_{it}$ and $z_{it}$, like in Design 2.) In Table 3, we report the simulated probabilities of rejecting the null in (4.5) under a nominal size of 0.05, for $\gamma = 0, \pm 1,$ and $\pm 2$. From Table 3, we have the following observations: (i) A sample size of 500 seems to be somewhat too small for this design, as indicated by the simulated sizes being noticeably below the targeted size of 0.05 at $N = 500$. (ii) As we increase the sample size to $N = 1,000$, the simulated sizes become reasonably close to 0.05 with $\mu_N = N^{1/4}$ and $N^{1/3}$. (iii) The rejection probabilities increase noticeably as $\gamma$ deviates away from zero, for all choices of $\mu_N$ and both $N = 500$ and $1,000$. This indicates good power performance. (iv) Similar to what we observe in the previous designs, $\mu_N = N^{1/2}$

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

seems to be a bit too large for the penalty weight, which leads to severe undersizing even at $N = 1,000$ for the current design. Nevertheless, $\mu_N = N^{1/2}$ still provides good power performance.

## 5. EMPIRICAL APPLICATION

In this section, we apply our method to study Engel curves for major nondurable expenditures in China, using data from the CFPS for the period 2010 to 2014. The CFPS is similar to the U.K. Family Expenditure Survey (UKFES), but is conducted every other year. Specifically, the CFPS data are collected in 2010, 2012, and 2014, which produces a three-period ($T = 3$) balanced panel data set of $6,627$ households.

According to the panel data from CFPS and its consumption categorization, on average, food (including dining) expenditures take the largest share of total nondurable expenditures (averaging at 41.54%), which is followed by medical and health care expenditures (averaging at 12.01%), expenditures on commuting and communication (averaging at 9.99%), and grocery expenditures (averaging at 9.63%). We study the Engel curves for these four major categories of consumption.

For household $i$ at period $t$, let $y_{f,it}$, $y_{m,it}$, $y_{c,it}$, $y_{g,it}$ be the share of its total nondurable expenditures spent on food including dining (FD), medical and health care (MH), commuting and communication (CC), grocery (GY), respectively. Let $x_{it}$ be the log of total annual nondurable expenditures, and let $z_{it}$ be total household annual income. The Engel curves are assumed to take the following additive form:

$$y_{j,it} = g_j(x_{it}) + \lambda_{j,i}' F_{j,t} + u_{j,it}, \tag{5.1}$$

for $t \in \{1,2,3\}$ and $j \in \{f,m,c,g\}$. $F_{j,t}$, $\lambda_{j,i}$, and $u_{j,it}$ are unobservable terms, representing the vector of factors, the vector of factor loadings, and other heterogeneity, respectively. Some observations are dropped because they have key variables out of reasonable range.[7] For studying Engel curves for FD and MH, we also condition on households who exhibited positive consumption of both categories. These restrictions yield a three-period panel of $3,811$ cross-sectional observations for FD and MH (consisting of $1,787$ urban households and $2,024$ rural households). Similarly, for studying Engel curves for CC and GY, we condition on households who exhibited positive consumption of both categories, which yields a three-period panel of $4,584$ cross-sectional observations for CC and GY (consisting of $2,287$ urban households and $2,297$ rural households).

We set the support for $x_{it}$ as $\mathcal{X} = [7,14]$, which includes all observations for all four categories of consumption. Since $\mathcal{X}$ is compact, we set $\zeta_0 = 0$ (i.e., no tail control needed). For nonparametric inferences, we employ B-spline of order 3 on $[7,14]$, with 2 interior knots which are the 33.3th percentile and the 66.7th

[7] We drop observations with extremely low total annual expenditures (≤2,188 CNY, which corresponds to the 0.01 quantile of total expenditures distribution) or extremely low total household annual income (≤3,000 CNY, which equals the poverty line per capita set by the State Council Leading Group Office of Poverty Alleviation and Development of China in 2015)

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

percentile of the corresponding $x_{it}$ sample. Finally, we conduct hypothesis tests with $\mu_N = N^{1/4}$ and $\mu_N = N^{1/3}$ for constructing the corresponding bootstrap statistics as suggested by our Monte Carlo simulations.

### 5.1. Testing for Log-Linearity

Since the seminal work of Deaton and Muellbauer (1980), a log-linear specification (i.e., linear in the log of total nondurable expenditures) has been commonly adopted to parameterize Engel curves in the literature. The use of log-linear Engel curves to estimate and correct bias in directly measured macroeconomic indicators is very prevalent in empirical studies. For instance, a popular way to construct an alternative measure of household income or expenditure relies on the log-linear Engel curves as a key assumption to infer incomes or expenditures (see Pissarides and Weber, 1989; Browning and Crossly, 2009; Aguiar and Bils, 2015; Hurst, Li, and Pugsley (2014), among others). Besides, there are papers focusing on estimating CPI bias based on the log-linear form of Engel curves (see Hamilton, 2001; Nakamura, Steinsson, and Liu, 2016, among others). On the other hand, there are also papers advocating advantages of building nonparametric Engel curves over the parametric ones for studies of demand (see, e.g., Blundell, Browning, and Crawford, 2003; Blundell, Chen, and Kristensen, 2007).

In our empirical study, we first examine whether the log-linear relationship could adequately describe the Engel curves for major nondurable expenditures in China. Under a potential lack of point-identification, the linear specification can be tested through the hypothesis specified in (4.1) that we previously studied in our Monte Carlo simulations, that is,

$$\mathbb{H}_0 : \Theta_I \cap \Theta_{R_L} \neq \emptyset, \quad \Theta_{R_L} \equiv \left\{\theta \in \Theta : g(x) = a + bx \text{ for some } (a,b) \in \mathbb{R}^2\right\}. \tag{5.2}$$

To account for potential differences in consumption pattern/habit between urban and rural households, we conduct tests of the above hypotheses using the whole sample, the urban subsample, and the rural subsample, respectively. To take into account the issue of multiple hypotheses testing, we also consider the method of Benjamini and Hochberg (1995, BH hereafter) to control the false discovery rate (FDR). Suppose that we consider testing $m$ hypotheses, and we order the individual $p$-values from the smallest to the largest as $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(m)}$ with their corresponding null hypotheses labeled accordingly as $\mathbb{H}_{0(1)}, \mathbb{H}_{0(2)}, \ldots, \mathbb{H}_{0(n)}$. The BH-adjusted $p$ -values for testing $\mathbb{H}_{0(k)}$ is then given by $\min(p_{(k)} \frac{m}{k}, 1)$.

In Table 4, we report the usual bootstrap $p$-values of corresponding test statistics, based on 1,000 bootstrap repetitions, for each of the four consumption categories (FD, MH, CC, and GY), and for specifying the number of factors as $R = 1$ or 2. Each of these $p$-values is computed based on a separate test of (5.2) regarding a specific Engel curve (of FD, MH, CC, or GY) under a specific (sub)sample (i.e., all, urban, or rural). To control the FDR, we also report the BH-adjusted $p$-values by taking into account the multiple testing issue for testing the four consumption

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

**Table 4.** Bootstrap $p$-values for testing the log-linear null for the Engel curves.

| Cat. | $\mu_N$ | No. of factors $R=1$ | | | | | | No. of factors $R=2$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | All | | Urban | | Rural | | All | | Urban | | Rural | |
| | | $p$ | BH-$p$ | $p$ | BH-$p$ | $p$ | BH-$p$ | $p$ | BH-$p$ | $p$ | BH-$p$ | $p$ | BH-$p$ |
| FD | $N^{1/4}$ | 0.002 | 0.003 | 0.001 | 0.004 | 0.887 | 0.887 | 0.530 | 0.707 | 0.456 | 0.912 | 0.930 | 0.930 |
| MH | $N^{1/4}$ | 0.009 | 0.009 | 0.639 | 0.639 | 0.158 | 0.211 | 0.261 | 0.522 | 0.741 | 0.988 | 0.255 | 0.510 |
| CC | $N^{1/4}$ | 0.000 | 0.000 | 0.003 | 0.006 | 0.000 | 0.000 | 0.065 | 0.260 | 0.421 | 1 | 0.185 | 0.740 |
| GY | $N^{1/4}$ | 0.000 | 0.000 | 0.016 | 0.021 | 0.009 | 0.018 | 0.784 | 0.784 | 0.855 | 0.855 | 0.810 | 1 |
| FD | $N^{1/3}$ | 0.005 | 0.007 | 0.001 | 0.004 | 0.890 | 0.890 | 0.635 | 0.847 | 0.490 | 0.980 | 0.960 | 0.960 |
| MH | $N^{1/3}$ | 0.010 | 0.010 | 0.741 | 0.741 | 0.163 | 0.217 | 0.273 | 0.546 | 0.796 | 1 | 0.400 | 0.800 |
| CC | $N^{1/3}$ | 0.000 | 0.000 | 0.006 | 0.012 | 0.002 | 0.008 | 0.077 | 0.308 | 0.465 | 1 | 0.230 | 0.920 |
| GY | $N^{1/3}$ | 0.002 | 0.004 | 0.028 | 0.037 | 0.012 | 0.024 | 0.808 | 0.808 | 0.872 | 0.872 | 0.790 | 1 |

*Note*: FD, MH, CC, and GY abbreviate food, medical and heath care, commuting and communication, and grocery, respectively. $p$ and BH-$p$ denote the usual $p$-value and BH-adjusted $p$-value, respectively.

categories. According to Table 4, when setting $R=1$ and applying the conventional 5% significance level, we obtain the following testing results: (i) For FD, our tests reject the null except for the rural subsample. (ii) For MH, interestingly, our tests fail to reject the null when conditioning on either urban or rural households, yet are able to reject the null for the whole sample (i.e., without conditioning on urban or rural households). (iii) For both CC and GY, our tests reject the null regardless of whether we condition on urban households, rural households, or not. (iv) When the multiple testing issue is accounted for, the findings in (i)–(iii) are also true. We obtain the same testing results regardless of whether $\mu_N = N^{1/4}$ or $N^{1/3}$, and the $p$-values seem to be insensitive to these two choices of $\mu_N$. In short, these results suggest nonlinearity for some Engel curves and noticeable difference in consumption pattern/habit on FD and MH between urban and rural households.

Also according to Table 4, when setting $R=2$, however, we fail to reject the null of log-linear specification for the Engel curves for all cases under our investigation, regardless of whether one controls the FDR or not. A possible explanation is that the IFE terms encompass unobserved heterogeneity of more flexible forms under $R=2$ than that under $R=1$; and consequently, under $R=2$, the log-linear relationship might suffice to adequately describe the Engel curves, with heterogeneity being more flexibly taken care of by the IFEs. As noted by Santos (2012), failing to reject the null that there are log-linear Engel curves in the identified set does not necessarily justify adopting such a parametric specification. If the model is partially identified, then even when log-linear specifications are indeed in $\Theta_I$, there is no guarantee that the true model is one of them. Therefore, confidence intervals constructed under the log-linear assumption may asymptotically exclude the true parameter of interest.

### 5.2. Confidence Interval for $g(\bar{x})$

Next, we examine the robustness of our testing results regarding the log-linear specification by comparing 95% confidence intervals for food Engel curves at the sample average (across both $N$ and $T$) $\bar{x}$ with and without assuming log-linearity and by setting $R = 1$ or 2 and $\mu_N = N^{1/3}$. Define

$$\Theta_{R_\gamma} \equiv \{\theta \in \Theta : g(\bar{x}) = \gamma\}. \tag{5.3}$$

We obtain the confidence intervals for $g(\bar{x})$ under the log-linear specification by inverting tests of a series of null hypotheses, indexed by $\gamma$, defined as follows:

$$\mathbb{H}_{0,\gamma} : \Theta_I \cap \left(\Theta_{R_L} \cap \Theta_{R_\gamma}\right) \neq \varnothing, \tag{5.4}$$

where $\Theta_{R_L}$ is defined in (5.2). We also obtain the nonparametric confidence intervals $g(\bar{x})$ by inverting tests of the following null hypotheses:

$$\mathbb{H}_{0,\gamma} : \Theta_I \cap \Theta_{R_\gamma} \neq \varnothing, \tag{5.5}$$

where $\Theta_{R_\gamma}$ is defined by (5.3). For a given $\gamma \in [0,1]$, test of either (5.4) or (5.5) is conducted with 200 bootstrap repetitions.[8] To test for the nonparametric null hypothesis in (5.5), we need to set $C_g$. We find our results insensitive to its value as long as it is not too small. Here, we report the results under $C_g = 15$. For comparison purpose, we also construct confidence intervals using the method developed by Santos (2012), as well as standard IV confidence intervals. To make Santos' (2012) method applicable here, we pool the panel into a large cross-section data set, ignoring any potential fixed effects (which effectively turns into a situation with $T = 1$ and $R = 0$). Standard IV confidence intervals are constructed based on the pooled data set, too. To make a direct comparison between our method and Santos' (2012), we also construct confidence intervals based on the pooled data set using our method. All these confidence intervals are reported in Table 5. Here, we mainly focus on results from the urban subsample for a detailed discussion. According to Table 5, for food consumption by urban households, while the log-linear confidence interval based on Santos (2012) is somewhat larger than the standard IV one, we end up with an empty set constructing log-linear confidence interval based on the pooled data set using our method. Moreover, based on the original panel data set, when setting $R < 2$, we obtain empty confidence intervals even under the nonparametric specification. Conducting further tests (to be discussed in the next subsection) on specifications using our method confirms our finding of these empty sets: for food consumption by urban households, based on the pooled data set, our test rejects the null of a log-linear specification; based on the original panel data set, our test rejects both the null of a log-linear specification and that of a nonparametric specification when setting $R < 2$.[9] In the

[8] Note that the construction of each confidence interval requires constructing a series of tests over a grid of $\gamma \in [0,1]$, which can be computationally heavy, especially under the nonparametric specification. So we pick a relative moderate number of bootstrap repetitions here to avoid overly high computational cost.

[9] When $R = 0$, one can continue to implement our testing procedure in the absence of the nuisance parameter $\phi$.

**Table 5.** 95% confidence intervals of $g(\bar{x})$ for the food Engel curves.

| | All | Urban | Rural |
|---|---|---|---|
| $R = 2$, nonparametric | [0.000, 1.000] | [0.000, 1.000] | [0.000, 1.000] |
| $R = 2$, log-linear | [0.000, 1.000] | [0.110, 0.890] | [0.000, 1.000] |
| $R = 1$, nonparametric | Empty | Empty | [0.308, 0.633] |
| $R = 1$, log-linear | Empty | Empty | [0.390, 0.585] |
| $R = 0$, nonparametric | Empty | Empty | Empty |
| $R = 0$, log-linear | Empty | Empty | Empty |
| Pooled, nonparametric | [0.325, 0.530] | [0.387, 0.550] | [0.258, 0.511] |
| Pooled, log-linear | [0.406, 0.412] | Empty | [0.400, 0.410] |
| Pooled, nonpara., Santos | [0.119, 0.810] | [0.252, 0.712] | [0.000, 0.791] |
| Pooled, log-linear, Santos | [0.378, 0.440] | [0.366, 0.533] | [0.377, 0423] |
| Pooled, standard IV | [0.411, 0.418] | [0.419, 0.429] | [0.401, 0.411] |

*Note*: Given $R$, a log-linear CI for $g(\bar{x})$ is empty if the null $\mathbb{H}_{0,\gamma} : \Theta_I \cap \Theta_{R_L} \neq \emptyset$ is rejected at 5% level. Similarly, a nonparametric CI for $g(\bar{x})$ is empty if the null $\mathbb{H}_{0,\gamma} : \Theta_I \cap \Theta_R \neq \emptyset$ is rejected at 5% level.

next subsection, we focus on the joint specification of functional forms and the number of factors.

Interestingly, several confidence intervals are given by [0, 1] in Table 5 under $R = 2$. This suggests that the inference does not provide any informative/powerful results regarding $g(\bar{x})$ when we allow for both a very flexible heterogeneity specification ($R = 2$) and a lack of point-identification. Overall, Table 5, in particular, its first six rows, can be interpreted as results from a sensitivity analysis which demonstrates how the degree of informativeness/powerfulness of our inference procedure varies in response to changes in the strength of the model assumptions. These results pretty much reflect the law of decreasing credibility as coined by Manski (2003). Stronger assumptions yield inferences that may be more powerful but less credible, which is a dilemma faced by empirical researchers as they decide what assumption to maintain. Here, we also take Manski's (2003) view that statistical theory cannot resolve the dilemma but can clarify its nature. That being said, Table 5 suggests our inference procedure still provides informative results under the already quite general setting of $R = 1$, nonparametric $g(\cdot)$ and partial identification.

## 5.3. Further Investigation on Heterogeneity and Specification

We further investigate the specification of the functional form of $g(\cdot)$ and the number of factors ($R$), again focusing on food consumption. Ignoring the index

$j$ for different expenditure categories, we can rewrite the Engel curves in (5.1) as

$$y_{it} = g_{it}(x_{it}) + u_{it}, \quad \textbf{(5.6)}$$

$$g_{it}(x_{it}) = g(x_{it}) + \lambda_i' F_t, \quad \textbf{(5.7)}$$

where $t = 1,2,3$, the subscripts on $g_{it}(\cdot)$ capture potential heterogeneity cross individual and time, while (5.7) assumes the heterogeneity to take a special form of a common part $g(\cdot)$ augmented with an additive IFE term. A specification under the (5.6) and (5.7) framework is characterized by the combination of two specifications: (i) the functional form specification on $g(\cdot)$, and (ii) the specification on the number of factors $R$. The less restrictive a functional form specification on $g(\cdot)$, the more flexible/general the model is w.r.t. the common part relationship. The larger the number of factors, the more flexible/general the model is w.r.t. unobserved heterogeneity. It is easy to see that, within the (5.6) and (5.7) framework, the model achieves its maximum flexibility/generality when $g(\cdot)$ is treated nonparametrically and $R$ is set to 2.[10] We use our method to test the specification of a variety of these combinations, based on 500 bootstrap replications.[11]

Both the usual $p$-values and BH-adjusted $p$-values obtained from these tests are reported in Table 6. For the BH-adjusted $p$-values, we consider six hypotheses for the proposed method, two hypotheses for the standard IV method for the pooled data, and two hypotheses for the Santos' (2012) method for the pooled data. As suggested by Table 6, even with a nonparametric specification on $g(\cdot)$, the model does not suffice to adequately describe the Engel curve for food consumption among urban households in China when setting $R < 2$, no matter whether one controls the FDR or not.[12] Interestingly, in comparison, our test fails to reject a log-linear specification on Engel curves among rural households when setting $R = 1$. (Note that our test still rejects a nonparametric specification for the rural population when setting $R = 0$.) In contrast, Banks, Blundell, and Lewbel (1997) pool a panel data set into a cross-sectional one and their study suggests that a log-quadratic specification suffices to adequately describe most Engel curves. Admittedly, such a comparison is only indirect because Banks, Blundell, and Lewbel (1997) use a different data set, one obtained from the UKFES, for their study. Our findings suggest: (i) There is a greater degree of heterogeneity in food consumption patterns among urban households in China. (ii) There is a lesser degree of heterogeneity in food consumption patterns among rural households, compared with that among urban households in China. (iii) Even a nonparametric specification on $g(\cdot)$, as

[10]Recall that the maximum number of factors allowed is $T-1$ for our method to work.

[11]In other words, here, we view these tests as jointly testing for the specification on $R$ and the specification on the functional form of $g(\cdot)$. While for obtaining CI for $g(\bar{x})$, we take a narrow view and interpret corresponding tests as only testing for the specification on the functional form of $g(\cdot)$, assuming any given specification on $R$ to be true.

[12]As mentioned earlier in this section, we employ a specific B-spline (i.e., of order 3 on $[7,14]$ with two interior knots) to approximate $g(\cdot)$ for corresponding tests where $g(\cdot)$ is supposed to be treated nonparametrically. This practice shares the same spirit with many existing nonparametric testing procedures. In an utterly strict sense, what is really tested here is where the specific cubic B-spline with two interior knots is adequate to describe the Engel curve for the given finite sample.

**Table 6.** $p$-Values for testing joint specification on $g(\cdot)$ and $R$

| | All | | Urban | | Rural | |
|---|---|---|---|---|---|---|
| | $p$ | BH-$p$ | $p$ | BH-$p$ | $p$ | BH-$p$ |
| $R = 2$, nonparametric | 0.756 | 0.756 | 0.956 | 0.956 | 0.982 | 0.982 |
| $R = 2$, log-linear | 0.636 | 0.763 | 0.494 | 0.593 | 0.958 | 1 |
| $R = 1$, nonparametric | 0.020 | 0.030 | 0.006 | 0.012 | 0.620 | 1 |
| $R = 1$, log-linear | 0.004 | 0.008 | 0.002 | 0.004 | 0.886 | 1 |
| $R = 0$, nonparametric | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| $R = 0$, log-linear | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Pooled, nonparametric | 0.874 | 0.874 | 0.988 | 0.988 | 0.902 | 0.902 |
| Pooled, log-linear | 0.286 | 0.572 | 0.004 | 0.008 | 0.384 | 0.768 |
| Pooled, nonpara., Santos | 0.756 | 0.756 | 0.930 | 0.930 | 0.220 | 0.220 |
| Pooled, log-linear, Santos | 0.068 | 0.136 | 0.834 | 1 | 0.192 | 0.384 |

*Note*: $p$ and BH-$p$ denote the usual $p$-value and BH-adjusted $p$-value, respectively.

general as it is, might still be insufficient to compensate for an inadequate handling of heterogeneity to make the whole model a correctly specified one, in the case of which a larger $R$ for the additive IFEs, or even nonseparable heterogeneity (in the form of one-way/two-way/interactive effects), may be required to make the model correctly specified. (iv) When a panel data set is available, using methods that fully extract information from the panel structure, such as ours, could potentially provide more informative results than those obtained based on cross-sectional data sets, or based on panel data sets but treated as pooled cross-sectional ones.

## 6. CONCLUSION

In this paper, we propose a statistical inference procedure for partially identified nonparametric panel data models with endogeneity and IFEs. Even though the original identified set is specified through a set of conditional moment restrictions under the weak exogeneity assumption, we are able to translate it into an equivalent set of unconditional moment restrictions by using the novel MDD measure for the distance between a conditional mean object and zero. We construct the test statistic based on such a measure which is associated with a second-order $U$-process in the limit that is degenerate under the null and nondegenerate under the alternative. We establish a tight asymptotic distributional upper bound for the resultant test statistic under the null and show that it is divergent at rate-$N$ under the global alternative. To obtain the critical values for our test, we also propose a version of multiplier bootstrap and establish its asymptotic validity. Simulations show that our test behaves well in finite samples. We apply our method to study Engel curves for several major nondurable expenditures in China by using a panel dataset from CFPS.

The paper can be extended in various directions. First, our panel data model is of a nonparametric nature in the presence of IFEs and we have a single nonparametric object of interest. It is also interesting to consider more general nonparametric panel data models with more than one nonparametric project (e.g., additive models) or semiparametric panel data models with both nonparametric and parametric components that are of interest. Second, it remains unclear how to determine the number of factors in our framework. Difficulty arises because one cannot apply existing methods (e.g., Bai and Ng, 2002; Onatski, 2010; Ahn and Horenstein, 2013; Jin, Miao, and Su, 2021) that are developed under the large $N$ and large $T$ setup to our framework with large $N$ and fixed $T$. Further complication is due to the partial identification nature of nonparametric panels. Third, it is possible to extend the current theoretical framework to conditional moment inequality models through the introduction of some slackness parameter. This will greatly broaden the scope of the current paper. We leave the extensions for future research.

# APPENDIX

The appendix contains the proofs of the main results in the paper. In proving these results, we make use of several lemmas whose proofs can be found in the Supplementary Material.

## A. PROOFS OF THE MAIN RESULTS

Let $\mathrm{MDD}(\varepsilon|W) = \left[\mathrm{MDD}(\varepsilon|W)^2\right]^{1/2}$. To prove the main results, we make use of the following lemmas.

LEMMA A.1. *Let $Z$ be a real random vector s.t. $\mathbb{E}|Z| < \infty$. For any real-valued random variables $W_1$ and $W_2$, if $\mathrm{MDD}(W_1|Z)^2 = 0$ a.s., then $\mathrm{MDD}(W_2 - W_1|Z)^2 = \mathrm{MDD}(W_2|Z)^2$ and $\mathbb{E}[(W_2 - W_1)(W_2^\dagger - W_1^\dagger) \times \left|Z - Z^\dagger\right|] = \mathbb{E}\left[W_2 W_2^\dagger \left|Z - Z^\dagger\right|\right]$, where $W_1^\dagger, W_2^\dagger$, and $Z^\dagger$ are independent copies of $W_1$, $W_2$, and $Z$, respectively.*

LEMMA A.2. *Let $Z$ be a real random vector s.t. $\mathbb{E}|Z| < \infty$. Let $\mathcal{W}$ be a set of real-valued random variables with uniformly bounded second moment, that is, $\sup_{W \in \mathcal{W}} \mathbb{E}\left(W^2\right) < \infty$. Then there exists a finite constant $b$, s.t. for any $W_1, W_2 \in \mathcal{W}$, it holds that*

$$\left|\mathrm{MDD}(W_1|Z)^2 - \mathrm{MDD}(W_2|Z)^2\right| \le b\mathrm{MDD}(W_1 - W_2|Z) \le 2b[\mathrm{MDD}(W_1|Z) + \mathrm{MDD}(W_2|Z)].$$

LEMMA A.3. *The parameter space $\Theta$ is compact under the norm $\|\cdot\|_c$ as defined by (2.9). Consequently, there exists a constant $B_c < \infty$ s.t. for all $g \in \mathcal{G}$, $\sup_{x \in \mathcal{X}} \left|D^\lambda g(x)\right| \le B_c$ for any vector of nonnegative integers $\lambda$ with $\langle\lambda\rangle \le \frac{d}{2}$. In particular, for all $g \in \mathcal{G}$, $\sup_{x \in \mathcal{X}} |g(x)| \le B_c$.*

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

LEMMA A.4. *Let Assumptions 2.1 and 3.1(i) hold. Define* $\mathcal{Q}(\theta) \equiv \sum_{s=1}^{T-R} \mathrm{MDD}[m_s(Y, X, \theta)|\underline{z}_s]^2$. *Then* $\mathcal{Q}(\cdot)$ *is Lipschitz continuous w.r.t.* $\|\cdot\|_{L^2}$ *in* $\Theta$.

LEMMA A.5. *Consider a generic econometric model* $Q(\theta) = 0$, *the identified set of which is characterized by* $\Theta_I \equiv \{\theta \in \Theta : Q(\theta) = 0\}$. *Suppose the following conditions hold: (i)* $Q(\cdot) \geq 0$ *and* $\Theta$ *is compact under (pseudo-)metric* $\mathrm{d}(\cdot,\cdot)$; *(ii)* $\Theta_N \subseteq \Theta$ *are closed and s.t.* $\exists\, \Pi_N\theta$ *for each* $\theta \in \Theta$ *s.t.* $\mathrm{d}(\Pi_N\theta, \theta) = o(1)$ *and* $\sigma_N \equiv \sup_{\theta^0 \in \Theta_I} \mathrm{d}\left(\Pi_N\theta^0, \theta^0\right) = o(1)$; *(iii)* $\sup_{\theta\in\Theta_N} |Q_N(\theta) - Q(\theta)| = O_p(b_N)$ *for some* $b_N = o(1)$; *(iv)* $\exists$ *positive constants* $a_1$ *and* $a_2$ *s.t.* $a_1 \mathrm{d}(\theta, \Theta_I)^2 \leq Q(\theta) \leq a_2 \mathrm{d}(\theta, \Theta_I)^2$. *Then for* $\hat{\theta}_N \in \underset{\theta\in\Theta_N}{\mathrm{argmin}} Q_N(\theta)$, *it holds that* $\mathrm{d}(\hat{\theta}_N, \Theta_I) = O_p\left(\max\{\sigma_N, b_N^{1/2}\}\right)$.

LEMMA A.6. *(i) Let Assumptions 2.1, 2.2, 3.1–3.3(i) hold. For any* $\hat{\theta}_N \in \underset{\theta\in\Theta_N\cap\Theta_R}{\mathrm{argmin}}\ S_N(\theta)$, *it holds that* $\mathrm{d}_{\|\cdot\|_{L^2}}\left(\hat{\theta}_N, \Theta_I \cap \Theta_R\right) = o_p(1)$. *(ii) If, in addition, Assumptions 3.3(ii) and 3.4 hold, then it holds that* $\mathrm{d}_{\|\cdot\|_{L^2}}\left(\hat{\theta}_N, \Theta_I \cap \Theta_R\right) = O_p\left(\varrho_N \mathrm{d}_w\left(\hat{\theta}_N, \Theta_I \cap \Theta_R\right) + \delta_{s,N}\right)$.

**Proof of Lemma 2.1.** Recall from (2.10) that $\Theta_I = \{\theta = \left(\phi', g\right)' \in \boldsymbol{\Phi} \times \mathcal{G} : \mathbb{E}\left[m_s(Y_i, \phi_s, \mathbf{g}(X_i))\,|\underline{z}_{is}\right] = 0$ a.s. for $s = 1, \ldots, T-R\}$. Let

$$\tilde{\Theta}_I \equiv \left\{\theta = (\phi', g)' \in \Theta : \begin{array}{l} \text{For some } R\text{-dimensional random vector } \lambda_i, \text{ it holds} \\ \mathbb{E}\left[y_{it} - g(x_{it}) - \lambda_i'\phi_t|\underline{z}_{it}\right] = 0 \text{ a.s. for } t = 1, \ldots, T-R \\ \mathbb{E}\left[y_{it} - g(x_{it}) - \lambda_i'\left(-\iota_{t-(T-R)}\right)|\underline{z}_{it}\right] = 0 \text{ a.s. for } t = T-R+1, \ldots, T \end{array}\right\}, \tag{A.1}$$

where $\iota_r$ is the $r$th column of the $R \times R$ identity matrix. We need to show that $\Theta_I = \tilde{\Theta}_I$.

For any given $\tilde{\theta} = (\tilde{\phi}', \tilde{g})' \in \tilde{\Theta}_I$, it holds that

$$\begin{pmatrix} \mathbb{E}\left[y_{i1} - \tilde{g}(x_{i1}) - \lambda_i'\tilde{\phi}_1\ |\underline{z}_{i1}\right] \\ \vdots \\ \mathbb{E}\left[y_{i,T-R} - \tilde{g}(x_{i,T-R}) - \lambda_i'\tilde{\phi}_{T-R}\ |\underline{z}_{i,T-R}\right] \\ \mathbb{E}\left[y_{i,T-R+1} - \tilde{g}(x_{i,T-R+1}) - \lambda_i'(-\iota_1)\ |\underline{z}_{i,T-R+1}\right] \\ \vdots \\ \mathbb{E}\left[y_{iT} - \tilde{g}(x_{iT}) - \lambda_i'(-\iota_R)\ |\underline{z}_{iT}\right] \end{pmatrix} = \mathbf{0} \text{ a.s.} \tag{A.2}$$

By (2.6), multiplying both sides of (A.2) by the $(T-R) \times T$ matrix $H(\tilde{\phi})' \equiv (I_{T-R}, \tilde{\Phi})$yields

$$\begin{pmatrix} \mathbb{E}\left[m_1\left(Y_i, \tilde{\phi}_1, \tilde{\mathbf{g}}(X_i)\right)|\underline{z}_{i1}\right] \\ \vdots \\ \mathbb{E}\left[m_{T-R}\left(Y_i, \tilde{\phi}_{T-R}, \tilde{\mathbf{g}}(X_i)\right)|\underline{z}_{i,T-R}\right] \end{pmatrix} = \mathbf{0} \text{ a.s.},$$

which clearly implies that $\tilde{\theta} \in \Theta_I$. Since this holds for any $\tilde{\theta} \in \tilde{\Theta}_I$, it holds that $\tilde{\Theta}_I \subseteq \Theta_I$.

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

Next, for any given $\theta^0 = \left(\phi^{0\prime}, g^0\right)' \in \Theta_I$, it holds that

$$\begin{pmatrix} \mathbb{E}\left[m_1\left(Y_i, \phi_1^0, \mathbf{g}^0(X_i)\right)|\underline{z}_{i1}\right] \\ \vdots \\ \mathbb{E}\left[m_{T-R}\left(Y_i, \phi_{T-R}^0, \mathbf{g}^0(X_i)\right)|\underline{z}_{i,T-R}\right] \end{pmatrix} = \mathbf{0} \text{ a.s.},$$

or, equivalently,

$$\begin{pmatrix} \mathbb{E}\left\{y_{i1} - g^0(x_{i1}) + \sum_{r=1}^R \phi_{1r}^0\left[y_{i,T-R+r} - g^0\left(x_{i,T-R+r}\right)\right]|\underline{z}_{i1}\right\} \\ \vdots \\ \mathbb{E}\left\{y_{i,T-R} - g^0\left(x_{i,T-R}\right) + \sum_{r=1}^R \phi_{T-R,r}^0\left[y_{i,T-R+r} - g^0\left(x_{i,T-R+r}\right)\right]|\underline{z}_{i1}\right\} \end{pmatrix} = \mathbf{0} \text{ a.s.} \quad \textbf{(A.3)}$$

Let $\lambda_i^0 \equiv (y_{i,T-R+1} - g^0\left(x_{i,T-R+1}\right), \ldots, y_{i,T} - g^0\left(x_{i,T}\right))'$. Then by (A.3) and the fact that $\lambda_i^{0\prime}(-\iota_r) = y_{i,T-R+r} - g^0\left(x_{i,T-R+r}\right)$, it holds that

$$\begin{pmatrix} \mathbb{E}\left[y_{i1} - g^0(x_{i1}) - \lambda_i^{0\prime}\phi_1^0 \ |\underline{z}_{i1}\right] \\ \vdots \\ \mathbb{E}\left[y_{i,T-R} - g^0(x_{i,T-R}) - \lambda_i^{0\prime}\phi_{T-R}^0 \ |\underline{z}_{i,T-R}\right] \\ \mathbb{E}\left[y_{i,T-R+1} - g^0(x_{i,T-R+1}) - \lambda_i^{0\prime}(-\iota_1) \ |\underline{z}_{i,T-R+1}\right] \\ \vdots \\ \mathbb{E}\left[y_{iT} - g^0(x_{iT}) - \lambda_i^{0\prime}(-\iota_R) \ |\underline{z}_{iT}\right] \end{pmatrix} = \mathbf{0} \text{ a.s.}, \quad \textbf{(A.4)}$$

which clearly implies that $\theta^0 \in \tilde{\Theta}_I$. Since it holds for any $\theta^0 \in \Theta_I$, it holds that $\Theta_I \subseteq \tilde{\Theta}_I$. This completes the proof of the lemma. □

**Proof of Lemma 3.1.** Note that for any real-valued random variable $W$ and real vector-valued random variable $Z$,

$$\text{MDD}(W|Z)^2 = \text{MDD}_o(W|Z)^2 + [\mathbb{E}(W)]^2 \mathbb{E}\left|Z - Z^\dagger\right|,$$

which follows directly from the definitions of $\text{MDD}_o$ and MDD. By Definition 3.1, we have that for any $\theta_1 = \left(\phi_1', g_1\right)' \in \Theta$ and $\theta_2 = \left(\phi_2', g_2\right)' \in \Theta$,

$$\begin{aligned} \text{d}_W(\theta_1, \theta_2)^2 &= \sum_{s=1}^{T-R} \text{MDD}\left[m_s(Y, X, \theta_1) - m_s(Y, X, \theta_2)|\underline{z}_s\right]^2 \\ &= \sum_{s=1}^{T-R} \text{MDD}_o\left[m_s(Y, X, \theta_1) - m_s(Y, X, \theta_2)|\underline{z}_s\right]^2 \\ &\quad + \sum_{s=1}^{T-R} \{\mathbb{E}[m_s(Y, X, \theta_1) - m_s(Y, X, \theta_2)]\}^2 \mathbb{E}\left|\underline{z}_s - \underline{z}_s^\dagger\right|. \end{aligned} \quad \textbf{(A.5)}$$

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

The rest of the proof is organized into three parts. In Part I, we show the existence of a constant $c_1 < \infty$ s.t.

$$\sum_{s=1}^{T-R} \mathrm{MDD}_o \left[ m_s(Y,X,\theta_1) - m_s(Y,X,\theta_2) | \underline{z}_s \right]^2 \leq c_1 \left\| \theta_1 - \theta_2 \right\|_{L^2}^2 .$$

In Part II, we show the existence of a constant $c_2 < \infty$ s.t.

$$\sum_{s=1}^{T-R} \{\mathbb{E}[m_s(Y,X,\theta_1) - m_s(Y,X,\theta_2)]\}^2 \, \mathbb{E}\left| \underline{z}_s - \underline{z}_s^{\dagger} \right| \leq c_2 \left\| \theta_1 - \theta_2 \right\|_{L^2}^2$$

which, together with the results from Part I, complete the proof Lemma 3.1(i). And in Part III, we prove Lemma 3.1(ii).

**Part I.** The compactness of $\boldsymbol{\Phi}$ according to Assumption 2.1(i) implies that $B_{\boldsymbol{\Phi}} \equiv \sup_{\phi \in \boldsymbol{\Phi}} |\phi| < \infty$. By Lemma A.3, for all $g \in \mathcal{G}$, it holds that $\sup_{x \in \mathcal{X}} |g(x)| \leq B_c < \infty$. Note that $m_s(Y,X,\theta) = [y_s - g(x_s)] + \sum_{r=1}^{R} \phi_{s,r}\left[ y_{T-R+r} - g\left(x_{T-R+r}\right)\right]$. Then for any $\theta_1 = \left(\phi_1', g_1\right)' \in \Theta$ and $\theta_2 = \left(\phi_2', g_2\right)' \in \Theta$, we have

$$\begin{aligned} m_s(Y,X,\theta_1) - m_s(Y,X,\theta_2) = & -[g_1(x_s) - g_2(x_s)] + \sum_{r=1}^{R} (\phi_{1,s,r} - \phi_{2,s,r})[y_{T-R+r} - g_1\left(x_{T-R+r}\right)] \\ & - \sum_{r=1}^{R} \phi_{2,s,r}\left[ g_1\left(x_{T-R+r}\right) - g_2\left(x_{T-R+r}\right)\right]. \end{aligned} \tag{A.6}$$

Then by the triangle inequality, we have

$$\begin{aligned} & |m_s(Y,X,\theta_1) - m_s(Y,X,\theta_2)| \\ \leq & \, |g_1(x_s) - g_2(x_s)| + \sum_{r=1}^{R} \left| \phi_{1,s,r} - \phi_{2,s,r} \right| |y_{T-R+r} - g_1(x_{T-R+r})| + \sum_{r=1}^{R} \left| \phi_{2,s,r} \right| |g_1(x_{T-R+r}) - g_2(x_{T-R+r})| \\ \leq & \, |g_1(x_s) - g_2(x_s)| + (|Y| + B_c) \sum_{r=1}^{R} \left| \phi_{1,s,r} - \phi_{2,s,r} \right| + B_{\boldsymbol{\Phi}} \sum_{r=1}^{R} |g_1(x_{T-R+r}) - g_2(x_{T-R+r})| \\ \leq & \, |g_1(x_s) - g_2(x_s)| + (|Y| + B_c) R |\phi_1 - \phi_2| + B_{\boldsymbol{\Phi}} \sum_{r=1}^{R} |g_1(x_{T-R+r}) - g_2(x_{T-R+r})| . \end{aligned}$$

It follows that for $s = 1, \ldots, T-R$,

$$\begin{aligned} & [m_s(Y,X,\theta_1) - m_s(Y,X,\theta_2)]^2 \\ \leq & \, 3 \left\{ [g_1(x_s) - g_2(x_s)]^2 + (|Y| + B_c)^2 R^2 |\phi_1 - \phi_2|^2 + R B_{\boldsymbol{\Phi}}^2 \sum_{r=1}^{R} \left[ g_1\left(x_{T-R+r}\right) - g_2\left(x_{T-R+r}\right)\right]^2 \right\} \\ \leq & \, B_{1m} \left\{ [g_1(x_s) - g_2(x_s)]^2 + \sum_{r=1}^{R} \left[ g_1\left(x_{T-R+r}\right) - g_2\left(x_{T-R+r}\right)\right]^2 \right\} + 3(|Y| + B_c)^2 R^2 |\phi_1 - \phi_2|^2 , \end{aligned} \tag{A.7}$$

where $B_{1m} = 3 \max\left\{ R B_{\boldsymbol{\Phi}}^2, 1 \right\}$, and the first inequality follows from the Cauchy–Schwarz inequality and Jensen inequality.

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

Denote by $\varphi_Z(s) \equiv \mathbb{E}\left[\exp\left(\mathbf{i}s'Z\right)\right]$ the characteristic function of $Z$. It holds that

$$\left|\mathrm{Var}\left(\exp\left(\mathbf{i}s'Z\right)\right)\right| = \left|\mathbb{E}\left[\exp\left(\mathbf{i}s'Z\right)^2\right] - \mathbb{E}\left[\exp\left(\mathbf{i}s'Z\right)\right]^2\right| = \left|\varphi_Z(2s) - \left[\varphi_Z(s)\right]^2\right|$$
$$\leq |\varphi_Z(2s)| + |\varphi_Z(s)|^2 \leq 2, \tag{A.8}$$

where the last inequality follows from the fact that $|\varphi_Z(\cdot)| \leq 1$. Now, by equation (2.4) in Su and Zheng (2017),

$$\begin{aligned}
&\mathrm{MDD}_o\left[m_s(Y,X,\theta_1) - m_s(Y,X,\theta_2)\,|\underline{z}_s\right]^2 \\
&= \int_{\mathbb{R}^{d_Z}} \left[\mathrm{Cov}\left(m_s(Y,X,\theta_1) - m_s(Y,X,\theta_2), \exp\left(\mathbf{i}s'Z\right)\right)\right]^2 q(s)\,ds \\
&\leq \mathrm{Var}\left(m_s(Y,X,\theta_1) - m_s(Y,X,\theta_2)\right) \int_{\mathbb{R}^{d_Z}} \left|\mathrm{Var}\left(\exp\left(\mathbf{i}s'Z\right)\right)\right| q(s)\,ds \\
&\leq 2\mathbb{E}\left\{\left[m_s(Y,X,\theta_1) - m_s(Y,X,\theta_2)\right]^2\right\} \int_{\mathbb{R}^{d_Z}} q(s)\,ds \\
&\leq 2c_q\left[B_{1m}\mathbb{E}\left\{\left[g_1(x_s) - g_2(x_s)\right]^2 + \sum_{r=1}^{R}\left[g_1(x_{T-R+r}) - g_2(x_{T-R+r})\right]^2\right\} + 3R^2\mathbb{E}[(|Y|+B_c)^2]\,|\phi_1-\phi_2|^2\right] \\
&\leq 2c_q\left[B_{1m}c_m(R+1)\,\|g_1-g_2\|_{L^2}^2 + B_{2m}\,|\phi_1-\phi_2|^2\right] \\
&\leq B_m\left\{\|g_1-g_2\|_{L^2}^2 + |\phi_1-\phi_2|^2\right\} = B_m\,\|\theta_1-\theta_2\|_{L^2}^2,
\end{aligned}$$

where $\mathbf{i} \equiv \sqrt{-1}$, $q(s) \equiv 1/\left[c\,|s|^{(1+d_Z)}\right]$, $c \equiv \pi^{(1+d_Z)/2}/\Gamma\left(\frac{1+d_Z}{2}\right)$, $\Gamma(\cdot)$ is the complete gamma function: $\Gamma(z) \equiv \int_0^\infty t^{(z-1)}\exp(-t)\,dt$, $c_q = \int_{\mathbb{R}^{d_Z}} q(s)\,ds < \infty$, $B_{2m} = 3R^2\mathbb{E}[(|Y|+B_c)^2] < \infty$, $c_m < \infty$ is a constant that depends on the density of $X$, and $B_m = 2c_q\max\{B_{1m}c_m(R+1), B_{2m}\} < \infty$. In the derivation above, the first inequality follows from Cauchy–Schwarz inequality, the second one holds by Jensen inequality and (A.8), and the third one holds by (A.7), and the fourth one holds by the boundedness of all density functions, according to Assumption 3.1(i). Consequently, we have that for any $\theta_1 \in \Theta$ and $\theta_2 \in \Theta$,

$$\sum_{s=1}^{T-R} \mathrm{MDD}_o\left[m_s(Y,X,\theta_1) - m_s(Y,X,\theta_2)\,|\underline{z}_s\right]^2 \leq c_1\,\|\theta_1-\theta_2\|_{L^2}^2, \tag{A.9}$$

where $c_1 \equiv (T-R)B_m < \infty$.

**Part II.** For any $\theta_1 \in \Theta$ and $\theta_2 \in \Theta$, we have

$$\begin{aligned}
&\{\mathbb{E}[m_s(Y,X,\theta_1) - m_s(Y,X,\theta_2)]\}^2\,\mathbb{E}\left|\underline{z}_s - \underline{z}_s^\dagger\right| \\
&\leq \{\mathbb{E}[m_s(Y,X,\theta_1) - m_s(Y,X,\theta_2)]\}^2\,\mathbb{E}\left|Z - Z^\dagger\right| \leq \mathbb{E}\left\{[m_s(Y,X,\theta_1) - m_s(Y,X,\theta_2)]^2\right\}\mathbb{E}\left|Z - Z^\dagger\right| \\
&\leq \left[B_{1m}\mathbb{E}\left\{\left[g_1(x_s) - g_2(x_s)\right]^2 + \sum_{r=1}^{R}\left[g_1(x_{T-R+r}) - g_2(x_{T-R+r})\right]^2\right\} + 3R^2\mathbb{E}[(|Y|+B_c)^2]\,|\phi_1-\phi_2|^2\right] \\
&\quad \times \mathbb{E}\left|Z - Z^\dagger\right| \\
&\leq \left[B_{1m}c_m(R+1)\,\|g_1-g_2\|_{L^2}^2 + B_{2m}\,|\phi_1-\phi_2|^2\right]\mathbb{E}\left|Z - Z^\dagger\right| \\
&\leq \tilde{B}_m\,\|\theta_1-\theta_2\|_{L^2}^2,
\end{aligned}$$

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

where $\tilde{B}_m = \max\{B_{1m}c_m(R+1), B_{2m}\}\mathbb{E}\left|Z - Z^\dagger\right| < \infty$, the second inequality holds by Jensen inequality, and the third one follows from (A.7). Consequently, we have that for any $\theta_1 \in \Theta$ and $\theta_2 \in \Theta$,

$$\sum_{s=1}^{T-R} \{\mathbb{E}[m_s(Y,X,\theta_1) - m_s(Y,X,\theta_2)]\}^2 \mathbb{E}\left|\underline{z}_s - \underline{z}_s^\dagger\right| \le c_2 \|\theta_1 - \theta_2\|_{L^2}^2, \tag{A.10}$$

where $c_2 \equiv (T-R)\tilde{B}_m < \infty$.

Combining (A.5), (A.9), and (A.10) yields that for any $\theta_1 \in \Theta$ and $\theta_2 \in \Theta$, $\mathrm{d}_w(\theta_1,\theta_2)^2 \le (c_1+c_2)\|\theta_1-\theta_2\|_{L^2}^2 = c^2\|\theta_1-\theta_2\|_{L^2}^2$, where $c \equiv \sqrt{c_1+c_2}$. This proves the first claim in Lemma 3.1.

**Part III.** To prove Lemma 3.1(ii), for any $\theta_1, \theta_2, \theta_3 \in \Theta$, define

$$W_{1,s} \equiv m_s(X,Y,\theta_1) - m_s(X,Y,\theta_3) \text{ and} W_{2,s} \equiv m_s(X,Y,\theta_2) - m_s(X,Y,\theta_3).$$

It follows from the second inequality result of Lemma A.2 that

$$\begin{aligned} \mathrm{MDD}\left(W_{1,s} - W_{2,s}|\underline{z}_s\right)^2 &\le 4\left[\mathrm{MDD}\left(W_{1,s}|\underline{z}_s\right) + \mathrm{MDD}\left(W_{2,s}|\underline{z}_s\right)\right]^2 \\ &\le 8\left[\mathrm{MDD}\left(W_{1,s}|\underline{z}_s\right)^2 + \mathrm{MDD}\left(W_{2,s}|\underline{z}_s\right)^2\right]. \end{aligned} \tag{A.11}$$

Noting that $W_{1,s} - W_{2,s} = m_s(X,Y,\theta_1) - m_s(X,Y,\theta_2)$, it follows from (A.11) and Definition 3.1 of $d_w(\cdot,\cdot)$ that

$$\begin{aligned} [d_w(\theta_1,\theta_2)]^2 &= \sum_{s=1}^{T-R} \mathrm{MDD}\left(W_{1,s} - W_{2,s}|\underline{z}_s\right)^2 \\ &\le 8\left[\sum_{s=1}^{T-R}\mathrm{MDD}\left(W_{1,s}|\underline{z}_s\right)^2 + \sum_{s=1}^{T-R}\mathrm{MDD}\left(W_{2,s}|\underline{z}_s\right)^2\right] \\ &= 8\left[\left[d_w(\theta_1,\theta_3)\right]^2 + \left[d_w(\theta_2,\theta_3)\right]^2\right] \\ &\le 8\left[d_w(\theta_1,\theta_3) + d_w(\theta_2,\theta_3)\right]^2, \end{aligned} \tag{A.12}$$

where the last inequality follows from the fact that the interaction term $2d_w(\theta_1,\theta_3)$ $d_w(\theta_2,\theta_3) \ge 0$. Lemma 3.1(ii) follows immediately from (A.12). □

To prove Theorem 3.1, we introduce some notations adopted from Arcones and Giné (1993) and de la Peña and Giné (1999, Chap. 5). Let $(S,\mathcal{S},P)$ be a probability space, and let $\{\xi_i\}_{i=1}^N$ be an i.i.d. sequence with probability law $P$. Let $\mathcal{F}$ be a class of measurable real functions on $S^m$. The $m$ th order $U$-process based on $P$ and indexed by $\mathcal{F}$ is

$$U_N^m(f) \equiv U_N^m(f;P) \equiv \frac{(N-m)!}{N!} \sum_{\mathbf{i}_m \in I_N^m} f\left(\xi_{i_1},\ldots,\xi_{i_m}\right), f \in \mathcal{F}, \tag{A.13}$$

where $\mathbf{i}_m \equiv (i_1,\ldots,i_m)$, $I_N^m = \{(i_1,\ldots,i_m):\ i_j \in \mathbb{N},\ 1 \le i_j \le N, \text{ and } i_j \ne i_k \text{ if } j \ne k\}$. We will repeatedly use the Hoeffding's decomposition of a $U$-statistic. The operator $\pi_{k,m} = \pi_{k,m}^P$ acts on $P^m$-integrable function $f: S^m \to \mathbb{R}$ as follows:

$$\pi_{k,m} f(\xi_1,\ldots,\xi_k) = \left(\delta_{\xi_1} - P\right)\cdots\left(\delta_{\xi_k} - P\right)P^{m-k}f, \tag{A.14}$$

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

where $\delta_{\xi_j}$ is the Dirac measure at the observation $\xi_j$. Note that $\pi_{k,m}f$ is a $P$-canonical function of $k$ variables. Then we have the following Hoeffding's decomposition

$$U_N^m(f) = \sum_{k=0}^{m} \binom{m}{k} U_N^m(\pi_{k,m} \circ S_m f), \tag{A.15}$$

where $S_m f$ is a symmetric version of $f$ : $S_m f(\xi_1, \ldots, \xi_k) = (m!)^{-1} \sum f\left(\xi_{i_1}, \ldots, \xi_{i_m}\right)$ with the sum extended over $m!$ permutations $(i_1, \ldots, i_m)$ of $\{1, \ldots, m\}$.

Given a pseudometric space $(\mathcal{F}, e)$, the $\varepsilon$-covering number of $(\mathcal{F}, e)$ is

$$\mathbb{N}(\varepsilon, \mathcal{F}, e) \equiv \min\left\{ n : \exists f_1, \ldots, f_n \in \mathcal{F} \text{ s.t. } \sup_{f \in \mathcal{F}} \min_{i \le n} e(f, f_i) \le \varepsilon \right\}.$$

We define $\mathbb{N}_{N,p}(\epsilon, \mathcal{F}) \equiv \mathbb{N}_{N,p}\left(\epsilon, \mathcal{F}, e_{N,p}\right)$ as the *random* $\varepsilon$-covering numbers of $(\mathcal{F}, e_{N,p})$, where $e_{N,p}(f,g) = \left\{U_N^m(|f-g|^p)\right\}^{1/p}$ where $p \ge 1$. Note that $e_{N,p}$ denotes the $L^p$ distance corresponding to the random measure that assigns mass $\frac{(N-m)!}{N!}$ to each of the points $(\xi_{i_1}, \ldots, \xi_{i_m}) \in S^m$, $\mathbf{i}_m \in I_N^m$. When $\mathcal{F}$ is a class of real symmetric measurable functions on $S^m$, we define pseudo-distances $e_{N,k,2}$ on $\mathcal{F}$ as follows:

$$e_{N,k,2}^2(f,g) \equiv \frac{N^k}{\binom{N}{k}} U_N^k \left(\pi_{k,m}(f-g)^2\right).$$

By the proof of Corollary 5.7 in Arcones and Giné (1993), there exist some positive finite constants $c_{k,r}$ such that for all $\varepsilon > 0$,

$$\mathbb{N}_{N,2}\left(\varepsilon, \pi_{k,m}\mathcal{F}\right) \le \prod_{r=0}^{k} \mathbb{N}\left(\frac{\varepsilon}{2(k+1)^{1/2} c_{k,2}}, \mathcal{F}, \ \|\cdot\|_{L^2\left(U_N^r \times P^{m-r}\right)}\right), \tag{A.16}$$

where for $r > 0$, $U_N^r \times P^{m-r}$ denotes the random probability measure

$$U_N^r \times P^{m-r} = \frac{(N-r)!}{N!} \sum_{\mathbf{i}_r \in I_N^r} \delta_{(\xi_{i_1}, \ldots, \xi_{i_r})} \times P^{m-r}$$

defined on $\left(S^m, \mathcal{S}^m\right)$ and for $r = 0$, $U_N^0 \times P^m$ just means $P^m$. Here $L^2\left(U_N^r \times P^{m-r}\right)$ defines the pseudometric on $S^m$ :

$$\|f-g\|_{L^2\left(U_N^r \times P^{m-r}\right)}^2 = U_N^r \times P^{m-r}(f-g)^2.$$

Note that

$$\mathbb{N}\left(\varepsilon, \mathcal{F}, \ \|\cdot\|_{L^2\left(U_N^0 \times P^m\right)}\right) \simeq 2P^m F/\varepsilon \text{ if } \varepsilon \le 2P^m F \text{ and equals } 1 \text{ otherwise.} \tag{A.17}$$

**Proof of Theorem 3.1.** Note that $S_{Ns}(\theta) = S_{Ns,1}(\theta) + S_{Ns,2}(\theta)$, where $S_{Ns,1}(\theta) = -\frac{1}{N}\sum_{1 \le i \ne j \le N} m_{is}(\theta) \times m_{js}(\theta) \kappa_{ij,s}$ and $S_{Ns,2}(\theta) = \frac{2}{N}\sum_{1 \le i \ne j \le N} m_{is}(\theta) \kappa_{ij,s} \frac{1}{N}\sum_{k=1}^{N} m_{ks}(\theta)$. Let $\bar{m}_{is}(\theta) = \mathbb{E}\left[m_{is}(\theta) | z_{is}\right]$ and $\tilde{m}_{is}(\theta) = m_{is}(\theta) - \bar{m}_{is}(\theta)$. Then

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

$$S_{Ns,1}(\theta) = -\frac{1}{N}\sum_{1\le i\ne j\le N}\left[\tilde{m}_{is}(\theta)+\bar{m}_{is}(\theta)\right]\left[\tilde{m}_{js}(\theta)+\bar{m}_{js}(\theta)\right]\kappa_{ij,s}$$
$$= -\frac{1}{N}\sum_{1\le i\ne j\le N}\tilde{m}_{is}(\theta)\tilde{m}_{js}(\theta)\kappa_{ij,s} - \frac{1}{N}\sum_{1\le i\ne j\le N}\bar{m}_{is}(\theta)\bar{m}_{js}(\theta)\kappa_{ij,s} - \frac{2}{N}\sum_{1\le i\ne j\le N}\tilde{m}_{is}(\theta)\bar{m}_{js}(\theta)\kappa_{ij,s},$$

and

$$S_{Ns,2}(\theta) = \frac{2}{N^2}\sum_{1\le i\ne j\le N}\sum_{k=1}^{N}\tilde{m}_{is}(\theta)\tilde{m}_{ks}(\theta)\kappa_{ij,s} + \frac{2}{N^2}\sum_{1\le i\ne j\le N}\sum_{k=1}^{N}\bar{m}_{is}(\theta)\bar{m}_{ks}(\theta)\kappa_{ij,s}$$
$$+\frac{2}{N^2}\sum_{1\le i\ne j\le N}\sum_{k=1}^{N}\left[\tilde{m}_{is}\bar{m}_{ks}(\theta)+\bar{m}_{is}(\theta)\tilde{m}_{ks}(\theta)\right]\kappa_{ij,s}.$$

Then $S_{Ns}(\theta) = \tilde{S}_{Ns,1}(\theta)+\tilde{S}_{Ns,2}(\theta)+\tilde{S}_{Ns,3}(\theta)$, where

$$\tilde{S}_{Ns,1}(\theta) = -\frac{1}{N}\sum_{1\le i\ne j\le N}\tilde{m}_{is}(\theta)\tilde{m}_{js}(\theta)\kappa_{ij,s} + \frac{2}{N^2}\sum_{1\le i\ne j\le N}\sum_{k=1}^{N}\tilde{m}_{is}(\theta)\tilde{m}_{ks}(\theta)\kappa_{ij,s},$$
$$\tilde{S}_{Ns,2}(\theta) = -\frac{1}{N}\sum_{1\le i\ne j\le N}\bar{m}_{is}(\theta)\bar{m}_{js}(\theta)\kappa_{ij,s} + \frac{2}{N^2}\sum_{1\le i\ne j\le N}\sum_{k=1}^{N}\bar{m}_{is}(\theta)\bar{m}_{ks}(\theta)\kappa_{ij,s},$$
$$\tilde{S}_{Ns,3}(\theta) = -\frac{2}{N}\sum_{1\le i\ne j\le N}\tilde{m}_{is}(\theta)\bar{m}_{js}(\theta)\kappa_{ij,s} + \frac{2}{N^2}\sum_{1\le i\ne j\le N}\sum_{k=1}^{N}\left[\tilde{m}_{is}\bar{m}_{ks}(\theta)+\bar{m}_{is}(\theta)\tilde{m}_{ks}(\theta)\right]\kappa_{ij,s}.$$

We prove parts (i) and (ii) of the theorem in turn.

**Part I. Proof of part (i)**.

When $\theta \in \Theta_I$, $\bar{m}_{is}(\theta) = 0$ for all $i = 1,\ldots,N$ and $s = 1,\ldots,T-R$. This implies that $\tilde{S}_{Ns,2}(\theta) = \tilde{S}_{Ns,3}(\theta) = 0$. We are left to study $\tilde{S}_{Ns,1}(\theta)$. For the second term in the definition of $\tilde{S}_{Ns,1}(\theta)$, we have

$$\frac{2}{N}\sum_{1\le i\ne j\le N}\sum_{k=1}^{N}\tilde{m}_{is}(\theta)\tilde{m}_{ks}(\theta)\kappa_{ij,s}$$
$$= \frac{2}{N^2}\sum_{1\le i\ne j\le N}\tilde{m}_{is}(\theta)^2\kappa_{ij,s} + \frac{2}{N^2}\sum_{1\le i\ne j\le N}\tilde{m}_{is}(\theta)\tilde{m}_{js}(\theta)\kappa_{ij,s} + \frac{2}{N^2}\sum_{1\le i\ne j\ne k\le N}\tilde{m}_{is}(\theta)\tilde{m}_{ks}(\theta)\kappa_{ij,s}$$
$$= \frac{2}{N^2}\sum_{1\le i\ne j\le N}\tilde{m}_{is}(\theta)^2\kappa_{ij,s} + \frac{2}{N^2}\sum_{1\le i\ne j\le N}\tilde{m}_{is}(\theta)\tilde{m}_{js}(\theta)\kappa_{ij,s} + \frac{(N-1)(N-2)}{N^2}N\mathbb{U}_{2Ns},$$

where $\mathbb{U}_{2Ns} = \binom{N}{3}^{-1}\sum_{1\le i<j\le k\le N}\psi_s\left(\xi_i,\xi_j,\xi_k;\theta\right)$ and $\psi_s\left(\xi_i,\xi_j,\xi_k;\theta\right) = \frac{1}{3}[\tilde{m}_{is}(\theta)\tilde{m}_{ks}(\theta)\kappa_{ij,s}+\tilde{m}_{is}(\theta) \times \tilde{m}_{js}(\theta)\kappa_{ik,s} + \tilde{m}_{js}(\theta)\tilde{m}_{ks}(\theta)\kappa_{jk,s} + \tilde{m}_{js}(\theta)\tilde{m}_{is}(\theta)\kappa_{jk,s} + \tilde{m}_{ks}(\theta)\tilde{m}_{is}(\theta)\kappa_{jk,s} + \tilde{m}_{ks}(\theta)\tilde{m}_{js}(\theta)\kappa_{ik,s}]$ is a symmetrized version of $\psi_{0s}\left(\xi_i,\xi_j,\xi_k;\theta\right) \equiv 2\tilde{m}_{is}(\theta)\tilde{m}_{ks}(\theta)\kappa_{ij,s}$. Note that

$$\mathbb{E}\left[\psi_s(\xi_1,\xi_2,\xi_3;\theta)\right] = 0,\ \mathbb{E}\left[\psi(\xi_1,\xi_2,\xi_3;\theta)\,|\xi_1\right] = 0 \text{ and}$$
$$\mathbb{E}\left[\psi_s(\xi_1,\xi_2,\xi_3;\theta)\,|\xi_1,\xi_2\right] = \frac{1}{3}\tilde{m}_{1s}(\theta)\tilde{m}_{2s}(\theta)\left[\mathbb{E}_3(\kappa_{13,s})+\mathbb{E}_3(\kappa_{23,s})\right] \equiv h_s^{(2)}(\xi_1,\xi_2;\theta).$$

Let $h_s^{(3)}(\xi_1,\xi_2,\xi_3;\theta) = \psi_s(\xi_1,\xi_2,\xi_3;\theta) - [h_s^{(2)}(\xi_1,\xi_2;\theta)+h_s^{(2)}(\xi_1,\xi_3;\theta)+h_s^{(2)}(\xi_2,\xi_3;\theta)]$. By Hoeffding's decomposition in (A.15) (see also Lee (1990, p. 26) and de la Peña and

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

Giné (1999, p. 137)), we have $\mathbb{U}_{2Ns}(\theta) = 3\mathbb{H}_{2Ns}(\theta) + \mathbb{H}_{3Ns}(\theta)$, where

$$\mathbb{H}_{2Ns}(\theta) = \binom{N}{2}^{-1} \sum_{1\leq i<j\leq N} h_s^{(2)}\left(\xi_i, \xi_j; \theta\right) \text{ and } \mathbb{H}_{3Ns}(\theta) = \binom{N}{3}^{-1} \sum_{1\leq i<j\leq k\leq N} h_s^{(3)}\left(\xi_i, \xi_j, \xi_k; \theta\right).$$

Similarly, we can write the first term in the definition of $\tilde{S}_{Ns,1}(\theta)$ as follows: $-\frac{1}{N}\sum_{1\leq i\neq j\leq N} \tilde{m}_{is}(\theta)\tilde{m}_{js}(\theta)\kappa_{ij,s} = \frac{N-1}{N} N\mathbb{U}_{1Ns}(\theta)$, where $\mathbb{U}_{1N}(\theta) = -\binom{N}{2}^{-1} \sum_{1\leq i\neq j\leq N} \tilde{m}_{is}(\theta)\tilde{m}_{js}(\theta)\kappa_{ij,s}$. Then we have

$$\begin{aligned}
\tilde{S}_{Ns,1}(\theta) &= \frac{N-1}{N} N\mathbb{U}_{1Ns}(\theta) + \frac{(N-1)(N-2)}{N^2} N\mathbb{U}_{2Ns} + \frac{2}{N^2}\sum_{1\leq i\neq j\leq N} \tilde{m}_{is}(\theta)^2\kappa_{ij,s} \\
&\quad + \frac{2}{N^2}\sum_{1\leq i\neq j\leq N} \tilde{m}_{is}(\theta)\tilde{m}_{js}(\theta)\kappa_{ij,s} \\
&= N\mathbb{U}_{Ns}(\theta) + N\mathbb{H}_{3Ns}(\theta) + \frac{2}{N^2}\sum_{1\leq i\neq j\leq N} \tilde{m}_{is}(\theta)^2\kappa_{ij,s} - \frac{3N-2}{N}\mathbb{U}_{1Ns}(\theta) - \frac{3N-2}{N}\mathbb{U}_{2Ns},
\end{aligned}$$

where $\mathbb{U}_{Ns}(\theta) = \mathbb{U}_{1Ns}(\theta) + 3\mathbb{H}_{2Ns}(\theta) = \binom{N}{2}^{-1}\sum_{1\leq i<j\leq N} h_s\left(\xi_i, \xi_j; \theta\right)$ and $h_s\left(\xi_i, \xi_j; \theta\right) = \tilde{m}_{is}(\theta)\tilde{m}_{js}(\theta) \times [\mathbb{E}_j(\kappa_{ij,s}) + \mathbb{E}_i(\kappa_{ij,s}) - \kappa_{ij,s}]$.

Let $\mathcal{X}$ and $\mathcal{Z}_s$ denote the supports of $x_{it}$ and $z_{is}$, respectively. Note that $m_s(Y, X; \theta) = H_s(\phi_s)'[Y - \mathbf{g}(X)] = [y_s - g(x_s)] + \sum_{r=1}^{R} \phi_{s,r}\left[y_{T-R+r} - g\left(x_{T-R+r}\right)\right]$ and $\phi_s = (\phi_{s,1}, \ldots, \phi_{s,R})'$ for $s = 1, \ldots, T-R$. Let $\theta_s = (\phi_s', g)'$. Define

$$\begin{aligned}
\mathcal{F}_{1s} \equiv \{m_s(\cdot,\cdot;\theta_s) : \mathbb{R}^T \times \mathcal{X}^T \to \mathbb{R} :\ & m_s(y, x; \theta_s) = [y_s - g(x_s)] \\
&+ \sum_{r=1}^{R} \phi_{s,r}\left[y_{T-R+r} - g\left(x_{T-R+r}\right)\right] \text{ for some } \theta_s = \left(\phi_s', g\right)' \in \Phi_s \times \mathcal{G}\},
\end{aligned} \quad \textbf{(A.18)}$$

where $\Phi_s \equiv \left\{\phi_s \in \mathbb{R}^R : \|\phi_s\| \leq c_\phi\right\}$ for some constant $c_\phi$, and $\mathcal{G} \equiv \{g \in W^s(\mathcal{X}) : \|g\|_s \leq C_g\}$. Similarly, let $\xi = (y', x', z')'$ and $S = \mathbb{R}^T \times \mathcal{X}^T \times \mathcal{Z}^T$. Define

$$\begin{aligned}
\mathcal{F}_{1s}^c \equiv \{\tilde{m}_s(\cdot;\theta_s) : S \to \mathbb{R} :\ & \tilde{m}_s(\xi;\theta_s) = [y_s - g(x_s)] - \mathbb{E}\left[(y_{is} - g(x_{is}))|z_{is} = z_s\right] \\
&+ \sum_{r=1}^{R} \phi_{s,r}\{\left[y_{T-R+r} - g\left(x_{T-R+r}\right)\right] - \mathbb{E}\left[(y_{i,T-R+r} - g\left(x_{i,T-R+r}\right))|z_{is} = z_s\right]\} \\
&\text{for some } \theta_s = \left(\phi_s', g\right)' \in \Phi_s \times \mathcal{G}\},
\end{aligned} \quad \textbf{(A.19)}$$

$$\begin{aligned}
\mathcal{F}_{2s} \equiv \{f_s(\cdot,\cdot;\theta_s) : S \times S \to \mathbb{R} :\ & f_s(\xi_1, \xi_2; \theta_s) = \tilde{m}_s(\xi_1;\theta_s)\tilde{m}_s(\xi_2;\theta_s)\breve{\kappa}_{12,s} \\
&\text{for some } \theta_s = \left(\phi_s', g\right)' \in \Phi_s \times \mathcal{G}\},
\end{aligned} \quad \textbf{(A.20)}$$

and

$$\begin{aligned}
\mathcal{F}_{3s} \equiv \{f_s(\cdot,\cdot,\cdot;\theta_s) : S \times S \times S \to \mathbb{R} :\ & f_s(\xi_1, \xi_2, \xi_3; \theta_s) = \tilde{m}_s(\xi_1;\theta_s)\tilde{m}_s(\xi_2;\theta_s)\breve{\kappa}_{12,s} \\
&+ \tilde{m}_s(\xi_1;\theta_s)\tilde{m}_s(\xi_3;\theta_s)\breve{\kappa}_{13,s} + \tilde{m}_s(\xi_2;\theta_s)\tilde{m}_s(\xi_3;\theta_s)\breve{\kappa}_{23,s} \\
&\text{for some } \theta_s = \left(\phi_s', g\right)' \in \Phi_s \times \mathcal{G}\},
\end{aligned} \quad \textbf{(A.21)}$$

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

where, for example, $\underline{z}_{is} = \left(z_{i1}', \ldots, z_{is}'\right)'$ and $\breve{\kappa}_{ij,s} = \mathbb{E}_j(\kappa_{ij,s}) + \mathbb{E}_i(\kappa_{ij,s}) - \kappa_{ij,s}$. The compactness of $\mathbf{\Phi}$ according to Assumption 2.1(i) implies that $B_{\mathbf{\Phi}} \equiv \sup_{\phi \in \mathbf{\Phi}} |\phi| < \infty$. By Lemma A.3, for all $g \in \mathcal{G}$, it holds that $\sup_{x \in \mathcal{X}} |g(x)| \le B_c < \infty$. Then for any $\theta \in \Theta$ and $s = 1, \ldots, T-R$, we have

$$\begin{aligned} |m_s(Y,X,\theta)| &\le [|y_s| + |g(x_s)|] + \sum_{r=1}^{R} |\phi_{s,r}| \left[|y_{T-R+r}| + \left|g\left(x_{T-R+r}\right)\right|\right] \\ &\le |Y| + B_c + B_{\mathbf{\Phi}} R[|Y| + B_c] = (B_{\mathbf{\Phi}} R + 1)[|Y| + B_c] \le K[|Y| + 1] \equiv F_1(Y), \end{aligned} \tag{A.22}$$

where $Y = (y_1, \ldots, y_T)'$, $X = (x_1, \ldots, x_T)'$, $\phi_{s,r}$ denote the $r$th element in $\phi_s$, and $K$ is a generic positive constant that may vary across lines. By (A.6),

$$\begin{aligned} |m_s(Y,X,\theta_1) - m_s(Y,X,\theta_2)| &\le (|Y| + B_c) R |\phi_1 - \phi_2| + (RB_{\mathbf{\Phi}} + 1)\|g_1 - g_2\|_\infty \\ &\le K(|Y| + 1)\left\{|\phi_1 - \phi_2| + \|g_1 - g_2\|_\infty\right\}. \end{aligned}$$

It follows that the class $\mathcal{F}_{1s}$ is Lipschitz in $\Phi_s \times \mathcal{G}$ (w.r.t.) the norm $|\cdot| + \|\cdot\|_\infty$. Then by Theorem 2.7.11 in van der Vaart and Wellner (1996), we have

$$\begin{aligned} \mathbb{N}_{[]}\left(\epsilon \|F_1\|, \mathcal{F}_{1s}, \|\cdot\|_{L^2}\right) &\le \mathbb{N}(\epsilon/2, \Phi_s \times \mathcal{G}, |\cdot| + \|\cdot\|_\infty) \le \mathbb{N}(\epsilon/4, \Phi_s, |\cdot|)\,\mathbb{N}(\epsilon/4, \mathcal{G}, \|\cdot\|_\infty) \\ &\le K\left(\frac{4}{\epsilon}\right)^R \exp\left[\left(\frac{4}{\epsilon}\right)^\nu\right], \end{aligned} \tag{A.23}$$

where the first inequality follows from Theorem 2.7.11 in van der Vaart and Wellner (1996) and the last one follows from Lemma A.3 in Santos (2012), which indicates that $\mathbb{N}(\epsilon, \mathcal{G}, \|\cdot\|_\infty) \le K \exp\left(\left(\frac{1}{\epsilon}\right)^\nu\right)$ with $\nu$ being defined in (2.8). This also implies that

$$\mathbb{N}_{[]}\left(2\epsilon\|F\|, \mathcal{F}_{1s}^c, \|\cdot\|_{L^2}\right) \le K\left(\frac{4}{\epsilon}\right)^{2R} \exp\left[2\left(\frac{4}{\epsilon}\right)^\nu\right]. \tag{A.24}$$

Let $F_2(\xi_1, \xi_2) = K(|y_1| + 1)(|y_2| + 1)(|z_1| + |z_2| + 1)$ with $\xi_i = (y_i', x_i', z_i')' \in S$. By arguments as used in the proof of Theorem 6 in Andrews (1994), there is a finite positive constant $c_0$ such that

$$\mathbb{N}_{[]}(2c_0\epsilon\|F_2\|, \mathcal{F}_{2s}, \|\cdot\|) \le \left[\mathbb{N}_{[]}(2\epsilon\|F_1\|, \mathcal{F}_{2s}, \|\cdot\|)\right]^2 \le K\left(\frac{4}{\epsilon}\right)^{4R} \exp\left[4\left(\frac{4}{\epsilon}\right)^\nu\right]. \tag{A.25}$$

We verify the conditions in Theorem 5.6 of Arcones and Giné (1993, AG hereafter). First, by Assumption 3.1(i) and (ii),

$$\begin{aligned} \mathbb{E}\left([F_2(\xi_1, \xi_2)]^2\right) &\le 2K^2 E\left\{[(|Y_1| + 1)(|Y_2| + 1)|Z_1|]^2 + [(|Y_1| + 1)(|Y_2| + 1)|Z_2|]^2\right\} \\ &\le 4K^2 \mathbb{E}\left[(|Y_1| + 1)^2 |Z_1|^2\right] \mathbb{E}\left[(|Y_2| + 1)^2\right] \le \infty. \end{aligned}$$

This verifies Condition (a) in Theorem 5.6 of AG. Applying (A.16) with $m = k = 2$ yields

$$\mathbb{N}_{N,2}(\varepsilon, \mathcal{F}_{2s}) = \mathbb{N}_{N,2}\left(\varepsilon, \pi_{2,2}\mathcal{F}_{2s}\right) \le \prod_{r=0}^{2} \mathbb{N}\left(\frac{\varepsilon}{2\sqrt{3}c_{2,r}}, \mathcal{F}_{2s}, \|\cdot\|_{L^2(U_N^r \times P^{2-r})}\right).$$

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

By (A.17), $\int_0^\delta \log \mathbb{N}\left(\frac{\varepsilon}{2\sqrt{3}c_{2,0}}, \mathcal{F}_{2s},\ \|\cdot\|_{L^2\left(U_N^0 \times P^2\right)}\right) d\epsilon \preceq \int_0^\delta \log \frac{1}{\varepsilon} d\epsilon$. Note that

$$\begin{aligned}
&\int_0^\delta \log \mathbb{N}\left(\frac{\varepsilon}{2\sqrt{3}c_{2,1}}, \mathcal{F}_{2s},\ \|\cdot\|_{L^2\left(U_N^1 \times P^1\right)}\right) d\varepsilon \\
&= 2\sqrt{3}c_{2,1}\left[U_N^1(P^1 F_2^2)\right]^{1/2} \int_0^{\delta/\left[2\sqrt{3}c_{2,1}U_N^1(P^1 F^2)\right]^{1/2}} \\
&\quad \times \log N\left(\varepsilon\left[U_N^1(P^1 \bar{F}^2)\right]^{1/2}, \mathcal{F}_{2s},\ \|\cdot\|_{L^2(U_N^1 \times P^1)}\right) d\varepsilon \\
&\preceq \left[U_N^1(P^1 F_2^2)\right]^{1/2} \int_0^{\delta/\left[2\sqrt{3}c_{2,1}U_N^1(P^1 F_2^2)\right]^{1/2}} \left[\log\left(\frac{4}{\varepsilon}\right) + \left(\frac{4}{\varepsilon}\right)^{\nu}\right] d\varepsilon \\
&\preceq \left[U_N^1(P^1 F_2^2)\right]^{1/2} \int_0^{\delta/\left[2\sqrt{3}c_{2,1}U_N^1(P^1 F_2^2)\right]^{1/2}} \varepsilon^{-\nu} d\varepsilon \\
&\preceq \left[U_N^1(P^1 F_2^2)\right]^{\nu/2} \delta^{1-\nu},
\end{aligned}$$

where the first equality follows from the change of variables, the first inequality holds by (A.25), and the second inequality follows from the fact that the integrand is dominated by the term $(\varepsilon/4)^{-\nu}$ in the neighborhood of 0, $\nu < 1$ by Assumption 2.1(iii), and $\int_0^{\delta'} \log(1/\varepsilon)\, d\varepsilon < \infty$ for any $\delta' < \infty$. Similarly, we have

$$\begin{aligned}
&\int_0^\delta \log \mathbb{N}\left(\frac{\varepsilon}{2\sqrt{3}c_{2,2}}, \mathcal{F}_{2s},\ \|\cdot\|_{L^2(U_N^2)}\right) d\varepsilon \\
&= 2\sqrt{3}c_{2,2}\left[U_N^2(\bar{F}^2)\right]^{1/2} \int_0^{\delta/\left[2\sqrt{3}c_{2,2}U_N^2(F^2)\right]^{1/2}} \log \mathbb{N}\left(\varepsilon\left[U_N^2(F^2)\right]^{1/2}, \mathcal{F},\ \|\cdot\|_{L^2(U_N^2)}\right) d\varepsilon \\
&\preceq \left[U_N^2(P^2 F_2^2)\right]^{1/2} \int_0^{\delta/\left[2\sqrt{3}c_{2,2}U_N^2(F^2)\right]^{1/2}} \left[\log\left(\frac{4}{\varepsilon}\right) + \left(\frac{4}{\varepsilon}\right)^{\nu}\right] d\varepsilon \\
&\preceq \left[U_N^2(F_2^2)\right]^{1/2} \int_0^{\delta/\left[2\sqrt{3}c_{2,2}U_N^2(F^2)\right]^{1/2}} \varepsilon^{-\nu} d\varepsilon \preceq \left[U_N^2(F_2^2)\right]^{\nu/2} \delta^{1-\nu}.
\end{aligned}$$

Then

$$\begin{aligned}
&\lim_{\delta\to 0} \limsup_{N\to\infty} \mathbb{E}^o\left[\int_0^\delta \log \mathbb{N}_{N,2}(\varepsilon, \mathcal{F}_{2s})\, d\epsilon\right] \\
&= \lim_{\delta\to 0} \limsup_{N\to\infty} \mathbb{E}^*\left[\int_0^\delta \sum_{r=0}^{2} \log \mathbb{N}\left(\frac{\varepsilon}{2\sqrt{3}c_{2,r}}, \mathcal{F}_{2s},\ \|\cdot\|_{L^2\left(U_N^r \times P^{2-r}\right)}\right) d\varepsilon\right] \\
&\preceq \lim_{\delta\to 0} \limsup_{N\to\infty} \left[\int_0^\delta \log \frac{1}{\varepsilon} d\epsilon + \mathbb{E}\left\{\left[U_N^1(P^1 F_2^2)\right]^{\nu/2} + \left[U_N^2(F_2^2)\right]^{\nu/2}\right\} \delta^{1-\nu}\right] = 0,
\end{aligned}$$

where the last equality follows from the fact that $\mathbb{E}\{\left[U_N^2(F_2^2)\right]^{\nu/2}\} \leq \left\{\mathbb{E}\left[U_N^2(F_2^2)\right]\right\}^{\nu/2} = \{\mathbb{E}([F_2(\xi_1,\xi_2)]^2\}^{\nu/2} < \infty$ by Jensen inequality and similarly $\mathbb{E}\{\left[U_N^1(P^1F_2^2)\right]^{\nu/2}\} < \infty$. This verifies condition (c) in Theorem 5.6 of AG. Next, notice that $\mathbb{N}\left(\frac{\varepsilon}{2\sqrt{3}c_{2,r}}, \mathcal{F}_{2s}, \|\cdot\|_{L^2\left(U_N^r \times P^{2-r}\right)}\right) = 1$ a.s. for $r = 0,1,2$ and for sufficiently large $\varepsilon$, say, $\varepsilon \geq \varepsilon_0$, by the total boundedness of $\Phi \times \mathcal{G}$and the law of large numbers for U-statistics. It follows that for some small $\epsilon > 0$ and by the above calculations,

$$\begin{aligned}
&\mathbb{E}^o\left|\int_0^\infty \log \mathbb{N}_{N,2}\left(\varepsilon, \pi_{2,2}\mathcal{F}_{2s}\right)d\varepsilon\right|^{1+\epsilon} \\
&\quad\preceq \mathbb{E}^o\left|\int_0^{\varepsilon_0} \log \mathbb{N}_{N,2}\left(\varepsilon, \pi_{2,2}\mathcal{F}_{2s}\right)d\varepsilon\right|^{1+\epsilon} \\
&\quad\preceq \left|\int_0^{\varepsilon_0} \log\frac{1}{\varepsilon}d\epsilon\right|^{1+\epsilon} + \mathbb{E}\left\{\left[U_N^1(P^1F_2^2)\right]^{(1+\epsilon)\nu/2} + \left[U_N^2(F_2^2)\right]^{(1+\epsilon)\nu/2}\right\} < \infty,
\end{aligned}$$

where $\mathbb{E}^o$ denotes the outer-expectation associated with $\mathbb{E}$, the last inequality holds by choosing $\epsilon$ sufficiently small such that $(1+\epsilon)\nu/2 \leq 1$. This implies that the sequence $\left\{\int_0^\infty \log N_{N,2}(\varepsilon, \mathcal{F}_{2s})d\varepsilon\right\}_{N=1}^\infty$ is uniformly integrable. That is, condition (b) in Theorem 5.6 of AG is verified. Then by Theorem 5.6 of AG, we have $N\mathbb{U}_N(\theta) \Longrightarrow \mathbb{C}_s(\theta)$ in $L^\infty(\Theta)$.

Next, note that $\mathbb{H}_{3N}(\theta)$ is a third-order $P-$ canonical $U-$process with the envelope function for the kernel in the definition of $\mathbb{H}_{3N}(\theta)$ given by $F_3(\xi_1,\xi_2,\xi_3) = K\{(|y_1|+1)(|y_2|+1)(|z_1|+|z_2|+1) + (|y_1|+1)(|y_3|+1)(|z_1|+|z_3|+1) + (|y_2|+1)(|y_3|+1)(|z_2|+|z_3|+1)\}$. Following the analysis of $\mathbb{U}_N(\theta)$, it is easy to show that $\mathbb{E}[F_3(\xi_1,\xi_2,\xi_3)^2] < \infty$ under Assumption 3.1 (i) and (ii),

$$\lim_{\delta\to 0}\limsup_{N\to\infty}\mathbb{E}^o\left[\int_0^\delta \left[\log \mathbb{N}_{N,2}(\varepsilon, \mathcal{F}_{3s})\right]^{3/2}d\epsilon\right] = 0$$

and the sequence $\left\{\int_0^\infty \left[\log N_{N,2}(\varepsilon,\mathcal{F}_{3s})\right]^{3/2}d\varepsilon\right\}_{N=1}^\infty$ is uniformly integrable. Here, we use the fact that $\frac{3}{2}(1+\epsilon)\nu/2 \leq 1$ for sufficiently small $\epsilon$. As a result, $N^{3/2}\mathbb{H}_{3N}(\theta)$ converges to a Gaussian chaos process and $\sup_{\theta\in\Theta}|N\mathbb{H}_{3N}(\theta)| = O_P\left(N^{-1/2}\right)$. Our condition is sufficient to ensure the uniform law of large numbers to hold for the $U$-process with kernel function associated with $\tilde{m}_{is}(\theta)^2\kappa_{ij,s}$. As a result, we have

$$\frac{2}{N^2}\sum_{1\leq i\neq j\leq N}\tilde{m}_{is}(\theta)^2\kappa_{ij,s} = 2\mathbb{E}\left[\tilde{m}_{1s}(\theta)^2\kappa_{12,s}\right] + o_p(1) \equiv \mathbb{B}_s(\theta) + o_p(1) \text{ uniformly in } \theta\in\Theta.$$

Following the analysis of $\mathbb{U}_{Ns}(\theta)$, we can also show that both $N\mathbb{U}_{1Ns}(\theta)$ and $N\mathbb{U}_{2Ns}(\theta)$ converge to Gaussian chaos processes. Consequently, we have

$$\tilde{S}_{Ns,1}(\theta) \Longrightarrow \mathbb{B}_s(\theta) + \mathbb{C}_s(\theta). \qquad \textbf{(A.26)}$$

When $\theta \in \Theta_I$, we also have $S_{Ns}(\theta) \Longrightarrow \mathbb{B}_s(\theta) + \mathbb{C}_s(\theta)$ given the fact that $\tilde{S}_{Ns,\ell}(\theta) = 0$ for $\ell = 2,3$ in this case. As a result, we have $S_N(\theta) \Longrightarrow \mathbb{B}(\theta) + \mathbb{C}(\theta)$.

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

**Part II. Proof of part (ii)**.

When $\theta \notin \Theta_I$, (A.26) continues to hold. By the law of large numbers for $U$-processes and the definition of MDD, we have

$$\begin{aligned}\frac{1}{N}\tilde{S}_{Ns,2}(\theta) &= -\frac{1}{N^2}\sum_{1\le i\neq j\le N}\bar{m}_{is}(\theta)\bar{m}_{js}(\theta)\kappa_{ij,s}+\frac{2}{N^3}\sum_{1\le i\neq j\le N}\sum_{k=1}^{N}\bar{m}_{is}(\theta)\bar{m}_{ks}(\theta)\kappa_{ij,s}\\ &= -\mathbb{E}\left[\bar{m}_{1s}(\theta)\bar{m}_{2s}(\theta)\kappa_{12,s}\right]+2\mathbb{E}\left[\bar{m}_{1s}(\theta)\kappa_{12,s}\right]\mathbb{E}\left[\bar{m}_{2s}(\theta)\right]+o_P(1)\\ &= \text{MDD}\left[m_s(Y,X,\theta)\,|\underline{z}_s\right]^2+o_P(1)\ \text{uniformly in}\ \theta\in\Theta\backslash\Theta_I.\end{aligned}$$

In fact, applying Hoeffding decomposition to $\frac{1}{N}\tilde{S}_{Ns,2}(\theta)$ and arguments as used in Part I, we can strengthen $o_P(1)$ to $O_P\left(N^{-1/2}\right)$ in the last claim.

Now, define $\mathbb{U}_{3N}(\theta)=\frac{(N-2)!}{N!}\sum_{1\le i\neq j\le N}\tilde{m}_{is}(\theta)\bar{m}_{js}(\theta)\kappa_{ij,s}$ and $\mathbb{U}_{4N}(\theta)=\frac{(N-3)!}{N!}\sum_{1\le i\neq j\neq k\le N}[\tilde{m}_{is}\bar{m}_{ks}(\theta)+\bar{m}_{is}(\theta)\tilde{m}_{ks}(\theta)]\kappa_{ij,s}$. It is easy to see that $\mathbb{U}_{3N}(\theta)$ and $\mathbb{U}_{4N}(\theta)$ are nondegenerate second- and third-order $U$ processes, respectively. One can easily apply symmetrization and similar calculations as above to verify the entropy condition in Theorem 4.10 of AG holds to conclude both $N^{1/2}\mathbb{U}_{3N}(\theta)$ and $N^{1/2}\mathbb{U}_{4N}(\theta)$ converge to Gaussian processes. This implies that

$$\begin{aligned}&\sup_{\theta\in\Theta\backslash\Theta_I}\frac{1}{N^{1/2}}\left|\tilde{S}_{Ns,3}(\theta)\right|\\ &=\sup_{\theta\in\Theta\backslash\Theta_I}\left|-\frac{2}{N^{3/2}}\sum_{1\le i\neq j\le N}\tilde{m}_{is}(\theta)\bar{m}_{js}(\theta)\kappa_{ij,s}\right.\\ &\quad\left.+\frac{2}{N^{5/2}}\sum_{1\le i\neq j\le N}\sum_{k=1}^{N}\left[\tilde{m}_{is}\bar{m}_{ks}(\theta)+\bar{m}_{is}(\theta)\tilde{m}_{ks}(\theta)\right]\kappa_{ij,s}\right|\\ &\preceq\sup_{\theta\in\Theta\backslash\Theta_I}\left|N^{1/2}\mathbb{U}_{3N}(\theta)\right|+\sup_{\theta\in\Theta\backslash\Theta_I}\left|N^{1/2}\mathbb{U}_{4N}(\theta)\right|=O_p(1).\end{aligned}$$

Consequently, we have $\frac{1}{N}S_N(\theta)=\sum_{s=1}^{T-R}\sum_{\ell=1}^{3}\frac{1}{N}\tilde{S}_{Ns,\ell}(\theta)=\sum_{s=1}^{T-R}\text{MDD}\ [m_s(Y,X,\theta)|\underline{z}_s]^2+O_P(N^{-1/2})$ uniformly in $\theta\in\Theta\backslash\Theta_I$. □

**Proof of Theorem 3.2.** In this proof Conditions (i)–(iv) listed in Lemma A.5 are referred to as C(i)–C(iv), respectively. Let $Q(\theta)\equiv\sum_{s=1}^{T-R}\text{MDD}\left[m_s(Y,X,\theta)\,|\underline{z}_s\right]^2$ and $Q_N(\theta)\equiv\frac{1}{N}S_N(\theta)=\sum_{s=1}^{T-R}\frac{1}{N}S_{Ns}(\theta)$. Our goal is to show that, over the restricted parameter space $\Theta\cap\Theta_R$ under $\mathrm{d}_w(\cdot,\cdot)$, $Q(\cdot)$ and $Q_N(\cdot)$ as specified above satisfy C(i)–C(iv) in Lemma A.5.

We first prove that $\mathrm{d}_w(\hat{\theta}_N,\Theta_I)=O_p\left(\max\{\delta_{w,N},N^{-1/4}\}\right)=O_p\left(N^{-1/4}\right)$ and then argue that such a rate can be improved to $O_p(\varrho_N N^{-1/2})$ by iterative arguments.

Due to the non-negativity of MDD, $Q(\cdot)\ge 0$. By Lemma A.3, $\Theta$ is compact under $\|\cdot\|_c$ and hence is compact under $\mathrm{d}_w(\cdot,\cdot)$, which is weaker than $\|\cdot\|_c$. Since $\Theta_R$ is closed due to the continuity of $L(\cdot)$ under Assumption 2.2, $\Theta\cap\Theta_R$ is also compact under $\mathrm{d}_w(\cdot,\cdot)$. So C(i) is satisfied. Assumption 3.3(i) and (ii) guarantee C(ii) to hold with $\sigma_N=\delta_{w,N}=o\left(N^{-1/2}\right)$. C(iii) holds according to Theorem 3.1 with $b_N=N^{-1/2}$. Obviously, C(iv) holds with

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

$a_1 = a_2 = 1$ by Lemma A.1 and the fact that $\mathrm{MDD}\left[m_s\left(Y,X,\theta^0\right)|\underline{z}_s\right] = 0$ for any $\theta^0 \in \Theta_I$. Then by Lemma A.5, we have

$$\mathrm{d}_w\left(\hat{\theta}_N, \Theta_I\right) = O_p\left(\max\left\{\delta_{w,N}, N^{-1/4}\right\}\right) = O_p\left(N^{-1/4}\right).$$

This, in conjunction with Assumption 3.4 and Lemma A.6, implies that

$$\mathrm{d}_{\|\cdot\|_{L^2}}(\hat{\theta}_N, \Theta_I) = O_p(\varrho_N \mathrm{d}_w(\hat{\theta}_N, \Theta_I \cap \Theta_R) + \delta_{s,N}) = O_p(\varrho_N N^{-1/4}).$$

For any $\epsilon > 0$, there exists a constant $K_\epsilon > 0$ such that $\Pr\left(\mathrm{d}_{\|\cdot\|_{L^2}}(\hat{\theta}_N, \Theta_I) \leq K_\epsilon \varrho_N N^{-1/4}\right) \geq 1-\epsilon$. Let $\tilde{\Theta}_N = \{\theta : \mathrm{d}_{\|\cdot\|_{L^2}}(\theta, \Theta_I) \leq K_\epsilon \varrho_N N^{-1/4}\}$. Now, we can consider minimization of $Q_N(\theta)$ over $\theta \in \tilde{\Theta}_N$. Using arguments as used in the proof of Theorem 3.1 and the expressions in (A.33)–(A.38) in the proof of Theorem 3.3, we can show that

$$\sup_{\theta \in \tilde{\Theta}_N} |Q_N(\theta) - Q(\theta)| = N^{-1/2} O_p\left(\varrho_N N^{-1/4}\right) = O_p\left(\varrho_N N^{-3/4}\right) \tag{A.27}$$

by showing that $\sup_{\theta \in \tilde{\Theta}_N}\left|\frac{1}{N}\rho_{\ell Ns}\left(\theta^0, \theta - \theta^0\right) - \rho_{\ell s}\left(\theta^0, \theta-\theta^0\right)\right| = O_p\left(\varrho_N N^{-3/4}\right)$ for $\ell = 1,2,3$ and $\sup_{\theta^0 \in \Theta_I}\left|\frac{1}{N}S_N\left(\theta^0\right)\right| = O_P\left(N^{-1}\right)$. The last claim holds by Theorem 3.1(i). Next, we argue that the first claim holds for $\ell = 1$ only as the other two cases can be studied analogously. Note that with $\Delta = \theta - \theta^0$,

$$\begin{aligned}
&\frac{1}{N}\rho_{1Ns}\left(\theta^0, \theta-\theta^0\right) - \rho_{1s}\left(\theta^0, \theta - \theta^0\right) \\
&\quad = -\left\{\frac{1}{N^2}\sum_{1\leq i\neq j\leq N} \partial m_{is}[\Delta]\,\partial m_{js}[\Delta]\,\kappa_{ij,s} - \mathbb{E}\left\{\partial m_s[\Delta]\,\partial m_s^\dagger[\Delta]\left|\underline{z}_s - \underline{z}_s^\dagger\right|\right\}\right\} \\
&\qquad + 2\left\{\frac{1}{N^3}\sum_{1\leq i\neq j\leq N}\sum_{k=1}^N \partial m_{is}[\Delta]\,\kappa_{ij,s}\,\partial m_{ks}[\Delta] - \mathbb{E}\left\{\partial m_s[\Delta]\left|\underline{z}_s - \underline{z}_s^\dagger\right|\right\}\mathbb{E}\left[\partial m_s^\dagger[\Delta]\right]\right\} \\
&\quad \equiv -D_{1s}(\theta) + 2D_{2s}(\theta),
\end{aligned}$$

where we suppress the dependence of $D_{1s}(\theta)$ and $D_{2s}(\theta)$ on $\theta^0$. Let $\varkappa_s\left(\xi_i, \xi_j; \Delta\right) = \partial m_{is}[\Delta]\,\partial m_{js}[\Delta]\,\kappa_{ij,s}$, $c_s(\Delta) = \mathbb{E}_i\mathbb{E}_j[\varkappa_s\left(\xi_i,\xi_j;\Delta\right)]$, and $\varkappa_{1s}(\xi_i;\Delta) = \mathbb{E}_j[\varkappa_s\left(\xi_i,\xi_j;\Delta\right)] - c_s(\Delta)$, where $\mathbb{E}_j$ denotes expectation w.r.t. $\xi_j$ alone. Let $\tilde{\varkappa}_s\left(\xi_i,\xi_j;\Delta\right) = \varkappa_s\left(\xi_i,\xi_j;\Delta\right) - \varkappa_{1s}(\xi_i;\Delta) - \varkappa_{1s}\left(\xi_j;\Delta\right) + c_s(\Delta)$. By Hoeffding decomposition, we have

$$\begin{aligned}
D_{1s}(\theta) &= \frac{2}{N(N-1)}\sum_{1\leq i<j\leq N}\left\{\partial m_{is}[\Delta]\,\partial m_{js}[\Delta]\,\kappa_{ij,s} - \mathbb{E}\left[\partial m_{is}[\Delta]\,\partial m_{js}[\Delta]\,\kappa_{ij,s}\right]\right\} + O_p\left(N^{-1}\right) \\
&= \frac{2}{N}\sum_{i=1}^N \varkappa_{1s}(\xi_i;\Delta) + \frac{2}{N(N-1)}\sum_{1\leq i<j\leq N}\tilde{\varkappa}_s\left(\xi_i,\xi_j;\Delta\right) + O_p\left(N^{-1}\right),
\end{aligned} \tag{A.28}$$

where $O_p\left(N^{-1}\right)$ holds uniformly in $\theta$ (and $\theta^0$). Noting that the second term in (A.28) is a degenerate second-order $U$-process, we can readily follow the proof of Theorem 3.1(i) and show that it is $O_P\left(N^{-1}\right)$ uniformly in $\Delta = \theta - \theta^0 \in \Theta - \theta^0$. For the first term in (A.28), we can apply the expression of $\frac{\partial m_s(Y,X,\theta^0)}{\partial\theta}[\Delta]$ in (A.33) and entropy calculations as used in the proof of Theorem 3.1(i) to show that

$$\begin{aligned}
&\sup_{\theta\in\tilde{\Theta}_N}\left|\frac{1}{N}\sum_{i=1}^{N}\varkappa_{1s}\left(\xi_i;\theta-\theta^0\right)\right| \\
&= \sup_{\theta\in\tilde{\Theta}_N}\left|\frac{1}{N}\sum_{i=1}^{N}\left\{\mathbb{E}_j\left[\partial m_{is}\left[\theta-\theta^0\right]\partial m_{js}\left[\theta-\theta^0\right]\kappa_{ij,s}\right]\right.\right. \\
&\quad \left.\left.-\mathbb{E}_i\mathbb{E}_j\left[\partial m_{is}\left[\theta-\theta^0\right]\partial m_{js}\left[\theta-\theta^0\right]\kappa_{ij,s}\right]\right\}\right| \\
&= N^{-1/2}O_p\left(\varrho_N N^{-1/4}\right) = O_p\left(\varrho_N N^{-3/4}\right).
\end{aligned}$$

Then $\sup_{\theta\in\tilde{\Theta}_N}|D_{1s}(\theta)| = O_p\left(\varrho_N N^{-3/4}\right)$. Analogously, we can show that $\sup_{\theta\in\tilde{\Theta}_N}|D_{2s}(\theta)| = O_p\left(\varrho_N N^{-3/4}\right)$. It follows that $\sup_{\theta\in\tilde{\Theta}_N}\left|\frac{1}{N}\rho_{1Ns}\left(\theta^0,\theta-\theta^0\right)-\rho_{1s}\left(\theta^0,\theta-\theta^0\right)\right| = O_p\left(\varrho_N N^{-3/4}\right)$ and (A.27) holds. Then we can apply Lemma A.5 with $b_N = \varrho_N N^{-3/4}$ to conclude

$$\mathrm{d}_w\left(\hat{\theta}_N,\Theta_I\right) = O_p\left(\max\left\{\delta_{w,N},\varrho_N^{1/2}N^{-3/8}\right\}\right) = O_p(\varrho_N^{1/2}N^{-3/8}). \tag{A.29}$$

Now, given the first iteration result in (A.29), we can focus on $\tilde{\Theta}_N^{(1)} = \{\theta : \mathrm{d}_{\|\cdot\|_{L^2}}(\theta,\Theta_I) \le K_\epsilon\varrho_N\varrho_N^{1/2}N^{-3/8}\}$ and show that

$$\sup_{\theta\in\tilde{\Theta}_N^{(1)}}|Q_N(\theta)-Q(\theta)| = N^{-1/2}O_p\left(\varrho_N\varrho_N^{1/2}N^{-3/8}\right) = O_p\left(\varrho_N^{3/2}N^{-7/8}\right)$$

which, in conjunction with Lemma A.5 implies that

$$\mathrm{d}_w\left(\hat{\theta}_N,\Theta_I\right) = O_p\left(\max\left\{\delta_{w,N},\varrho_N^{3/4}N^{-7/16}\right\}\right) = O_p(\varrho_N^{3/4}N^{-7/16}). \tag{A.30}$$

Repeating such an arguments for any finite $m$ times, we can obtain

$$\mathrm{d}_w\left(\hat{\theta}_N,\Theta_I\right) = N^{-1/4}O_p\left(\left(N^{-1/4}\varrho_N\right)^{\sum_{j=1}^m\frac{1}{2^j}}\right).$$

By choosing $m = m(\epsilon)$ sufficiently large, we obtain $\mathrm{d}_w(\hat{\theta}_N,\Theta_I) = o_p(N^{-\frac{1}{2}+\frac{\epsilon}{4}}\varrho_N^{1-\epsilon})$ for any fixed small positive number $\epsilon > 0$. □

**Proof of Theorem 3.3.** We organize this proof into three parts. In Part I, we establish that, to calculate the test statistic, we only need to minimize over a neighborhood of $\Theta_I\cap\Theta_R$; in Part II, we establish a stochastic upper bound for the test statistic, which still takes the form of a minimization over $\Theta_I\cap\Theta_R$; and in Part III, we derive the asymptotic distribution (under the null) of the stochastic upper bound established in Part II to complete the proof. Note that continuity and compactness imply that all minimums are indeed attained.

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

**Part I.** Pick $\delta_N \downarrow 0$ such that

$$\delta_N = o\left(N^{-1/4}\right) \text{ and } \max\left\{\delta_{s,N}, N^{-1/2}\right\} = o(\delta_N). \quad \textbf{(A.31)}$$

Define $B_N^{\delta_N}\left(\theta^0\right) \equiv \left\{\theta \in \Theta_N \cap \Theta_R : \left\|\theta - \theta^0\right\|_{L^2} \leq \delta_N\right\}$. Then we have

$$\min_{\theta\in\Theta_N\cap\Theta_R} S_N(\theta) \leq \inf_{\theta^0\in\Theta_I\cap\Theta_R}\left[\min_{\theta\in B_N^{\delta_N}(\theta^0)} S_N(\theta)\right] \quad \textbf{(A.32)}$$

by set inclusion, noting the facts that

$$\inf_{\theta^0\in\Theta_I\cap\Theta_R}\left[\min_{\theta\in B_N^{\delta_N}(\theta^0)} S_N(\theta)\right] = \min_{\theta\in \bigcup_{\theta^0\in\Theta_I\cap\Theta_R} B_N^{\delta_N}(\theta^0)} S_N(\theta),$$

and that

$$\bigcup_{\theta^0\in\Theta_I\cap\Theta_R} B_N^{\delta_N}\left(\theta^0\right) \subseteq \Theta_N \cap \Theta_R.$$

To proceed, for the functions $m_s(Y,X,\cdot) : \Theta \to \mathbb{R}$, $s = 1,\ldots,T-R$, we denote by $\frac{\partial m_s(Y,X,\theta^0)}{\partial\theta}[\Delta] \equiv \frac{\partial m_s(Y,X,\theta^0+\tau\Delta)}{\partial\tau}|_{\tau=0}$, and $\frac{\partial^2 m_s(Y,X,\theta^0)}{\partial\theta^2}[\Delta,\Delta] \equiv \frac{\partial^2 m_s(Y,X,\theta^0+\tau\Delta)}{\partial\tau^2}|_{\tau=0}$ its first- and second-order pathwise derivatives at $\theta^0$ in the direction of $\Delta$, respectively. And straightforward pathwise derivative calculations reveal the following facts, which we are going to utilize to complete Part I of the proof:

$$\begin{aligned}
\frac{\partial m_s\left(Y,X,\theta^0\right)}{\partial\theta}[\Delta] &= \sum_{t=1}^{R}\left(\phi_{s,t} - \phi_{s,t}^0\right)\left[y_{T-R+t} - g^0\left(x_{T-R+t}\right)\right] - \left[g(x_s) - g^0(x_s)\right] \\
&\quad - \sum_{t=1}^{R}\phi_{s,t}^0\left[g\left(x_{T-R+t}\right) - g^0\left(x_{T-R+t}\right)\right], \\
\frac{\partial^2 m_s\left(Y,X,\theta^0\right)}{\partial\theta^2}[\Delta,\Delta] &= -2\sum_{t=1}^{R}\left(\phi_{s,t} - \phi_{s,t}^0\right)\left[g\left(x_{T-R+t}\right) - g^0\left(x_{T-R+t}\right)\right], \\
m_s(Y,X,\theta) &= m_s\left(Y,X,\theta^0\right) + \frac{\partial m_s\left(Y,X,\theta^0\right)}{\partial\theta}[\Delta] + \frac{1}{2}\frac{\partial^2 m_s\left(Y,X,\theta^0\right)}{\partial\theta^2}[\Delta,\Delta],
\end{aligned} \quad \textbf{(A.33)}$$

for $\Delta = \theta - \theta^0$.

Continuing with (A.32),

$$\begin{aligned}
&\min_{\theta\in\Theta_N\cap\Theta_R} S_N(\theta) \\
&\leq \inf_{\theta^0\in\Theta_I\cap\Theta_R}\left[\min_{\theta\in B_N^{\delta_N}(\theta^0)} \sum_{s=1}^{T-R} S_{Ns}(\theta)\right]
\end{aligned}$$

$$= \inf_{\theta^0 \in \Theta_I \cap \Theta_R} \left[ \min_{\theta \in B_N^{\delta_N}(\theta^0)} \sum_{s=1}^{T-R} \left[ S_{Ns}\left(\theta^0\right) + \rho_{1Ns}\left(\theta^0, \Delta\right) + \rho_{2Ns}\left(\theta^0, \Delta\right) + \rho_{3Ns}\left(\theta^0, \Delta\right) + \rho_{4Ns}\left(\theta^0, \Delta\right) \right] \right] + o_p(1), \qquad \textbf{(A.34)}$$

where $\Delta = \theta - \theta^0$,

$$\rho_{1Ns}\left(\theta^0, \Delta\right) \equiv -\frac{1}{N} \sum_{1 \le i \ne j \le N} \partial m_{is}[\Delta] \partial m_{js}[\Delta] \kappa_{ij,s} + \frac{2}{N^2} \sum_{1 \le i \ne j \le N} \sum_{k=1}^{N} \partial m_{is}[\Delta] \kappa_{ij,s} \partial m_{ks}[\Delta] - \frac{1}{2N} \sum_{1 \le i \ne j \le N} \left[ m_{is} \partial^2 m_{js}[\Delta] + \partial^2 m_{is}[\Delta] m_{js} \right] \kappa_{ij,s} + \frac{1}{N^2} \sum_{1 \le i \ne j \le N} \sum_{k=1}^{N} \left[ m_{is} \partial^2 m_{ks}[\Delta] + \partial^2 m_{is}[\Delta] m_{ks} \right] \kappa_{ij,s}, \qquad \textbf{(A.35)}$$

$$\rho_{2Ns}\left(\theta^0, \Delta\right) \equiv -\frac{1}{2N} \sum_{1 \le i \ne j \le N} \partial m_{is}[\Delta] \partial^2 m_{js}[\Delta] \kappa_{ij,s} - \frac{1}{2N} \sum_{1 \le i \ne j \le N} \partial^2 m_{is}[\Delta] \partial m_{js}[\Delta] \kappa_{ij,s} + \frac{1}{N^2} \sum_{1 \le i \ne j \le N} \sum_{k=1}^{N} \left[ \partial m_{is}[\Delta] \partial^2 m_{ks}[\Delta] + \partial^2 m_{is}[\Delta] \partial m_{ks}[\Delta] \right] \kappa_{ij,s}, \qquad \textbf{(A.36)}$$

$$\rho_{3Ns}\left(\theta^0, \Delta\right) \equiv -\frac{1}{4N} \sum_{1 \le i \ne j \le N} \partial^2 m_{is}[\Delta] \partial^2 m_{js}[\Delta] \kappa_{ij,s} + \frac{1}{2N^2} \sum_{1 \le i \ne j \le N} \sum_{k=1}^{N} \partial^2 m_{is}[\Delta] \kappa_{ij,s} \partial^2 m_{ks}[\Delta], \qquad \textbf{(A.37)}$$

$$\rho_{4Ns}\left(\theta^0, \Delta\right) \equiv -\frac{1}{N} \sum_{1 \le i \ne j \le N} \left[ m_{is} \partial m_{js}[\Delta] + \partial m_{is}[\Delta] m_{js} \right] \kappa_{ij,s} + \frac{2}{N^2} \sum_{1 \le i \ne j \le N} \sum_{k=1}^{N} [m_{is} \partial m_{ks}[\Delta] + \partial m_{is}[\Delta] m_{ks}] \kappa_{ij,s}, \qquad \textbf{(A.38)}$$

with $m_{is} \equiv m_{is}\left(\theta^0\right) \equiv m_s\left(Y_i, X_i, \theta^0\right)$, $\partial m_{is}[\Delta] \equiv \frac{\partial m_s(Y_i, X_i, \theta^0)}{\partial \theta}[\Delta]$, and $\partial^2 m_{is}[\Delta] \equiv \frac{\partial^2 m_s(Y_i, X_i, \theta^0)}{\partial \theta^2}[\Delta, \Delta]$. Note that the last equality in (A.34) is obtained by plugging the third equation in ( A.33) (with $\Delta = \theta - \theta^0$) into

$$S_{Ns}(\theta) = \frac{1}{N} \sum_{1 \le i \ne j \le N} m_s(Y_i, X_i, \theta) m_s\left(Y_j, X_j, \theta\right) \kappa_{ij,s} + \frac{2}{N^2} \sum_{1 \le i \ne j \le N} m_s(Y_i, X_i, \theta) \kappa_{ij,s} \sum_{k=1}^{N} m_s(Y_k, X_k, \theta).$$

Let $m_s \equiv m_s\left(Y, X, \theta^0\right)$, $\partial m_s[\Delta] \equiv \frac{\partial m_s(Y, X, \theta^0)}{\partial \theta}[\Delta]$, and $\partial^2 m_s[\Delta] \equiv \frac{\partial^2 m_s(Y, X, \theta^0)}{\partial \theta^2}[\Delta, \Delta]$. Define $m_s^\dagger$, $\partial m_s^\dagger[\Delta]$, and $\partial^2 m_s^\dagger[\Delta]$ analogously with $(Y, X)$ replaced by their independent copy $(Y^{\dagger}, X^{\dagger})$. The population analogues of $N^{-1}\rho_{\ell Ns}\left(\theta^0,\Delta\right)$, $\ell = 1,2,3,4$, are respectively given by

$$\begin{aligned}
\rho_{1s}\left(\theta^0,\Delta\right) \equiv& -\mathbb{E}\left\{\partial m_s[\Delta]\partial m_s^{\dagger}[\Delta]\left|\underline{z}_s-\underline{z}_s^{\dagger}\right|\right\}+2\mathbb{E}\left\{\partial m_s[\Delta]\left|\underline{z}_s-\underline{z}_s^{\dagger}\right|\right\}\mathbb{E}\left[\partial m_s^{\dagger}[\Delta]\right] \\
&-\frac{1}{2}\mathbb{E}\left\{\left[m_s\partial^2 m_s^{\dagger}[\Delta]+\partial^2 m_s[\Delta]m_s^{\dagger}\right]\left|\underline{z}_s-\underline{z}_s^{\dagger}\right|\right\} \\
&+\mathbb{E}\left\{m_s\left|\underline{z}_s-\underline{z}_s^{\dagger}\right|\right\}\mathbb{E}\left\{\partial^2 m_s^{\dagger}[\Delta]\right\}+\mathbb{E}\left\{\partial^2 m_s[\Delta]\left|\underline{z}_s-\underline{z}_s^{\dagger}\right|\right\}\mathbb{E}[m_s^{\dagger}], \\
\rho_{2s}\left(\theta^0,\Delta\right) \equiv& -\frac{1}{2}\mathbb{E}\left\{\partial m_s[\Delta]\partial^2 m_s^{\dagger}[\Delta]\left|\underline{z}_s-\underline{z}_s^{\dagger}\right|\right\}-\frac{1}{2}\mathbb{E}\left\{\partial^2 m_s[\Delta]\partial m_s^{\dagger}[\Delta]\left|\underline{z}_s-\underline{z}_s^{\dagger}\right|\right\} \\
&+\mathbb{E}\left\{\partial m_s[\Delta]\left|\underline{z}_s-\underline{z}_s^{\dagger}\right|\right\}\mathbb{E}\left[\partial^2 m_s^{\dagger}[\Delta]\right]+\mathbb{E}\left\{\partial^2 m_s[\Delta]\left|\underline{z}_s-\underline{z}_s^{\dagger}\right|\right\}\mathbb{E}\left[\partial m_s^{\dagger}[\Delta]\right], \\
\rho_{3s}\left(\theta^0,\Delta\right) \equiv& -\frac{1}{4}\mathbb{E}\left\{\partial^2 m_s[\Delta]\partial^2 m_s^{\dagger}[\Delta]\left|\underline{z}_s-\underline{z}_s^{\dagger}\right|\right\}+\frac{1}{2}\mathbb{E}\left\{\partial^2 m_s[\Delta]\left|\underline{z}_s-\underline{z}_s^{\dagger}\right|\right\}\mathbb{E}\left[\partial^2 m_s^{\dagger}[\Delta]\right], \\
\rho_{4s}\left(\theta^0,\Delta\right) \equiv& -\mathbb{E}\left\{\left[m_s\partial m_s^{\dagger}[\Delta]+\partial m_s[\Delta]m_s^{\dagger}\right]\left|\underline{z}_s-\underline{z}_s^{\dagger}\right|\right\} \\
&+2\mathbb{E}\left\{m_s\left|\underline{z}_s-\underline{z}_s^{\dagger}\right|\right\}\mathbb{E}\left\{\partial m_s^{\dagger}[\Delta]]\right\}+2\mathbb{E}\left\{\partial m_s[\Delta]\left|\underline{z}_s-\underline{z}_s^{\dagger}\right|\right\}\mathbb{E}[m_s^{\dagger}].
\end{aligned}$$

Then we can verify algebraically that

$$\begin{aligned}
\sum_{\ell=1}^{4}\rho_{\ell s}\left(\theta^0,\theta-\theta^0\right) &= \sum_{\ell=1}^{4}\rho_{\ell s}\left(\theta^0,\theta-\theta^0\right)+\mathrm{MDD}\left[m_s\left(Y,X,\theta^0\right)|\underline{z}_s\right]^2 \\
&= \mathrm{MDD}\left\{m_s\left(Y,X,\theta^0\right)+\left[m_s(Y,X,\theta)-m_s\left(Y,X,\theta^0\right)\right]|\underline{z}_s\right\}^2 \\
&= \mathrm{MDD}\left[m_s(Y,X,\theta)|\underline{z}_s\right]^2, \qquad \textbf{(A.39)}
\end{aligned}$$

where the first equality follows from the fact that $\mathrm{MDD}\left[m_s\left(Y,X,\theta^0\right)|\underline{z}_s\right]^2 = 0$, and the second equality is verified by using the fact that

$$\mathrm{MDD}(A+B|W)^2 = -\mathbb{E}\left[(A+B)(A^{\dagger}+B^{\dagger})\left|W-W^{\dagger}\right|\right]+2\mathbb{E}\left[(A+B)\left|W-W^{\dagger}\right|\right]\mathbb{E}(A^{\dagger}+B^{\dagger})$$

with $A = m_s\left(Y,X,\theta^0\right)$, $B = m_s(Y,X,\theta)-m_s\left(Y,X,\theta^0\right)$, $W = \underline{z}_s$, and $(A^{\dagger},B^{\dagger},W^{\dagger})$ being the independent copy of $(A,B,W)$.

Using arguments analogous to those in the proof of Theorem 3.1(ii), we can show that for any given $\theta^0 \in \Theta_I$,

$$\frac{1}{N}\rho_{1Ns}\left(\theta^0,\Delta\right) = \rho_{1s}\left(\theta^0,\Delta\right)+O_p\left(N^{-1/2}\right)$$

uniformly in $\Delta = \theta-\theta^0 \in \Theta-\theta^0 \equiv \left\{\tilde{\theta}-\theta^0 : \tilde{\theta}\in\Theta\right\}$. It follows that $\sqrt{N}\left[\frac{1}{N}\rho_{1Ns}\left(\theta^0,\Delta\right)-\rho_{1s}\left(\theta^0,\Delta\right)\right] = O_p(1)$ uniformly in $\Delta = \theta-\theta^0\in\Theta-\theta^0$, which in turn implies the uniform asymptotic $\|\cdot\|_{L^2}$-equicontinuity in probability in the sense that $\forall\,\epsilon_N \downarrow 0$, it holds that

$$\sup_{\|\theta_1-\theta_2\|_{L^2}\le\epsilon_N}\sqrt{N}\left|\left[\frac{1}{N}\rho_{1Ns}\left(\theta^0,\theta_1\right)-\rho_{1s}\left(\theta^0,\theta_1\right)\right]-\left[\frac{1}{N}\rho_{1Ns}\left(\theta^0,\theta_2\right)-\rho_{1s}\left(\theta^0,\theta_2\right)\right]\right| = o_p(1). \qquad \textbf{(A.40)}$$

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

It is easy to see algebraically that $a\rho_{1Ns}\left(\theta^{0},\Delta\right)=\rho_{1Ns}\left(\theta^{0},a^{\frac{1}{2}}\Delta\right)$ and $a\rho_{1s}\left(\theta^{0},\Delta\right)=\rho_{1s}\left(\theta^{0},a^{\frac{1}{2}}\Delta\right)$ for any scalar constant $a\geq 0$. These algebraic properties imply that

$$\sup_{\|\theta_1-\theta_2\|_{L^2}\leq\delta_N}\left|\left[\rho_{1Ns}\left(\theta^{0},\theta_1\right)-N\rho_{1s}\left(\theta^{0},\theta_1\right)\right]-\left[\rho_{1Ns}\left(\theta^{0},\theta_2\right)-N\rho_{1s}\left(\theta^{0},\theta_2\right)\right]\right|$$
$$=\sup_{\left\|N^{\frac{1}{4}}\theta_1-N^{\frac{1}{4}}\theta_2\right\|_{L^2}\leq N^{\frac{1}{4}}\delta_N}\sqrt{N}\left|\left[\frac{1}{N}\rho_{1Ns}(\theta^{0},N^{\frac{1}{4}}\theta_1)-\rho_{1s}(\theta^{0},N^{\frac{1}{4}}\theta_1)\right]\right.$$
$$\left.-\left[\frac{1}{N}\rho_{1Ns}(\theta^{0},N^{\frac{1}{4}}\theta_2)-\rho_{1s}(\theta^{0},N^{\frac{1}{4}}\theta_2)\right]\right|=o_p(1),\tag{A.41}$$

where the last equality holds because $N^{\frac{1}{4}}\delta_N\downarrow 0$ according to (A.31). This, in conjunction with the fact that $\rho_{1Ns}\left(\theta^{0},0\right)=\rho_{1s}\left(\theta^{0},0\right)=0$, implies that

$$\sup_{\theta:\,\|\theta-\theta^{0}\|_{L^2}\leq\delta_N}\left|\rho_{1Ns}\left(\theta^{0},\theta-\theta^{0}\right)-N\rho_{1s}\left(\theta^{0},\theta-\theta^{0}\right)\right|=o_p(1).\tag{A.42}$$

Analogously to the analysis of $\rho_{1Ns}\left(\theta^{0},\Delta\right)$, it can be shown that

$$\sqrt{N}\left[\frac{1}{N}\rho_{2Ns}\left(\theta^{0},\Delta\right)-\rho_{2s}\left(\theta^{0},\Delta\right)\right]=O_p(1)\text{ uniformly in }\Delta=\theta-\theta^{0}\in\Theta-\theta^{0}\text{ and}$$
$$\sup_{\theta:\,\|\theta-\theta^{0}\|_{L^2}\leq\delta_N}\left|\rho_{2Ns}\left(\theta^{0},\theta-\theta^{0}\right)-N\rho_{2s}\left(\theta^{0},\theta-\theta^{0}\right)\right|=o_p(1),\tag{A.43}$$

where we use the fact that $a\rho_{2Ns}\left(\theta^{0},\Delta\right)=\rho_{2Ns}\left(\theta^{0},a^{1/3}\Delta\right)$ and $a\rho_{2s}\left(\theta^{0},\Delta\right)=\rho_{2s}\left(\theta^{0},a^{1/3}\Delta\right)$ for any scalar constant $a\geq 0$, $\rho_{2Ns}\left(\theta^{0},0\right)=\rho_{2s}\left(\theta^{0},0\right)=0$, and that $N^{\frac{1}{6}}\delta_N\downarrow 0$. And it can be shown that

$$\sqrt{N}\left[\frac{1}{N}\rho_{3Ns}\left(\theta^{0},\Delta\right)-\rho_{3s}\left(\theta^{0},\Delta\right)\right]=O_p(1)\text{ uniformly in }\Delta=\theta-\theta^{0}\in\Theta-\theta^{0}\text{ and}$$
$$\sup_{\theta:\,\|\theta-\theta^{0}\|_{L^2}\leq\delta_N}\left|\rho_{3Ns}\left(\theta^{0},\theta-\theta^{0}\right)-N\rho_{3s}\left(\theta^{0},\theta-\theta^{0}\right)\right|=o_p(1),\tag{A.44}$$

where we use the fact that $a\rho_{3Ns}\left(\theta^{0},\Delta\right)=\rho_{3Ns}\left(\theta^{0},a^{1/4}\Delta\right)$ and $a\rho_{3s}\left(\theta^{0},\Delta\right)=\rho_{3s}\left(\theta^{0},a^{1/4}\Delta\right)$ for any scalar constant $a\geq 0$, $\rho_{3Ns}\left(\theta^{0},0\right)=\rho_{3s}\left(\theta^{0},0\right)=0$, and that $N^{\frac{1}{8}}\delta_N\downarrow 0$.

To recap, up to this point, we have established that,

$$\left|\rho_{lNs}\left(\theta^{0},\theta-\theta^{0}\right)-N\rho_{ls}\left(\theta^{0},\theta-\theta^{0}\right)\right|=o_p(1),\text{ uniformly in }\theta\in B_N^{\delta_N}\left(\theta^{0}\right),$$

for $l = 1,2,3$. These results validate our passing from $\rho_{lNs}$ to $N\rho_{ls}$, for $l = 1,2,3$, in $\theta \in B_N^{\delta_N}\left(\theta^0\right)$ later. However, it would be invalid to pass from $\rho_{4Ns}$ to $N\rho_{4s}$ in $\theta \in B_N^{\delta_N}\left(\theta^0\right)$ in the same way, because

$$\begin{aligned}
&\sup_{\theta:\ \|\theta-\theta^0\|_{L^2}\le\delta_N} \left|\rho_{4Ns}\left(\theta^0,\theta-\theta^0\right) - N\rho_{4s}\left(\theta^0,\theta-\theta^0\right)\right| \\
&= \sup_{\theta:\ \|\theta-\theta^0\|_{L^2}\le\delta_N} \sqrt{N}\left|\frac{1}{N}\rho_{4Ns}\left(\theta^0,N^{1/2}\left(\theta-\theta^0\right)\right) - N\rho_{4s}\left(\theta^0,N^{1/2}\left(\theta-\theta^0\right)\right)\right| \\
&= \sup_{\theta:\ \|N^{1/2}(\theta-\theta^0)\|_{L^2}\le N^{1/2}\delta_N} \sqrt{N}\left|\frac{1}{N}\rho_{4Ns}\left(\theta^0,N^{1/2}\left(\theta-\theta^0\right)\right) - N\rho_{4s}\left(\theta^0,N^{1/2}\left(\theta-\theta^0\right)\right)\right| \\
&\neq o_p(1), \qquad \textbf{(A.45)}
\end{aligned}$$

where the equality in (A.45) follows from the fact that $a\rho_{4Ns}\left(\theta^0,\Delta\right) = \rho_{4Ns}\left(\theta^0,a\Delta\right)$ and $a\rho_{4s}\left(\theta^0,\Delta\right) = \rho_{4s}\left(\theta^0,a\Delta\right)$ for any scalar constant $a \ge 0$. In the final step of (A.45), one is unable to establish the subject to be $o_p(1)$ because there is no guarantee that $N^{1/2}\delta_N = o_p(1)$, that is, $\delta_N$ does not shrink to 0 fast enough (it is only $o_p\left(N^{-1/4}\right)$, as specified in (A.31)).

Combining (A.39) and (A.42)–(A.44), we have

$$\begin{aligned}
\sum_{\ell=1}^{4}\rho_{\ell Ns}\left(\theta^0,\theta-\theta^0\right) &= N\sum_{\ell=1}^{4}\rho_{\ell s}\left(\theta^0,\theta-\theta^0\right) + \left[\rho_{4Ns}\left(\theta^0,\theta-\theta^0\right) - N\rho_{4s}\left(\theta^0,\theta-\theta^0\right)\right] + o_p(1) \\
&= N\cdot \mathrm{MDD}\left[m_s(Y,X,\theta)\,|\underline{z}_s\right]^2 + \widetilde{\rho}_{4Ns}\left(\theta^0,\theta-\theta^0\right) + o_p(1), \qquad \textbf{(A.46)}
\end{aligned}$$

where each $o_p(1)$ holds uniformly in $\theta \in B_N^{\delta_N}\left(\theta^0\right)$, and

$$\widetilde{\rho}_{4Ns}\left(\theta^0,\theta-\theta^0\right) \equiv \rho_{4Ns}\left(\theta^0,\theta-\theta^0\right) - N\rho_{4s}\left(\theta^0,\theta-\theta^0\right).$$

Plugging (A.46) in (A.34) yields

$$\begin{aligned}
&\min_{\theta\in\Theta_N\cap\Theta_R} S_N(\theta) \\
&\le \inf_{\theta^0\in\Theta_I\cap\Theta_R}\left[\min_{\theta\in B_N^{\delta_N}(\theta^0)}\sum_{s=1}^{T-R}\left[S_{Ns}\left(\theta^0\right) + N\cdot\mathrm{MDD}\left[m_s(Y,X,\theta)\,|\underline{z}_s\right]^2 + \widetilde{\rho}_{4Ns}\left(\theta^0,\theta-\theta^0\right)\right]\right] \\
&\quad + o_p(1) \\
&= \inf_{\theta^0\in\Theta_I\cap\Theta_R}\left[S_N\left(\theta^0\right) + \min_{\theta\in B_N^{\delta_N}(\theta^0)}\sum_{s=1}^{T-R}\left[N\cdot\mathrm{MDD}\left[m_s(Y,X,\theta)\,|\underline{z}_s\right]^2 + \widetilde{\rho}_{4Ns}\left(\theta^0,\theta-\theta^0\right)\right]\right] \\
&\quad + o_p(1). \qquad \textbf{(A.47)}
\end{aligned}$$

**Part II.** In this part, we establish a stochastic upper bound for $\min_{\theta\in\Theta_N\cap\Theta_R} S_N(\theta)$, by examining smaller neighborhoods of $\theta^0$ (compared with $B_N^{\delta_N}\left(\theta^0\right)$), namely $B_N^{\widetilde{\delta}_N}\left(\theta^0\right)$.

Specifically, recalling that $\delta_{s,N} = \sup_{\theta \in \Theta_I \cap \Theta_R} \left\| \Pi_N \theta - \theta \right\|_{L^2} = o\left(N^{-1/2}\right)$ by Assumption 3.3(i), we pick a positive sequence $\tilde{\delta}_N \downarrow 0$ such that

$$\tilde{\delta}_N \leq \delta_N, \;\; \tilde{\delta}_N = o\left(N^{-1/2}\right), \text{ and } \delta_{s,N} = o\left(\tilde{\delta}_N\right). \tag{A.48}$$

Let $B_N^{\tilde{\delta}_N}\left(\theta^0\right) \equiv \left\{\theta \in \Theta_N \cap \Theta_R : \left\|\theta - \theta^0\right\|_{L^2} \leq \tilde{\delta}_N\right\}$. It is helpful to note that $B_N^{\tilde{\delta}_N}\left(\theta^0\right)$ defined similarly to $B_N^{\delta_N}\left(\theta^0\right)$, but with a smaller radius $\tilde{\delta}_N$.

Analogously to the analysis of $\rho_{1Ns}\left(\theta^0, \Delta\right)$ shown in (A.40)–(A.42), it can be shown that

$$\sqrt{N}\left[\frac{1}{N}\rho_{4Ns}\left(\theta^0, \Delta\right) - \rho_{4s}\left(\theta^0, \Delta\right)\right] = O_p(1) \text{ uniformly in } \Delta = \theta - \theta^0 \in \Theta - \theta^0,$$

and

$$\begin{aligned}
&\sup_{\theta:\, \left\|\theta - \theta^0\right\|_{L^2} \leq \tilde{\delta}_N} \left|\widetilde{\rho}_{4Ns}\left(\theta^0, \theta - \theta^0\right)\right| \\
&= \sup_{\theta:\, \left\|\theta - \theta^0\right\|_{L^2} \leq \tilde{\delta}_N} \left|\rho_{4Ns}\left(\theta^0, \theta - \theta^0\right) - N\rho_{4s}\left(\theta^0, \theta - \theta^0\right)\right| \\
&= \sup_{\theta:\, \left\|\theta - \theta^0\right\|_{L^2} \leq \tilde{\delta}_N} \sqrt{N}\left|\frac{1}{N}\rho_{4Ns}\left(\theta^0, \sqrt{N}\left(\theta - \theta^0\right)\right) - N\rho_{4s}\left(\theta^0, \sqrt{N}\left(\theta - \theta^0\right)\right)\right| \\
&\leq \sup_{\|\Delta\|_{L^2} \leq N^{\frac{1}{2}}\tilde{\delta}_N} \sqrt{N}\left|\frac{1}{N}\rho_{4Ns}\left(\theta^0, \Delta\right) - N\rho_{4s}\left(\theta^0, \Delta\right)\right| \\
&= o_p(1),
\end{aligned} \tag{A.49}$$

where we use the fact that $a\rho_{4Ns}\left(\theta^0, \Delta\right) = \rho_{4Ns}\left(\theta^0, a\Delta\right)$ and $a\rho_{4s}\left(\theta^0, \Delta\right) = \rho_{4s}\left(\theta^0, a\Delta\right)$ for any scalar constant $a \geq 0$, $\rho_{4Ns}\left(\theta^0, 0\right) = \rho_{4s}\left(\theta^0, 0\right) = 0$, and that $N^{\frac{1}{2}}\tilde{\delta}_N \downarrow 0$.

Continuing with (A.47),

$$\begin{aligned}
&\min_{\theta \in \Theta_N \cap \Theta_R} S_N(\theta) \\
&\leq \inf_{\theta^0 \in \Theta_I \cap \Theta_R}\left[S_N\left(\theta^0\right) + \min_{\theta \in B_N^{\delta_N}(\theta^0)} \sum_{s=1}^{T-R}\left[N \cdot \mathrm{MDD}\left[m_s(Y, X, \theta) \,|\, \underline{z}_s\right]^2 + \widetilde{\rho}_{4Ns}\left(\theta^0, \theta - \theta^0\right)\right]\right] \\
&\quad + o_p(1). \\
&\leq \inf_{\theta^0 \in \Theta_I \cap \Theta_R}\left[S_N\left(\theta^0\right) + \min_{\theta \in B_N^{\tilde{\delta}_N}(\theta^0)} \sum_{s=1}^{T-R}\left[N \cdot \mathrm{MDD}\left[m_s(Y, X, \theta) \,|\, \underline{z}_s\right]^2 + \widetilde{\rho}_{4Ns}\left(\theta^0, \theta - \theta^0\right)\right]\right] \\
&\quad + o_p(1)
\end{aligned}$$

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

$$= \inf_{\theta^0 \in \Theta_I \cap \Theta_R} \left[ S_N\left(\theta^0\right) + \min_{\theta \in B_N^{\tilde{\delta}_N}(\theta^0)} N \sum_{s=1}^{T-R} \text{MDD}\left[m_s(Y,X,\theta)\,|\underline{z}_s\right]^2 \right] + o_p(1), \qquad \textbf{(A.50)}$$

where the inequality in (A.50) follows from that $\tilde{\delta}_N \leq \delta_N$ as specified in (A.48), and the last step in (A.50) follows from (A.49).

**Part III.** Next, we first show that the term $\min_{\theta \in B_N^{\tilde{\delta}_N}(\theta^0)} N \sum_{s=1}^{T-R} \text{MDD}\left[m_s(Y,X,\theta)\,|\underline{z}_s\right]^2$ in (A.50) can be dropped asymptotically and then derive the asymptotic null distribution of $\inf_{\theta^0 \in \Theta_I \cap \Theta_R} S_N\left(\theta^0\right)$. The non-negativity of MDD implies that

$$\min_{\theta \in B_N^{\tilde{\delta}_N}(\theta^0)} N \sum_{s=1}^{T-R} \text{MDD}\left[m_s(Y,X,\theta)\,|\underline{z}_s\right]^2 \geq 0. \qquad \textbf{(A.51)}$$

According to the third condition in (A.48), that is, $\sup_{\theta \in \Theta_I \cap \Theta_R} \left\| \Pi_N \theta - \theta \right\|_{L^2} = o\left(\tilde{\delta}_N\right)$, it holds that $\left\| \Pi_N \theta^0 - \theta^0 \right\|_{L^2} \leq \tilde{\delta}_N$, or equivalently, $\Pi_N \theta^0 \in B_N^{\tilde{\delta}_N}\left(\theta^0\right)$ for all $\theta^0 \in \Theta_I \cap \Theta_R$ and large enough $N$. Therefore,

$$\min_{\theta \in B_N^{\tilde{\delta}_N}(\theta^0)} N \sum_{s=1}^{T-R} \text{MDD}\left[m_s(Y,X,\theta)\,|\underline{z}_s\right]^2 \leq N \sum_{s=1}^{T-R} \text{MDD}\left[m_s\left(Y,X,\Pi_N\theta^0\right)|\underline{z}_s\right]^2 \qquad \textbf{(A.52)}$$

for all $\theta^0 \in \Theta_I \cap \Theta_R$ and large enough $N$. Then for all $\theta^0 \in \Theta_I \cap \Theta_R$,

$$\begin{aligned}
0 &\leq N \sum_{s=1}^{T-R} \text{MDD}\left[m_s\left(Y,X,\Pi_N\theta^0\right)|\underline{z}_s\right]^2 \\
&= N \sum_{s=1}^{T-R} \text{MDD}\left[m_s\left(Y,X,\Pi_N\theta^0\right) - m_s\left(Y,X,\theta^0\right)|\underline{z}_s\right]^2 \\
&= N \text{d}_w\left(\Pi_N\theta^0,\theta^0\right)^2 \leq \left[N^{1/2} \sup_{\theta^0 \in \Theta_I \cap \Theta_R} \text{d}_w\left(\Pi_N\theta^0,\theta^0\right)\right]^2 = o(1),
\end{aligned}$$

where the first equality follows from Lemma A.1 and the fact that $\text{MDD}\left[m_s\left(Y,X,\theta^0\right)|\underline{z}_s\right]^2 = 0$, and the last equality follows from Assumption 3.3(ii). It follows that

$$\sup_{\theta^0 \in \Theta_I \cap \Theta_R} N \sum_{s=1}^{T-R} \text{MDD}\left[m_s\left(Y,X,\Pi_N\theta^0\right)|\underline{z}_s\right]^2 = o(1). \qquad \textbf{(A.53)}$$

Combining (A.51)–(A.53) yields

$$\sup_{\theta^0 \in \Theta_I \cap \Theta_R} \min_{\theta \in B_N^{\tilde{\delta}_N}(\theta^0)} N \sum_{s=1}^{T-R} \text{MDD}\left[m_s(Y,X,\theta)\,|\underline{z}_s\right]^2 = o(1). \qquad \textbf{(A.54)}$$

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

As a result, substituting (A.54) into (A.50) delivers

$$\begin{aligned}
&\min_{\theta\in\Theta_N\cap\Theta_R} S_N(\theta) \\
&\quad \le \inf_{\theta^0\in\Theta_I\cap\Theta_R}\left\{S_N\left(\theta^0\right)+\min_{\theta\in B_N^{\tilde{\delta}_N}(\theta^0)} N\left[\sum_{s=1}^{T-R}\mathrm{MDD}\left[m_s(Y,X,\theta)\,|\underline{z}_s\right]^2\right]\right\}+o_p(1) \\
&\quad = \inf_{\theta^0\in\Theta_I\cap\Theta_R} S_N\left(\theta^0\right)+o_p(1). \qquad \textbf{(A.55)}
\end{aligned}$$

Now, we show that if we write $\min_{\theta\in\Theta_N\cap\Theta_R} S_N(\theta)=\inf_{\theta^0\in\Theta_I\cap\Theta_R} S_N\left(\theta^0\right)-c_N+o_p(1)$, we must have $0\le c_N\le \bar{c}_N$, where

$$\bar{c}_N \equiv -\inf_{\theta^0\in\Theta_I\cap\Theta_R}\min_{\theta\in B_N^{\delta_N}(\theta^0)}\sum_{s=1}^{T-R}\widetilde{\rho}_{4Ns}\left(\theta^0,\theta-\theta^0\right). \qquad \textbf{(A.56)}$$

Let $II\left(\theta^0\right)=\min_{\theta\in B_N^{\delta_N}(\theta^0)}\sum_{s=1}^{T-R} N\cdot\mathrm{MDD}\left[m_s(Y,X,\theta)\,|\underline{z}_s\right]^2$. As above, we can also show that $\sup_{\theta^0\in\Theta_I\cap\Theta_R} II\left(\theta^0\right)=o_p(1)$. Note that

$$\begin{aligned}
&\min_{\theta\in\Theta_N\cap\Theta_R} S_N(\theta) \\
&\quad = \inf_{\theta^0\in\Theta_I\cap\Theta_R}\left[S_N\left(\theta^0\right)+\min_{\theta\in B_N^{\delta_N}(\theta^0)}\sum_{s=1}^{T-R}\left[N\cdot\mathrm{MDD}\left[m_s(Y,X,\theta)\,|\underline{z}_s\right]^2+\widetilde{\rho}_{4Ns}\left(\theta^0,\theta-\theta^0\right)\right]\right] \\
&\qquad + o_p(1) \\
&\quad \ge \inf_{\theta^0\in\Theta_I\cap\Theta_R}\left[S_N\left(\theta^0\right)+II\left(\theta^0\right)+\min_{\theta\in B_N^{\delta_N}(\theta^0)}\sum_{s=1}^{T-R}\widetilde{\rho}_{4Ns}\left(\theta^0,\theta-\theta^0\right)\right]+o_p(1) \\
&\quad = \inf_{\theta^0\in\Theta_I\cap\Theta_R}\left[S_N\left(\theta^0\right)+\min_{\theta\in B_N^{\delta_N}(\theta^0)}\sum_{s=1}^{T-R}\widetilde{\rho}_{4Ns}\left(\theta^0,\theta-\theta^0\right)\right]+o_p(1) \\
&\quad \ge \inf_{\theta^0\in\Theta_I\cap\Theta_R} S_N\left(\theta^0\right)+\inf_{\theta^0\in\Theta_I\cap\Theta_R}\min_{\theta\in B_N^{\delta_N}(\theta^0)}\sum_{s=1}^{T-R}\widetilde{\rho}_{4Ns}\left(\theta^0,\theta-\theta^0\right)+o_p(1) \\
&\quad = \inf_{\theta^0\in\Theta_I\cap\Theta_R} S_N\left(\theta^0\right)-\bar{c}_N+o_p(1).
\end{aligned}$$

This implies that $0\le c_N\le\bar{c}_N$.

Lastly, by Theorem 3.1(i) and the extended continuous mapping (see, e.g., Theorem 1.11.1 in van der Vaart and Wellner (1996)), we have

$$\inf_{\theta^0\in\Theta_I\cap\Theta_R} S_N\left(\theta^0\right)\xrightarrow{\mathcal{L}}\inf_{\theta^0\in\Theta_I\cap\Theta_R}\sum_{s=1}^{T-R}\left[\mathbb{B}_s\left(\theta^0\right)+\mathbb{C}_s\left(\theta^0\right)\right],$$

which, in conjunction with (A.55), complete the proof. □

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

**Proof of Theorem 3.4.** Let $\delta_N$ be the same as the one specified by (A.31) in the proof of Theorem 3.3. By Theorem 3.1,

$$\sup_{\theta\in\Theta}\left|\frac{1}{N}S_N(\theta)-\sum_{s=1}^{T-R}\mathrm{MDD}\left[m_s(Y,X,\theta)\,|\underline{z}_s\right]^2\right|\xrightarrow{p}0. \tag{A.57}$$

Since $\Theta$ is compact under $\|\cdot\|_c$ by Lemma A.3, it follows from the theorem of the maximum and the continuous mapping theorem that

$$\min_{\theta\in\Theta\cap\Theta_R}\frac{1}{N}S_N(\theta)=\min_{\theta\in\Theta\cap\Theta_R}\sum_{s=1}^{T-R}\mathrm{MDD}\left[m_s(Y,X,\theta)\,|\underline{z}_s\right]^2+o_p(1). \tag{A.58}$$

Let $\tilde{\theta}_N\in\operatorname*{argmin}_{\theta\in\Theta\cap\Theta_R}S_N(\theta)$. Note that

$$\begin{aligned}
0&\le\min_{\theta\in\Theta_N\cap\Theta_R}\frac{1}{N}S_N(\theta)-\min_{\theta\in\Theta\cap\Theta_R}\frac{1}{N}S_N(\theta)\\
&\le\frac{1}{N}S_N\left(\Pi_N\tilde{\theta}_N\right)-\frac{1}{N}S_N\left(\tilde{\theta}_N\right)\\
&=\left|\sum_{s=1}^{T-R}\mathrm{MDD}\left[m_s\left(Y,X,\Pi_N\tilde{\theta}_N\right)|\underline{z}_s\right]^2-\sum_{s=1}^{T-R}\mathrm{MDD}\left[m_s\left(Y,X,\tilde{\theta}_N\right)|\underline{z}_s\right]^2\right|+o_p(1)\\
&\preceq\sum_{s=1}^{T-R}\left\{\mathrm{MDD}\left[m_s\left(Y,X,\Pi_N\tilde{\theta}_N\right)-m_s\left(Y,X,\tilde{\theta}_N\right)|\underline{z}_s\right]^2\right\}^{1/2}+o_p(1)\\
&\le(T-R)^{1/2}\left\{\sum_{s=1}^{T-R}\mathrm{MDD}\left[m_s\left(Y,X,\Pi_N\tilde{\theta}_N\right)-m_s\left(Y,X,\tilde{\theta}_N\right)|\underline{z}_s\right]^2\right\}^{1/2}+o_p(1)\\
&\preceq\left\|\Pi_N\tilde{\theta}_N-\tilde{\theta}_N\right\|_{L^2}+o_p(1)\\
&=o_p(1),
\end{aligned} \tag{A.59}$$

where the first equality holds because of the uniform asymptotic $\|\cdot\|_{L^2}$-equicontinuity of $\frac{1}{N}S_N(\theta)-\sum_{s=1}^{T-R}\mathrm{MDD}\left[m_s(Y,X,\theta)\,|\underline{z}_s\right]^2$ (implied by (A.57) or by Theorem 3.1 directly) and the fact that $\left\|\Pi_N\tilde{\theta}_N-\tilde{\theta}_N\right\|_{L^2}=o(\delta_N)$ with $\delta_N\downarrow 0$, the third inequality follows from Lemma A.2,[13] the fourth inequality holds by Cauchy–Schwarz inequality, and the last inequality holds by Lemma 3.1.

(A.58) and (A.59) imply

$$\frac{1}{N}\hat{S}_N=\min_{\theta\in\Theta_N\cap\Theta_R}\frac{1}{N}S_N(\theta)=\min_{\theta\in\Theta\cap\Theta_R}\sum_{s=1}^{T-R}\mathrm{MDD}\left[m_s(Y,X,\theta)\,|\underline{z}_s\right]^2+o_p(1).$$

[13]The finite moments requirement in Lemma A.2 can be easily verified by Assumption 3.1(ii), together with the result that $|m_s(Y,X,\theta)|\preceq|Y|+1$, which is established in (A.22) in the proof of Theorem 3.1.

If $\Theta_I \cap \Theta_R = \varnothing$, then $\forall\, \theta \in \Theta \cap \Theta_R$ it holds that $\sum_{s=1}^{T-R} \mathrm{MDD}\left[m_s(Y,X,\theta)\,|\underline{z}_s\right]^2 > 0$. The compactness of $\Theta \cap \Theta_R$ guarantees that

$$\min_{\theta \in \Theta \cap \Theta_R} \sum_{s=1}^{T-R} \mathrm{MDD}\left[m_s(Y,X,\theta)\,|\underline{z}_s\right]^2 > 0,$$

which completes the proof.[14] □

**Proof of Theorem 3.5.** Recall that $m_{is}^*(\theta) = m_{is}(\theta)\, v_i$ Note that $S_{Ns}^*(\theta) = S_{Ns,1}^*(\theta) + S_{Ns,2}^*(\theta)$, where $S_{Ns,1}^*(\theta) = -\frac{1}{N}\sum_{1\le i\neq j\le N} m_{is}^*(\theta)\, m_{js}^*(\theta)\,\kappa_{ij,s}$ and $S_{Ns,2}^*(\theta) = \frac{2}{N}\sum_{1\le i\neq j\le N} m_{is}^*(\theta)\,\kappa_{ij,s}\frac{1}{N}\sum_{k=1}^N m_{ks}^*(\theta)$. Note that

$$S_{Ns,2}^*(\theta) = \frac{2}{N^2}\sum_{1\le i\neq j\le N} m_{is}^*(\theta)^2\kappa_{ij,s} + \frac{2}{N^2}\sum_{1\le i\neq j\le N} m_{is}^*(\theta)\, m_{js}^*(\theta)\,\kappa_{ij,s} + \frac{(N-1)(N-2)}{N^2} N\mathbb{U}_{2Ns}^*,$$

where $\mathbb{U}_{2Ns}^* = \binom{N}{3}^{-1}\sum_{1\le i<j\le k\le N}\psi_s\left(\xi_i^*,\xi_j^*,\xi_k^*;\theta\right)$, $\xi_i^* = (\xi_i', v_i)'$ and $\psi_s^*(\xi_i^*,\xi_j^*,\xi_k^*;\theta) = \frac{1}{3}[m_{is}^*(\theta)\, m_{ks}^*(\theta)\,\kappa_{ij,s} + m_{is}^*(\theta)\, m_{js}^*(\theta)\,\kappa_{ik,s} + m_{js}^*(\theta)\, m_{ks}^*(\theta)\,\kappa_{jk,s} + m_{js}^*(\theta)\, m_{is}^*(\theta)\,\kappa_{jk,s} + m_{ks}^*(\theta)\, m_{is}^*(\theta)\,\kappa_{jk,s} + m_{ks}^*(\theta)\, m_{js}^*(\theta)\,\kappa_{ik,s}]$is a symmetrized version of $\psi_{0s}\left(\xi_i,\xi_j,\xi_k;\theta\right) \equiv 2m_{is}^*(\theta)\, m_{ks}^*(\theta)\,\kappa_{ij,s}$. Note that

$$\mathbb{E}^*\left[\psi_s\left(\xi_1^*,\xi_2^*,\xi_3^*;\theta\right)\right] = 0,\ \mathbb{E}^*\left[\psi_s\left(\xi_1^*,\xi_2^*,\xi_3^*;\theta\right)|\xi_1^*\right] = 0 \text{ and}$$
$$\mathbb{E}^*\left[\psi_s\left(\xi_1^*,\xi_2^*,\xi_3^*;\theta\right)|\xi_1^*,\xi_2^*\right] = \frac{1}{3} m_{1s}^*(\theta)\, m_{2s}^*(\theta)\,\mathbb{E}_3(\kappa_{13,s}+\kappa_{23,s}) \equiv h_s^{(2)}\left(\xi_1^*,\xi_2^*;\theta\right).$$

Let $h_s^{(3*)}\left(\xi_1^*,\xi_2^*,\xi_3^*;\theta\right) = \psi_s\left(\xi_1^*,\xi_2^*,\xi_3^*;\theta\right) - [h_s^{(2)}\left(\xi_1^*,\xi_2^*;\theta\right) + h_s^{(2)}\left(\xi_1^*,\xi_3^*;\theta\right) + h_s^{(2)}\left(\xi_2^*,\xi_3^*;\theta\right)]$. By Hoeffding's decomposition in (A.15), we have $\mathbb{U}_{2Ns}^*(\theta) = 3\mathbb{H}_{2Ns}^*(\theta) + \mathbb{H}_{3Ns}^*(\theta)$, where

$$\mathbb{H}_{2Ns}(\theta) = \binom{N}{3}^{-1}\sum_{1\le i<j\le N} h_s^{(2*)}\left(\xi_i^*,\xi_j^*;\theta\right) \text{ and } \mathbb{H}_{3Ns}(\theta) = \binom{N}{3}^{-1}\sum_{1\le i<j\le k\le N} h_s^{(3*)}\left(\xi_i^*,\xi_j^*,\xi_k^*;\theta\right),$$

where $h_s^{(2*)}\left(\xi_i^*,\xi_j^*;\theta\right) = \frac{1}{N}\sum_{k=1}^N h_s^{(2)}(\xi_i^*,\xi_j^*;\underline{z}_{k,s},\theta) = \frac{1}{3} m_{is}^*(\theta)\, m_{js}^*(\theta)\,\mathbb{E}_k(\kappa_{ik,s}+\kappa_{jk,s})$. Similarly, $S_{Ns,1}^*(\theta) = \frac{N-1}{N} N\mathbb{U}_{1Ns}^*(\theta)$, where $\mathbb{U}_{1N}^*(\theta) = -\binom{N}{2}^{-1}\sum_{1\le i\neq j\le N} m_{is}^*(\theta)\, m_{js}^*(\theta)\,\kappa_{ij,s}$. Then we have

$$S_{Ns}^*(\theta) = N\mathbb{U}_{Ns}^*(\theta) + N\mathbb{H}_{3Ns}^*(\theta) + \frac{2}{N^2}\sum_{1\le i\neq j\le N} m_{is}^*(\theta)^2\kappa_{ij,s} - \frac{3N-2}{N}\mathbb{U}_{1Ns}^*(\theta) - \frac{3N-2}{N}\mathbb{U}_{2Ns}^*,$$

where $\mathbb{U}_{Ns}^*(\theta) = \mathbb{U}_{1Ns}^*(\theta) + 3\mathbb{H}_{2Ns}^*(\theta) = \binom{N}{2}^{-1}\sum_{1\le i<j\le N} h_s^*(\xi_i^*,\xi_j^*;\theta)$ and $h_s^*(\xi_i^*,\xi_j^*;\theta) = m_{is}^*(\theta)\, m_{js}^*(\theta)\,\breve{\kappa}_{ij,s}$ with $\breve{\kappa}_{ij,s} = \mathbb{E}_k(\kappa_{ik,s}+\kappa_{jk,s}) - \kappa_{ij,s}$.

[14] Note that $\Theta_R$ is a linear subspace, so it is closed, which, in conjunction with the compactness of $\Theta$, implies the compactness of $\Theta \cap \Theta_R$.

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

Define

$$\mathcal{F}_{1s}^{*} \equiv \{m_s^{*}(\cdot;\theta_s) : (\mathbb{R}^T \times \mathcal{X}^T \times \mathbb{R}) \to \mathbb{R} : \ m_s^{*}((y,x,v);\theta_s) = \{[y_s - g(x_s)] + \sum_{r=1}^{R} \phi_{s,r}\left[y_{T-R+r} - g(x_{T-R+r})\right]\}v \text{ for some } \theta_s = (\phi_s', g)' \in \Phi_s \times \mathcal{G}\}, \quad \textbf{(A.60)}$$

$$\mathcal{F}_{2s}^{*} \equiv \{f_s^{*}(\cdot,\cdot;\theta_s) : S^{*} \times S^{*} \to \mathbb{R} : f_s^{*}(\xi_1^{*},\xi_2^{*};\theta_s) = m_s^{*}(\xi_1^{*};\theta_s)\, m_s^{*}(\xi_2^{*};\theta_s)\, \breve{\kappa}_{12,s} \text{ for some } \theta_s = (\phi_s', g)' \in \Phi_s \times \mathcal{G}\}, \quad \textbf{(A.61)}$$

and

$$\mathcal{F}_{3s}^{*} \equiv \{f_s^{*}(\cdot,\cdot,\cdot;\theta_s) : S^{*} \times S^{*} \times S^{*} \to \mathbb{R} : f_s^{*}(\xi_1^{*},\xi_2^{*},\xi_3^{*};\theta_s) = m_s^{*}(\xi_1;\theta_s)\, m_s^{*}(\xi_2;\theta_s)\, \breve{\kappa}_{12,s} + m_s^{*}(\xi_1;\theta_s)\, m_s^{*}(\xi_3;\theta_s)\, \breve{\kappa}_{13,s} + m_s^{*}(\xi_2;\theta_s)\, m_s^{*}(\xi_3;\theta_s)\, \breve{\kappa}_{23,s} \text{ for some } \theta_s = (\phi_s', g)' \in \Phi_s \times \mathcal{G}\}, \quad \textbf{(A.62)}$$

where $S^{*} = S \times \mathbb{R}$. It is easy to see the envelope functions for $\mathcal{F}_{1s}^{*}$, $\mathcal{F}_{2s}^{*}$, and $\mathcal{F}_{3s}^{*}$ are respectively given by $F_1^{*}(\xi^{*}) \equiv K(|y|+1)|v|$, $F_2^{*}(\xi_1^{*},\xi_2^{*}) = K(|y_1|+1)(|y_2|+1)\breve{\kappa}_{12,s}|v_1 v_2|$, and $F_3^{*}(\xi_1^{*},\xi_2^{*},\xi_3^{*}) = K\{(|y_1|+1)(|y_2|+1)\breve{\kappa}_{12,s}|v_1 v_2| + (|y_1|+1)(|y_3|+1)\breve{\kappa}_{13,s}|v_1 v_3| + (|y_2|+1)(|y_3|+1)\breve{\kappa}_{23,s}|v_2 v_3|\}$. Following the analysis in the proof of Theorem 3.1, we can readily show that

$$\log \mathbb{N}_{[]}\left(\epsilon \left\|F_\ell^{*}\right\|, \mathcal{F}_{\ell s}^{*}, \ \|\cdot\|_{L^2}\right) \preceq \ln\left(\frac{1}{\epsilon}\right) + \frac{1}{\epsilon} \text{ for } \ell = 1,2,3. \quad \textbf{(A.63)}$$

**Part I. Proof of part (i)**. It is easy to see that $\mathbb{E}^{*}[F_2^{*}(\xi_1^{*},\xi_2^{*})^2] < \infty$, verifying condition (a) in Theorem 5.6 of AG. As in the proof of Theorem 3.1, we can readily show that

$$\begin{aligned}
&\lim_{\delta\to 0}\limsup_{N\to\infty}\mathbb{E}^{o*}\left[\int_0^{\delta} \log \mathbb{N}_{N,2}\left(\varepsilon, \mathcal{F}_{2s}^{*}\right) d\epsilon\right] \\
&= \lim_{\delta\to 0}\limsup_{N\to\infty}\mathbb{E}^{o*}\left[\int_0^{\delta} \sum_{r=0}^{2} \log \mathbb{N}\left(\frac{\varepsilon}{2\sqrt{3}c_{2,r}}, \mathcal{F}_{2s}^{*}, \ \|\cdot\|_{L^2(U_N^r \times P^{*2-r})}\right) d\varepsilon\right] \\
&\preceq \lim_{\delta\to 0}\limsup_{N\to\infty}\left[\int_0^{\delta} \log\frac{1}{\varepsilon} d\epsilon + \mathbb{E}^{*}\left\{\left[U_N^1(P^{*1}F_2^2)\right]^{\nu/2} + \left[U_N^2(F_2^{*2})\right]^{\nu/2}\right\}\delta^{1-\nu}\right] = 0,
\end{aligned}$$

where $\mathbb{E}^{o*}$ is the outer-expectation associated with $\mathbb{E}^{*}$, the last equality follows from the fact that $\mathbb{E}^{*}\left\{\left[U_N^2(F_2^{*2})\right]^{\nu/2}\right\} \le \left\{\mathbb{E}^{*}\left[U_N^2(F_2^{*2})\right]\right\}^{\nu/2} = \{\mathbb{E}^{*}\left(F_2^{*2}\right)\}^{\nu/2} < \infty$ by Jensen inequality and similarly $\mathbb{E}^{*}\{\left[U_N^1(P^{*1}F_2^{*2})\right]^{\nu/2}\} < \infty$. This verifies condition (c) in Theorem 5.6 of AG. Note that $\mathbb{N}\left(\frac{\varepsilon}{2\sqrt{3}c_{2,r}}, \mathcal{F}_{2s}^{*}, \ \|\cdot\|_{L^2(U_N^r \times P^{*2-r})}\right) = 1$ a.s. for $r = 0,1,2$ and for sufficiently large $\varepsilon$, say, $\varepsilon \ge \varepsilon_0^{*}$. It follows that for some small $\epsilon > 0$ and by the above calculations,

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

$$\mathbb{E}^{o*}\left|\int_0^\infty \log \mathbb{N}_{N,2}\left(\varepsilon, \pi_{2,2}\mathcal{F}_{2s}^*\right) d\varepsilon\right|^{1+\epsilon}$$
$$\preceq \left|\int_0^{\varepsilon_0^*} \log \frac{1}{\varepsilon} d\epsilon\right|^{1+\epsilon} + \mathbb{E}\left\{\left[U_N^1(P^{*1}F_2^{*2})\right]^{(1+\epsilon)\nu/2} + \left[U_N^2(F_2^{*2})\right]^{(1+\epsilon)\nu/2}\right\} < \infty,$$

where the last inequality holds by choosing $\epsilon$ sufficiently small such that $(1+\epsilon)\,\nu/2 \leq 1$. This verifies the uniform integrability of the sequence $\left\{\int_0^\infty \log N_{N,2}\left(\varepsilon, \mathcal{F}_{2s}^*\right) d\varepsilon\right\}_{N=1}^\infty$ and thus condition (b) in Theorem 5.6 of AG.

Then by Theorem 5.6 of AG, we have $N\mathbb{U}_{Ns}^*(\theta) \Longrightarrow \mathbb{C}_s^*(\theta)$ in $L^\infty(\Theta)$, where $\mathbb{C}_s^*(\theta) = \mathbb{C}\left(h_s^*(\cdot,\cdot;\theta)\right)$ and $h_s^*\left(\xi_i^*, \xi_j^*; \theta_s\right) = m_s^*\left(\xi_i^*; \theta_s\right) m_s^*\left(\xi_j^*; \theta_s\right) \breve{\kappa}_{12,s} = m_{is}(\theta)\, m_{js}(\theta)\, v_i v_j \breve{\kappa}_{12,s}$. We now argue that $\left\{\mathbb{C}_s^*(\theta)\right\}$ share the same finite-sample distribution as that of $\{\mathbb{C}_s(\theta) = \mathbb{C}(h_s(\cdot,\cdot;\theta))\}$ when $\theta \in \Theta_I$. That is, $\{N\mathbb{U}_{Ns}^*(\theta_{(1)}), \ldots, N\mathbb{U}_{Ns}^*(\theta_{(L)})\}$ has the same limiting distribution as $\{N\mathbb{U}_{Ns}(\theta_{(1)}), \ldots, N\mathbb{U}_{Ns}^*(\theta_{(L)})\}$ for any finite $L$ when we restrict $\theta_{(1)}, \ldots, \theta_{(L)}$ to lie in $\Theta_I$. Without loss of generality, we can focus on the case $L = 1$.

Note that for $\theta \in \Theta_I$, $h_s(\xi_i, \xi_j; \theta) = m_{is}(\theta)\, m_{js}(\theta)\, \breve{\kappa}_{12,s}$ and $h_s^*(\xi_i^*, \xi_j^*; \theta_s) = h_s(\xi_i, \xi_j; \theta)\, v_i v_j$. Let $\{\lambda_k\}$ denote an enumeration of the positive eigenvalues of $\lambda\Phi(\cdot) = \mathbb{E}[h_s(\cdot, \xi_1; \theta)\, \Phi(\xi_1)]$ in decreasing order and according to their multiplicity. The corresponding orthonormal eigenfunctions are denoted by $\{\Phi_k(\cdot)\}_{k=1}^\infty$. It follows from a version of Mercer's theorem (e.g., Theorem 2 of Sun, 2005) that

$$h_s^{(K)}(\xi, \tilde{\xi}; \theta_s) = \sum_{k=1}^K \lambda_k \Phi_k(\xi)\, \Phi_k(\tilde{\xi}) \to \sum_{k=1}^\infty \lambda_k \Phi_k(\xi)\, \Phi_k(\tilde{\xi}) \equiv h_s(\xi, \tilde{\xi}; \theta_s)$$

for all $\xi$ and $\tilde{\xi}$ on the support of the probability law of $\xi_i$. Let

$$V_N = \frac{1}{N}\sum_{i=1}^N\sum_{j=1}^N h_s\left(\xi_i, \xi_j; \theta_s\right) \text{ and } V_N^* = \frac{1}{N}\sum_{i=1}^N\sum_{j=1}^N h_s\left(\xi_i, \xi_j; \theta_s\right) v_i v_j,$$
$$V_N^{(K)} = \frac{1}{N}\sum_{i=1}^N\sum_{j=1}^N h_s^{(K)}\left(\xi_i, \xi_j; \theta_s\right) \text{ and } V_N^{(K*)} = \frac{1}{N}\sum_{i=1}^N\sum_{j=1}^N h_s^{(K)}\left(\xi_i, \xi_j; \theta_s\right) v_i v_j.$$

Noting that $V_N - V_N^{(K)} \geq 0$, it is standard to show that

$$\mathbb{E}\left|V_N - V_N^{(K)}\right| = \sum_{k=K+1}^\infty \lambda_k \frac{1}{N}\sum_{i=1}^N \mathbb{E}\left[\Phi_k(\xi_i)^2\right] = \sum_{k=K+1}^\infty \lambda_k \to 0 \text{ as } K \to \infty \text{ and}$$
$$\mathbb{E}^*\left|V_N^* - V_N^{(K*)}\right| = \sum_{k=K+1}^\infty \lambda_k \frac{1}{N}\sum_{i=1}^N \mathbb{E}^*\left[\Phi_k(\xi_i)^2 v_i^2\right] = \sum_{k=K+1}^\infty \lambda_k \to 0 \text{ as } K \to \infty.$$

Let $\zeta_i = (\Phi_1(\xi_i), \ldots, \Phi_K(\xi_i))'$ and $\zeta_i^* = (\Phi_1(\xi_i), \ldots, \Phi_K(\xi_i))'\, v_i$. For any fixed $K$, it is trivial to show under Assumption 3.1 that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N \zeta_i \xrightarrow{\mathcal{L}} N(0, \mathbb{I}_K) \text{ and } \frac{1}{\sqrt{N}}\sum_{i=1}^N \zeta_i^* \xrightarrow{\mathcal{L}} N(0, \mathbb{I}_K).$$

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

Then by the continuous mapping theorem, we have

$$V_N^{(K)} = \sum_{k=1}^{K} \lambda_k \left( \frac{1}{N^{1/2}} \sum_{i=1}^{N} \Phi_k(\xi_i) \right)^2 \xrightarrow{\mathcal{L}} \sum_{k=1}^{K} \lambda_k W_k^2 \text{ and}$$
$$V_N^{(K*)} = \sum_{k=1}^{K} \lambda_k \left( \frac{1}{N^{1/2}} \sum_{i=1}^{N} \Phi_k(\xi_i) v_i \right)^2 \xrightarrow{\mathcal{L}} \sum_{k=1}^{K} \lambda_k W_k^2,$$

where $\{W_k\}$ is a sequence of independent $N(0,1)$ random variables. These results, in conjunction with Theorem 2 in Dehling, Durieu, and Volny (2009), imply that $V_N \xrightarrow{\mathcal{L}} \sum_{k=1}^{\infty} \lambda_k W_k^2$ and $V_N^* \xrightarrow{\mathcal{L}} \sum_{k=1}^{\infty} \lambda_k W_k^2$. Consequently, we have

$$N\mathbb{U}_{Ns}(\theta) = V_N - \frac{1}{N} \sum_{i=1}^{N} h_s(\xi_i, \xi_i; \theta) = V_N - \sum_{k=1}^{\infty} \lambda_k \frac{1}{N} \sum_{i=1}^{N} \Phi_k(\xi_i)^2 \xrightarrow{\mathcal{L}} \sum_{k=1}^{\infty} \lambda_k \left( W_k^2 - 1 \right)$$

and

$$N\mathbb{U}_{Ns}^*(\theta) = V_N^* - \frac{1}{N} \sum_{i=1}^{N} h_s^*(\xi_i^*, \xi_i^*; \theta_s) = V_N - \sum_{k=1}^{\infty} \lambda_k \frac{1}{N} \sum_{i=1}^{N} \Phi_k(\xi_i)^2 v_i^2 \xrightarrow{\mathcal{L}} \sum_{k=1}^{\infty} \lambda_k \left( W_k^2 - 1 \right).$$

That is, $N\mathbb{U}_{Ns}^*(\theta)$ share the same asymptotic distribution as $N\mathbb{U}_{Ns}(\theta)$ when $\theta \in \Theta_I$. As a result, we have $N\mathbb{U}_{Ns}^*(\theta) \Longrightarrow \mathbb{C}_s(\theta)$ in $L^\infty(\Theta_I)$.

Next, note that $\mathbb{H}_{3N}^*(\theta)$ is a third-order $P^*$-canonical $U$-process with the envelope function for its associated kernel given by $F_3^*$. Following the analysis of $\mathbb{U}_N^*(\theta)$, it is easy to show that $\mathbb{E}^*\left[F_3^*\left(\xi_1^*, \xi_2^*, \xi_3^*\right)\right] < \infty$, $\lim_{\delta\to 0} \limsup_{N\to\infty} \mathbb{E}^{o*}\left[\int_0^\delta \left[\log \mathbb{N}_{N,2}\left(\varepsilon, \mathcal{F}_{3s}^*\right)\right]^{3/2} d\epsilon\right] = 0$ and the sequence $\left\{\int_0^\infty \left[\log N_{N,2}\left(\varepsilon, \mathcal{F}_{3s}^*\right)\right]^{3/2} d\varepsilon\right\}_{N=1}^\infty$ is uniformly integrable. As a result, $N^{3/2}\mathbb{H}_{3N}^*(\theta)$ converges to a Gaussian chaos process and $\sup_{\theta\in\Theta}\left|N\mathbb{H}_{3N}^*(\theta)\right| = O_P\left(N^{-1/2}\right)$. In addition, by the uniform law of large numbers for $U$-statistics, $\frac{2}{N^2}\sum_{1\le i\neq j\le N} m_{is}^*(\theta)^2 \kappa_{ij,s} = 2\mathbb{E}^*\left[\tilde{m}_{1s}(\theta)^2 v_1^2 \kappa_{12,s}\right] + o_p(1) = 2\mathbb{E}^*\left[\tilde{m}_{1s}(\theta)^2 \kappa_{12,s}\right] = \mathbb{B}_s(\theta) + o_p(1)$uniformly in $\theta \in \Theta$. Following the analysis of $\mathbb{U}_{Ns}^*(\theta)$, we can also show that both $N\mathbb{U}_{1Ns}^*(\theta)$ and $N\mathbb{U}_{2Ns}^*(\theta)$ converge to Gaussian chaos processes. Consequently, we have

$$S_{Ns}^*(\theta) \Longrightarrow \mathbb{B}_s(\theta) + \mathbb{C}_s^*(\theta), \tag{A.64}$$

where the process $\left\{\mathbb{C}_s^*(\theta)\right\}$ coincides with $\{\mathbb{C}_s(\theta)\}$ on $\Theta_I$.

**Part II. Proof of part (ii)**. When $\theta \notin \Theta_I$, (A.64) continues to hold, which implies that

$$\mu_N^{-1} S_N^*(\theta) = \mu_N^{-1} \sum_{s=1}^{T-R} S_{Ns}^*(\theta) = O_P\left(\mu_N^{-1}\right) \text{ uniformly in } \theta \in \Theta \backslash \Theta_I.$$

https://doi.org/10.1017/S0266466623000403 Published online by Cambridge University Press

By Theorem 3.1(ii) and Assumption 3.3, we can show that

$$\min_{\theta\in\Theta_N\cap\Theta_R}\frac{1}{N}S_N(\theta)\xrightarrow{p}\min_{\theta\in\Theta\cap\Theta_R}\sum_{s=1}^{T-R}\left\{\text{MDD}\left[m_s(Y,X,\theta)\,|\underline{z}_s\right]\right\}^2.$$

Then under $\mathbb{H}_1:\ \Theta_I\cap\Theta_R=\varnothing$,

$$\mu_N^{-1}\hat{S}_N^{*}=\min_{\theta\in\Theta_N\cap\Theta_R}\left[\mu_N^{-1}S_N^{*}(\theta)+\frac{1}{N}S_N(\theta)\right]=\min_{\theta\in\Theta_N\cap\Theta_R}\left[\frac{1}{N}S_N(\theta)\right]+O_P\left(\mu_N^{-1}\right)$$
$$\xrightarrow{p}\min_{\theta\in\Theta\cap\Theta_R}\sum_{s=1}^{T-R}\left\{\text{MDD}\left[m_s(Y,X,\theta)\,|\underline{z}_s\right]\right\}^2.$$

This completes the proof of the theorem. □

## SUPPLEMENTARY MATERIAL

Shengjie Hong, Liangjun Su, and Yaqi Wang (December 2023): Supplement to "Inference in Partially Identified Panel Data Models with Interactive Fixed Effects," Econometric Theory Supplementary Material. To view, please visit https://doi.org/10.1017/S0266466623000403.

### *REFERENCES*


Adams, R. A., & Fournier, J. J. (2003). *Sobolev space*. Elsevier.

Aguiar, M., & Bils, M. (2015). Has consumption inequality mirrored income inequality? *American Economic Review*, 105(9), 2725–2756.

Ahn, S., & Horenstein, A. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, 81, 1203–1227.

Ahn, S. C., Lee, Y. H., & Schmidt, P. (2001). GMM estimation of linear panel data models with time-varying individual effects. *Journal of Econometrics*, 101, 219–255.

Ahn, S. C., Lee, Y. H., & Schmidt, P. (2013). Panel data models with multiple time-varying individual effects. *Journal of Econometrics*, 174, 1–14.

Ai, C., & Chen, X. (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica*, 71, 1795–1843.

Andrews, D. W. (1994). Empirical process methods in econometrics. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of econometrics*, vol. 4, (pp. 2248–2294), Elsevier.

Andrews, D. W., & Shi, X. (2013). Inference based on conditional moment inequalities. *Econometrica*, 81, 609–666.

Andrews, D. W., & Shi, X. (2014). Nonparametric inference based on conditional moment inequalities. *Journal of Econometrics*, 179, 31–45.

Arcones, M. A., & Giné, E. (1993). *Limit theorems for* U*-processes*. *Annals of Probability*, 21, 1494–1542.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4), 1229–1279.

Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70, 191–221.

Banks, J., Blundell, R., & Lewbel, A. (1997). Quadratic Engel curves and consumer demand. *Review of Economics and Statistics*, 79, 527–539.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57, 289–300.

Bierens, H. (1982). Consistent model specification tests. *Journal of Econometrics*, 20, 105–134.

Blundell, R. W., Browning, M., & Crawford, I. A. (2003). Nonparametric Engel curves and revealed preference. *Econometrica*, 71(1), 205–240.

Blundell, R. W., Chen, X., & Kristensen, D. (2007). Semi-nonparametric IV estimation of shape-invariant Engel curves. *Econometrica*, 75(6), 1613–1669.

Browning, M., & Crossly, T. (2009). Are two cheap, noisy measures better than one expensive, accurate one? *American Economic Review*, 99(2), 99–103.

Bücher, A., & Kojadinovic, I. (2009). A note on conditional versus joint unconditional weak convergence in bootstrap consistency results. *Journal of Theoretical Probability*, 32, 1145–1165.

Chen, X., & Pouzo, D. (2012). Estimation of nonparametric conditional moment models with possibly nonsmooth moments. *Econometrica*, 80, 277–321.

Chernozhukov, V., Newey, W., & Santos, A. (2023). Constrained conditional moment restriction models. *Econometrica*, 91(2), 709–736.

Coakley, J., Fuertes, A., & Smith, R. (2002). A principal components approach to cross-section dependence in panels. Working paper, Birkbeck College, University of London.

de la Peña, V. H., & Giné, E. (1999). *Decoupling: From dependence to independence*. Springer.

Deaton, A. S., & Muellbauer, J. (1980). An almost ideal demand system. *American Economic Review*, 70(3), 312–326.

Dehling, H., Durieu, O., & Volny, D. (2009). New techniques for empirical processes of dependent data. *Stochastic Processes and Iheir Applications*, 119, 3699–3718.

Dominguez, M. A., & Lobato, I. N. (2004). Consistent estimation of models defined by conditional moment restrictions. *Econometrica*, 72, 1601–1615.

Dong, C., Gao, J., & Peng, B. (2020). Varying–coefficient panel data models with nonstationarity and partially observed factor structure. *Journal of Business & Economic Statistics*, 39(3), 700–711.

Freyberger, J. (2018). Non-parametric panel data models with interactive fixed effects. *Review of Economic Studies*, 85(3), 1824–1851.

Freyberger, J., & Masten, M. A. (2019). A practical guide to compact infinite dimensional parameter spaces. *Econometric Reviews*, 38(9), 979–1006.

Greenaway-McGrevy, R., Han, C., & Sul, D. (2008). Estimating and testing idiosyncratic equations using cross-section dependent panel data. Working Paper, Department of Economics, University of Auckland.

Hamilton, B. W. (2001). Using Engel's law to estimate CPI bias. *American Economic Review*, 91(3), 619–630.

Holz-Eakin, D., Newey, W., & Rosen, H. (1988). Estimating vector autoregressions with panel data. *Econometrica*, 56, 1371–1395.

Hong, S. (2017). Inference in semiparametric conditional moment models with partial identification. *Journal of Econometrics*, 196, 156–179.

Hurst, E., Li, G., & Pugsley, B. (2014). Are household surveys like tax forms? Evidence from income underreporting of the self-employed. *Review of Economics and Statistics*, 96, 19–33.

Jin, S., Miao, K., & Su, L. (2021). On factor models with random missing: EM estimation, inference, and cross validation. *Journal of Econometrics*, 222, 745–777.

Kapetanios, G., & Pesaran, M. H. (2007). Alternative approaches to estimation and inference in large multifactor panels: Small sample results with an application to modelling of asset returns. In G. Phillips and E. Tzavalis (eds.), *The refinement of econometric estimation and test procedures: Finite sample and asymptotic analysis*. Cambridge University Press.

Kororok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*. Springer.

Lee, A. J. (1990). *U-statistics: Theory and practice*. CRC Press.

Leucht, A., & Neumann, M. H. (2013). Dependent wild bootstrap for degenerate U- and V-statistics. *Journal of Multivariate Analysis*, 117, 257–280.

Lu, X., & Su, L. (2016). Shrinkage estimation of dynamic panel data models with interactive fixed effects. *Journal of Econometrics*, 190, 148–175.

Manski, C. F. (2003). *Partial identification of probability distributions*. Springer.

Moon, H. R., & Weidner, M. (2015). Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica*, 83, 1543–1579.

Moon, H. R., & Weidner, M. (2017). Dynamic linear panel regression models with interactive fixed effects. *Econometric Theory*, 33, 158–195.

Nakamura, E., Steinsson, J., & Liu, M. (2016). Are Chinese growth and inflation too smooth? Evidence from Engel curves. *American Economic Journal: Macroeconomics*, 8, 113–144.

Newey, W. K., & Powell, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrica*, 71, 1565–1578.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *Review of Economics and Statistics*, 92, 1004–1016.

Pesaran, M. H. (2006). Estimation and inference in large heterogenous panels with multifactor error. *Econometrica*, 74, 967–1012.

Pesaran, M. H., & Tosetti, E. (2007). Large panels with common factors and spatial correlation. *Journal of Econometrics*, 161, 182–202.

Phillips, P. C. B., & Sul, D. (2003). Dynamic panel estimation and homogeneity testing under cross sectional dependence. *Econometrics Journal*, 6, 217–259.

Phillips, P. C. B., & Sul, D. (2007). Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence. *Journal of Econometrics*, 137, 162–188.

Pissarides, C. A., & Weber, G. (1989). An expenditure-based estimate of Britain's black economy. *Journal of Public Economics*, 39, 17–32.

Santos, A. (2012). Inference in nonparametric instrumental variables with partial identification. *Econometrica*, 80, 213–275.

Schumaker, L. (2007). *Spline functions: Basic theory*. Cambridge University Press.

Shao, X., & Zhang, J. (2014). Martingale divergence correlation and its use in high dimensional variable screening. *Journal of the American Statistical Association*, 109, 1302–1318.

Stinchcombe, M. B., & White, H. (1998). Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory*, 14, 295–325.

Su, L., & Jin, S. (2012). Sieve estimation of panel data models with cross section dependence. *Journal of Econometrics*, 169, 34–47.

Su, L., Jin, S., & Zhang, Y. (2015). Specification test for panel data models with interactive fixed effects. *Journal of Econometrics*, 186, 222–244.

Su, L., & Zhang, Y. (2018). Nonparametric dynamic panel data models with interactive fixed effects: Sieve estimation and specification testing. Working Paper, Singapore Management University.

Su, L., & Zheng, X. (2017). A Martingale-difference-divergence-based test for specification. *Economics Letters*, 156, 162–167.

Sun, H. (2005). Mercer theorem for RKHS on noncompact sets. *Journal of Complexity*, 21, 337–349.

van der Vaart, A. W., & Wellner, J. (1996). *Weak convergence and empirical processes: With applications to statistics*. Springer.