JACKKNIFE, SMALL BANDWIDTH AND HIGH-DIMENSIONAL ASYMPTOTICS

YUKITOSHI MATSUSHITA AND TAISUKE OTSU

Abstract. This paper sheds light on problems of statistical inference under alternative or nonstandard asymptotic frameworks from the perspective of jackknife empirical likelihood (JEL). Examples include small bandwidth asymptotics for semiparametric inference, many covariates asymptotics for regression models, and many-weak instruments asymptotics for instrumental variable regression. We first establish Wilks' theorem for the JEL statistic on a general semiparametric inference problem under the conventional asymptotics. We then show that the JEL statistics lose asymptotic pivotalness under the above nonstandard asymptotic frameworks, and argue that these phenomena are understood as emergence of Efron and Stein's (1981) bias of the jackknife variance estimator in the *first* order. Finally we propose a modification of JEL to recover asymptotic pivotalness under both the conventional and nonstandard asymptotics. Our modification works for all above examples and provides a unified framework to investigate nonstandard asymptotic problems.

1. Introduction

This paper sheds light on problems of statistical inference under alternative or nonstandard asymptotic frameworks from the perspective of jackknife empirical likelihood (JEL), initially proposed by Jing, Yuan and Zhou (2009) for one- and two-sample U-statistics. Examples of nonstandard asymptotics include small bandwidth asymptotics for semiparametric inference using average derivatives (e.g., Cattaneo, Crump and Jansson, 2014), many covariates asymptotics for regression models (e.g., Cattaneo, Jansson and Newey, 2018a, b), and many-weak instruments asymptotics for instrumental variable regression (e.g., Chao *et al.*, 2012). These nonstandard asymptotic frameworks are developed to provide better approximations for finite sample properties of statistics and more reliable inference methods. We investigate behaviors of the JEL statistics under such nonstandard asymptotics and develop a unified inference approach that is robust to both the conventional and nonstandard asymptotics.

In particular, we first consider a general semiparametric inference problem under the conventional asymptotics, and establish Wilks' theorem (i.e., convergence to a chi-squared distribution) for the JEL statistic. This is a natural extension of Jing, Yuan and Zhou (2009) toward semiparametric moment condition models, which are typically written by U-statistics with varying kernels. Next, we show that the JEL statistics lose asymptotic pivotalness under the above nonstandard asymptotic frameworks, and typically converge to quadratic forms of normal vectors. Our crucial finding is that the mismatch between the variance of the normal vectors and the weight matrix in these quadratic forms is understood as emergence of Efron and Stein's (1981)

Our research is supported by the JSPS KAKENHI (26780133, 18K01541) (Matsushita) and ERC Consolidator Grant (SNP 615882) (Otsu).

bias of the jackknife variance estimator in the *first* order. Under the conventional asymptotics, Efron and Stein (1981) presented a general higher-order bias formula for the jackknife variance estimator. Under the nonstandard asymptotics, however, both the linear and quadratic terms of U-statistics are of same order, and Efron and Stein's (1981) bias violates asymptotic pivotalness of the JEL statistic. Finally, based on this finding, we propose a modification of JEL to recover asymptotic pivotalness under both the conventional and nonstandard asymptotics. The basic idea is to incorporate leave-two-out adjustments as in Hinkley (1978) and Efron and Stein (1981) into the estimating equations to construct the JEL statistics. Our modification works for all above examples and provides a unified framework to investigate nonstandard asymptotic problems.

The literature on alternative or nonstandard asymptotic analysis is so broad that we limit ourselves to mention only closely related papers for the examples discussed in later sections. In a series of papers, Cattaneo, Crump and Jansson (2010, 2013, 2014) advocated the small bandwidths asymptotics to conduct robust statistical inference for semiparametric average derivative estimators. See also Cattaneo and Jansson (2018) for further developments on bootstrap inference. Kunitomo (1980) and Bekker (1994), and Chao and Swanson (2005) advocated the many instrument asymptotics and the many weak instrument asymptotics, respectively. Chao, *et al.* (2012) and Hausman, *et al.* (2012) established asymptotic normality of jackknife versions of the instrumental variable and limited information maximum likelihood estimators, respectively, under heteroskedasticity and many instruments. Since Huber (1973), there is rich literature on regression analysis with a growing number of covariates. Examples include Mammen (1993), El Karoui *et al.* (2013), Zheng *et al.* (2014), among others. The analyses in Sections 5 and 6 are closely related to Cattaneo, Jansson and Newey (2018a, b).

This paper also contributes to the literature of empirical likelihood (see Owen (2001) for a review). Since the seminal work by Jing, Yuan and Zhou (2009), JEL has been extended to various contexts, such as Wang, Peng and Qi (2013) for high dimensional means, Gong, Peng and Qi (2010) for ROC curves, Zhang and Zhao (2013) for transformation models, Peng, Qi and van Keilegom (2012) for copulas, and Zhong and Chen (2014) for regression imputation, among others. Under the conventional asymptotics, empirical likelihood inference has been studied by e.g., Bertail (2006), Zhu and Xue (2006), Hjort, McKeague and van Keilegom (2009), and Bravo, Escanciano and van Keilegom (2018).

This paper is organized as follows. Section 2 considers JEL inference for semiparametric moment condition models under the conventional asymptotics, and establishes Wilks' theorem for the JEL statistic. In Section 3, we focus on the density weighted average derivative and study properties of JEL inference under the small bandwidth asymptotics. Sections 4 and 5 study JEL inference under many-weak instruments asymptotics and many covariates asymptotics, respectively. Section 6 presents a numerical illustration.

2. Standard asymptotics

2.1. **Semiparametric model.** This section considers inference on parameters defined via semiparametric moment conditions under the conventional asymptotic framework. In particular, we are interested in a vector of parameters θ satisfying

$$E[g(Z, \theta, \mu(X))] = 0, \tag{2.1}$$

where X and Z are observables, g is a known function up to θ and μ, and μ is a vector of unknown functions. In this section, we focus on the case where $\mu(X)$ takes the form of the conditional expectation $E[Y|X]$ for some variables Y or its derivatives. Many inference problems are covered by this setup as illustrated by the following popular examples.

Example 1. [Average treatment effect] Let $Y(0)$ and $Y(1)$ be potential outcomes for a treatment $D = 0$ and 1, respectively. We observe $Z = (Y, X, D)$, where $Y = DY(1) + (1 - D)Y(0)$ and X are covariates. Under the so-called ignorability assumption (Rosenbaum and Rubin, 1983), the average treatment effect is identified as

$$\theta = E[Y(1) - Y(0)] = E[\mu_1(X) - \mu_0(X)],$$

where $\mu_d(X) = E[Y|X, D = d]$. This setup can be considered as a special case of (2.1) by setting $g(Z, \theta, \mu(X)) = \mu_1(X) - \mu_0(X) - \theta$. $\square$

Example 2. [Weighted average derivatives] Let $m(X) = E[Y|X]$ and w be a known weight function. The weighted average derivative of the regression function is defined as

$$\theta = E\left[w(X)\frac{\partial m(X)}{\partial X}\right].$$

This object is often used for estimation of single index models (e.g., Powell, Stock and Stoker, 1989) and some nonseparable models. This setup can be considered as a special case of (2.1) by setting $g(Z, \theta, \mu(X)) = w(X)\mu(X) - \theta$ with $\mu(x) = \partial m(x)/\partial x$. $\square$

Other examples include estimating equations for various semiparametric models, such as partially linear and varying coefficient models.

Suppose a preliminary estimator $\hat{\mu}$ for μ is available. Then the parameters θ can be estimated by solving the estimating equations

$$\frac{1}{n}\sum_{j=1}^{n} g(Z_j, \hat{\theta}, \hat{\mu}(X_j)) = 0.$$

Under certain regularity conditions, the influence function of $\hat{\theta}$ is given by (Newey, 1994a)

$$\psi(Z, X) = -E\left[\frac{\partial g(Z, \theta, \mu(X))}{\partial \theta'}\right]\left\{g(Z, \theta, \mu(X)) + E\left[\frac{\partial g(Z, \theta, \mu(X))}{\partial \mu'}\middle| X\right]\{Y - \mu(X)\}\right\}, \tag{2.2}$$

and the asymptotic variance of $\hat{\theta}$ is obtained by $Var(\psi(Z, X))$. To obtain the Wald-type confidence set for θ, we need to estimate the asymptotic variance $Var(\psi(Z, X))$ that involves analytical or often numerical derivatives of g and estimation of the conditional mean $E\left[\frac{\partial g(Z,\theta,\mu(X))}{\partial \mu'}\middle| X\right]$ and average derivatives $E\left[\frac{\partial g(Z,\theta,\mu(X))}{\partial \theta'}\right]$. We provide an alternative inference approach based on the JEL, which does not require estimation of nonparametric components in $Var(\psi(Z, X))$ nor even computation of the derivatives of g.

2.2. **Jackknife empirical likelihood.** We now introduce the JEL approach for the setup in (2.1). Here we focus on the case where $\mu(X)$ is estimated by the kernel estimator

$$\hat{\mu}(X_j) = \frac{1}{\hat{f}(X_j)} \frac{1}{n-1} \sum_{k \neq j} \frac{1}{h^d} K\left(\frac{X_j - X_k}{h}\right) Y_k,$$

where K is a kernel function, h is the bandwidth, and $\hat{f}(X_j) = \frac{1}{n-1} \sum_{k \neq j} \frac{1}{h^d} K\left(\frac{X_j - X_k}{h}\right)$ is an estimator for the density f of X.[1] For given θ, we construct the jackknife pseudo-values as

$$V_i(\theta) = nS(\theta) - (n-1)S^{(i)}(\theta), \tag{2.3}$$

where

$$S(\theta) = \frac{1}{n} \sum_{j=1}^{n} g(Z_j, \theta, \hat{\mu}(X_j)), \qquad S^{(i)}(\theta) = \frac{1}{n-1} \sum_{j \neq i} g(Z_j, \theta, \hat{\mu}^{(i)}(X_j)),$$

and $\hat{\mu}^{(i)}(X_j) = \frac{1}{\hat{f}(X_j)} \frac{1}{n-2} \sum_{k \neq i,j} K\left(\frac{X_j - X_k}{h}\right) Y_k$ is a leave-i-out counterpart of $\hat{\mu}(X_j)$. We treat the jackknife pseudo-values as if they are estimating equations for θ, and construct JEL as

$$\ell(\theta) = -2 \sup_{\{p_i\}_{i=1}^n} \sum_{i=1}^{n} \log(np_i), \quad \text{s.t. } p_i \geq 0, \ \sum_{i=1}^{n} p_i = 1, \ \sum_{i=1}^{n} p_i V_i(\theta) = 0.$$

By applying the Lagrange multiplier method, the dual form of $\ell(\theta)$ is written as

$$\ell(\theta) = 2 \sup_{\lambda} \sum_{i=1}^{n} \log(1 + \lambda' V_i(\theta)). \tag{2.4}$$

In practice we employ this dual formula to compute $\ell(\theta)$. The asymptotic property of the JEL statistic $\ell(\theta)$ is obtained as follows.

Theorem SP. *Under Assumption SP in Appendix, it holds $\ell(\theta) \xrightarrow{d} \chi^2_p$, where p is the dimension of θ.*

This theorem says that the JEL statistic $\ell(\theta)$ is asymptotically pivotal and converges to the χ^2_p distribution. Thus, the JEL confidence set of θ can be constructed by $\{c : \ell(c) \leq \chi^2_{p,\alpha}\}$, where $\chi^2_{p,\alpha}$ is the $(1-\alpha)$-th quantile of the χ^2_p distribution. In contrast to the Wald-type confidence set based on the influence function in (2.2), the JEL-based inference does not require estimation of nonparametric components nor evaluations of the derivatives of g. Also we do not have to derive the influence function for each application. The above construction of JEL is particularly attractive when computation of the estimator $\hat{\theta}$ is expensive. Indeed the JEL statistic $\ell(\theta)$ does not involve any point estimator of θ because we conduct jackknifing on the estimating equations rather than the estimator.

3. Small bandwidth asymptotics

3.1. **Density weighted average derivative.** In this section we focus on the density weighted average derivative

$$\theta = E\left[f(X) \frac{\partial \mu(X)}{\partial X}\right],$$

[1]Similar results can be established for local polynomial estimators.

where f is the density of X and $\mu(X) = E[Y|X]$. Using integration by parts, this parameter is alternatively written as $\theta = -2E\left[Y \frac{\partial f(X)}{\partial X}\right]$, and thus can be estimated by

$$\hat{\theta} = -\frac{2}{n} \sum_{j=1}^{n} Y_j \frac{\partial \hat{f}(X_j)}{\partial X}, \tag{3.1}$$

where $\hat{f}(X_j) = \frac{1}{n-1} \sum_{k \neq j} \frac{1}{h^d} K\left(\frac{X_j - X_k}{h}\right)$ is the (leave-one-out) kernel density estimator. Note that this estimator takes the form of the second-order U-statistic and admits the Hoeffding decomposition:

$$\hat{\theta} = \binom{n}{2}^{-1} \sum_{j=1}^{n} \sum_{k=j+1}^{n} U_{jk} = E[\hat{\theta}] + \frac{1}{n} \sum_{j=1}^{n} L_j + \binom{n}{2}^{-1} \sum_{j=1}^{n} \sum_{k=j+1}^{n} W_{jk}, \tag{3.2}$$

where $U_{jk} = -\frac{1}{h^{d+1}} \dot{K}\left(\frac{X_j - X_k}{h}\right)(Y_j - Y_k)$ with the derivative $\dot{K}$ of K, $L_j = 2\{E[U_{jk}|Z_j] - E[U_{jk}]\}$, and $W_{jk} = U_{jk} - (L_j + L_k)/2 - E[U_{jk}]$. Under standard conditions (listed in Assumption SB in Appendix), the bias term $E[\hat{\theta}] - \theta$ is of order $O(h^s)$, where s is smoothness of f as well as the order of the kernel, and the quadratic term $\binom{n}{2}^{-1} \sum_{j=1}^{n} \sum_{k=j+1}^{n} W_{jk}$ is of order $O_p(n^{-1} h^{-\frac{d}{2}-1})$. Thus, by imposing both $\sqrt{n} h^s \to 0$ and $nh^{d+2} \to \infty$, the limiting distribution of $\hat{\theta}$ is determined by the linear term in (3.2) (Powell, Stock and Stoker, 1989), that is

$$\sqrt{n}(\hat{\theta} - \theta) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} L_j + o_p(1) \xrightarrow{d} N(0, \Sigma),$$

where $\Sigma = E[L_j L_j']$. In order to robustify inference on θ against the choice of bandwidths, Cattaneo, Crump and Jansson (2014) relaxed the requirement $nh^{d+2} \to \infty$ (called the small bandwidth asymptotics) so that both the linear and quadratic terms in (3.2) play the dominant roles. In particular, they established

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \Sigma + 2\kappa^{-1}\Delta) \quad \text{under} \quad nh^{d+2} \to \kappa \in (0, \infty),$$
$$\sqrt{\binom{n}{2} h^{d+2}}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \Delta) \quad \text{under} \quad nh^{d+2} \to 0,$$

where $\Delta = \lim_{n\to\infty} h^{d+2} E[W_{jk} W_{jk}'] = 2E[Var(Y|X) f(X)] \int \dot{K}(u) \dot{K}(u)' du$ is the variance of the quadratic term in the Hoeffding decomposition (3.2). Cattaneo, Crump and Jansson (2014) advocated inference based on the case of $nh^{d+2} \to \kappa$ by estimating the asymptotic variance $\Sigma + 2\kappa^{-1}\Delta$.

3.2. **Jackknife empirical likelihood.** We apply the JEL method to the density weighted average derivative estimator $\hat{\theta}$ in (3.1). Based on the estimator, we construct the jackknife pseudo-values as in (2.3) with

$$S(\theta) = \hat{\theta} - \theta, \qquad S^{(i)}(\theta) = \hat{\theta}^{(i)} - \theta,$$

where $\hat{\theta}^{(i)}$ is the leave-i-out version of $\hat{\theta}$ in (3.1). The asymptotic property of the JEL statistic in (2.4) is obtained as follows.

Theorem SB1. *Consider the setup of this section and suppose Assumption SB in Appendix holds true. Then*

$$\ell(\theta) \xrightarrow{d} \begin{cases} \chi_d^2 & \text{under } nh^{d+2} \to \infty, \\ \xi'(\Sigma + 4\kappa^{-1}\Delta)\xi & \text{under } nh^{d+2} \to \kappa \in (0,\infty), \\ \frac{1}{2}\chi_d^2 & \text{under } nh^{d+2} \to 0, \end{cases}$$

where $\xi \sim N(0, \Sigma + 2\kappa^{-1}\Delta)$.

Similar to the estimator $\hat{\theta}$, the limiting distribution of the JEL statistic $\ell(\theta)$ depends on the condition on nh^{d+2}. If $nh^{d+2} \to 0$ or ∞, then the JEL statistic is asymptotically pivotal but obeys different limiting distributions. In particular, if we use the conventional χ_d^2 critical values for very small values of h, such inference tends to be conservative. For the knife edge case of $nh^{d+2} \to \kappa \in (0,\infty)$, the JEL statistic is no longer asymptotically pivotal and its limiting distribution depends on κ. It is interesting to note that discrepancy of the constants multiplied to $\kappa^{-1}\Delta$ in the variance of ξ and the term $\Sigma + 4\kappa^{-1}\Delta$ is analogous to the (second-order) bias in the conventional jackknife variance estimator (Efron and Stein, 1981). This Efron-Stein bias of the jackknife variance estimator is exactly due to mismatch of characterizing the quadratic term in the Hoeffding decomposition. Under the small bandwidth asymptotics, the Efron-Stein bias emerges in the *first* order.

It is desirable to modify JEL to have the same limiting distribution for all cases. To this end, we employ the bias correction method suggested by Efron and Stein (1981) and modify the JEL statistic as follows. Let $\hat{\theta}^{(i,j)}$ be the leave-(i,j)-out version of $\hat{\theta}$, and define

$$Q_{ij} = n\hat{\theta} - (n-1)(\hat{\theta}^{(i)} + \hat{\theta}^{(j)}) + (n-2)\hat{\theta}^{(i,j)}.$$

This term is used in Efron and Stein (1981) to correct the higher-order bias of the jackknife variance estimator.[2] Since Q_{ij} is asymptotically expressed as a function of W_{ij}'s but not L_i's (see, eq. (A.13)), it can be used to estimate the variance component Δ. We utilize this term to modify the JEL statistic as follows

$$\ell^m(\theta) = 2\sup_{\lambda} \sum_{i=1}^{n} \log(1 + \lambda' V_i^m(\theta)), \tag{3.3}$$

where $V_i^m(\theta) = V_i(\hat{\theta}) - \hat{\Gamma}\tilde{\Gamma}^{-1}\{V_i(\hat{\theta}) - V_i(\theta)\}$, and $\hat{\Gamma}$ and $\tilde{\Gamma}$ are given by

$$\hat{\Gamma}\hat{\Gamma}' = \frac{1}{n}\sum_{i=1}^{n} V_i(\hat{\theta})V_i(\hat{\theta})', \qquad \tilde{\Gamma}\tilde{\Gamma}' = \frac{1}{n}\sum_{i=1}^{n} V_i(\hat{\theta})V_i(\hat{\theta})' - \frac{1}{n}\sum_{i=1}^{n}\sum_{j=i+1}^{n} Q_{ij}Q_{ij}'.$$

Theorem SB2. *Consider the setup of this section. Under Assumption SB,* $\ell^m(\theta) \xrightarrow{d} \chi_d^2$ *(regardless of the condition on* nh^{d+2}*).*

---

[2] If θ is scalar, the bias corrected variance estimator is given by

$$\frac{n-1}{n}\sum_{i=1}^{n}(\hat{\theta}^{(i)} - \hat{\theta})^2 - \frac{1}{n(n+1)}\sum_{i=1}^{n}\sum_{j=i+1}^{n}(Q_{ij} - \bar{Q})^2,$$

where $\bar{Q} = \frac{2}{n(n-1)}\sum_{j=i+1}^{n} Q_{ij}$.

Therefore, the modified JEL $\ell^m(\theta)$ is asymptotically pivotal and follows the χ^2_d limiting distribution for all cases of nh^{d+2}. Note that the modified JEL inference only requires the estimators, $\hat{\theta}$, $\hat{\theta}^{(i)}$, and $\hat{\theta}^{(i,j)}$, and circumvents estimation of Σ and Δ, which contains nonparametric components and requires additional smoothing.

4. Many weak instruments asymptotics

4.1. **Instrumental variable regression.** In this section, we consider the instrumental variable regression model

$$\begin{aligned} Y &= X\theta + U, \\ X &= Z'\gamma_n + \epsilon, \end{aligned} \tag{4.1}$$

where Y and X are scalar observables, U and ϵ are scalar error terms, and Z is a K-dimensional vector of instrumental variables.[3] We assume Z is nonrandom (otherwise conditional on Z). For the coefficient vector γ_n, we assume

$$\gamma_n = n^{-1/2}\mu_n \pi,$$

where μ_n is a scalar sequence and π is a K-dimensional vector of constants. We are interested in the three cases: (i) K is fixed and $\mu_n = O(n^{1/2})$, (ii) $K \to \infty$ and $K/\mu_n^2 \to \alpha \in (0, \infty)$ as $n \to \infty$, and (iii) $K \to \infty$ and $K/\mu_n^2 \to \infty$ as $n \to \infty$. Case (i) is the conventional asymptotic framework. Cases (ii) and (iii) are designed to the situations where the researcher has access to many but possibly weak instrumental variables.

As an estimator of θ, we focus on the jackknife instrumental variables (JIV) estimator (Angrist, Imbens and Krueger, 1999):

$$\hat{\theta} = \left(\sum_{k=1}^{n} \sum_{l \neq k} X_k P_{kl} X_l \right)^{-1} \sum_{k=1}^{n} \sum_{l \neq k} X_k P_{kl} Y_l,$$

where $P_{kl} = Z_k' \left(\sum_{h=1}^{n} Z_h Z_h' \right)^{-1} Z_l$. It is known that the JIV estimator is robust to heteroskedasticity and many instruments in contrast to the LIML and 2SLS estimators. Let $\sigma_k^2 = E[U_k^2]$. Under Assumption MW in the appendix, the limiting distribution of the JIV estimator is derived as follows (Chao *et al.,* 2012):

$$\begin{aligned} \text{Case (i)} &: \quad \mu_n(\hat{\theta} - \theta) \xrightarrow{d} N(0, H^{-1}\Sigma H^{-1}), \\ \text{Case (ii)} &: \quad \mu_n(\hat{\theta} - \theta) \xrightarrow{d} N(0, H^{-1}\Sigma H^{-1} + \alpha H^{-1}\Psi H^{-1}), \\ \text{Case (iii)} &: \quad \frac{\mu_n^2}{\sqrt{K}}(\hat{\theta} - \theta) \xrightarrow{d} N(0, H^{-1}\Psi H^{-1}), \end{aligned}$$

[3]To simplify the presentation, we consider the case where X is scalar, but an extension to the vector case is relatively straightforward.

where

$$H = \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}(1-P_{kk})\pi' Z_k Z_k'\pi, \qquad \Sigma = \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\sigma_k^2(1-P_{kk})^2\pi' Z_k Z_k'\pi,$$

$$\Psi = \lim_{n\to\infty}\frac{1}{K}\sum_{k=1}^{n}\sum_{l\neq k}P_{kl}^2\{\sigma_k^2 E[\epsilon_l^2]+E[\epsilon_k U_k]E[\epsilon_l U_l]\}.$$

Based on this result, Chao *et al.* (2012) suggested a robust inference method by estimating the unknown components H, Σ, and Ψ.

4.2. **Jackknife empirical likelihood.** In this case, based on the first-order condition of the JIV estimator, we construct the jackknife pseudo-values as in (2.3) with

$$S(\theta)=\frac{1}{n(n-1)}\sum_{k=1}^{n}\sum_{l\neq k}X_k P_{kl}(Y_l-X_l\theta), \qquad S^{(i)}(\theta)=\frac{1}{(n-1)(n-2)}\sum_{k\neq i}\sum_{l\neq i,k}X_k P_{kl}(Y_l-X_l\theta).$$

The asymptotic property of the jackknife empirical likelihood statistic in (2.4) is obtained as follows. Let $\Xi=\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\sum_{l\neq k}\sigma_l^2 P_{lk}^2\pi' Z_k Z_k'\pi$.

Theorem MW1. *Consider the setup of this section. Under Assumption MW,*

$$\ell(\theta)\xrightarrow{d}\begin{cases}\chi_1^2 & \textit{under Case (i)},\\ \xi^2/(\Sigma+\Xi+2\alpha\Psi) & \textit{under Case (ii)},\\ \frac{1}{2}\chi_1^2 & \textit{under Case (iii)},\end{cases}$$

where $\xi\sim N(0,\Sigma+\alpha\Psi)$.

Similar to Theorem SB1, the JEL statistic is not asymptotically pivotal under the many weak instruments asymptotics of Case (ii). On the other hand, for Case (iii), where the instruments are even weaker than Case (ii), the JEL statistic recovers asymptotic pivotalness. The term $\frac{1}{2}$ appears by setting $\Sigma=\Xi=0$ for Case (ii). The additional term Ξ emerges due to the fact that the matrix $[P_{kl}]$ is not exactly the projection matrix for the leave-i-out counterpart $S^{(i)}(\theta)$.

It is desirable to modify JEL to have same χ_1^2 limiting distribution for all cases. Let

$$Q_{ij}=nS(\theta)-(n-1)\{S^{(i)}(\theta)+S^{(j)}(\theta)\}+(n-2)S^{(i,j)}(\theta),$$

where

$$S^{(i,j)}(\theta)=\frac{1}{(n-2)(n-3)}\sum_{k\neq i,j}\sum_{l\neq i,j,k}X_k P_{kl}(Y_l-X_l\theta),$$

is the leave-(i,j)-out version of $S(\theta)$. Then define the modified JEL statistic $\ell^m(\theta)$ as in (3.3).

Theorem MW2. *Consider the setup of this section. Under Assumption MW,* $\ell^m(\theta)\xrightarrow{d}\chi_1^2$ *(for all cases).*

Similar comments to Theorem SB2 apply. The modified JEL $\ell^m(\theta)$ follows the χ_1^2 limiting distribution for all cases without estimating the variance components Σ, Ξ, and Ψ.

5. Many regressors asymptotics

We consider the regression model

$$Y = X\theta + Z'\gamma_n + U, \tag{5.1}$$

where Y and X are scalar observables, Z is a K-dimensional vector of covariates, and U is an error term. We are concerned with inference on the scalar parameter θ under two scenarios, $\frac{K}{n} \to 0$ and $\frac{K}{n} \to \tau \in (0,1)$ as $n \to \infty$.

Let $P_{kl} = Z_k'(\sum_{h=1}^n Z_h Z_h')^{-1} Z_l$ and $M_{kl} = \mathbb{I}\{k=l\} - P_{kl}$. Also define $\tilde{X}_k = \sum_{l=1}^n M_{kl} X_l$. We construct the jackknife pseudo-values as in (2.3) with

$$\begin{aligned} S(\theta) &= \frac{1}{n}\sum_{k=1}^n \sum_{l=1}^n \tilde{X}_k M_{kl}(Y_l - X_l\theta), \\ S^{(i)}(\theta) &= \frac{1}{n-1}\sum_{k\neq i} \tilde{X}_k M_{kk}(Y_k - X_k\theta) + \frac{1}{n-2}\sum_{k\neq i}\sum_{l\neq i,k} \tilde{X}_k M_{kl}(Y_l - X_l\theta). \end{aligned}$$

Let

$$\begin{aligned} \Sigma &= p\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \tilde{X}_i^2 U_i^2, \qquad \Psi = p\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \left(\sum_{l\neq i} \tilde{X}_i M_{il} U_l\right)^2, \\ \Xi_1 &= p\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n (\tilde{X}_i P_{ii} Z_i'\gamma)^2, \qquad \Xi_2 = p\lim_{n\to\infty}\frac{1}{n^3}\sum_{i=1}^n \left(\sum_{k\neq i} \tilde{X}_k M_{kk} Z_k'\gamma\right)^2. \end{aligned}$$

The limiting distribution of the JEL statistic is obtained as follows.

Theorem MR1. *Consider the setup of this section. Under Assumption MR,*

$$\ell(\theta) \xrightarrow{d} \begin{cases} \chi_1^2 & \text{under } \frac{K}{n} \to 0, \\ \xi^2/(\Sigma + \Psi + \Xi_1 + \Xi_2) & \text{under } \frac{K}{n} \to \tau \in (0,1), \end{cases}$$

where $\xi \sim N(0, \Sigma)$.

Similar to Theorems SB1 and MW1, the JEL statistic is not asymptotically pivotal under the many regressors asymptotics with $\frac{K}{n} \to \tau \in (0,1)$. The term Ψ emerges due to mismatch of characterizing the quadratic term in the Hoeffding decomposition of $\frac{1}{n}\sum_{i=1}^n \tilde{X}_i U_i = \frac{1}{n}\sum_{i=1}^n \sum_{l=1}^n X_l M_{il} U_i$, which is analogous to the Efron-Stein bias. Again, the additional terms Ξ_1 and Ξ_2 emerge due to the fact that the matrix $[P_{kl}]$ is not exactly the projection matrix for the leave-i-out counterpart $S^{(i)}(\theta)$.

It is desirable to modify JEL to have same χ_1^2 limiting distribution for all cases. Let $\hat{\gamma}^{(i)}$ be the leave-i-out OLS estimator for γ from the regression of $Y_i - X_i\theta$ on Z_i, and[4]

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^n \tilde{X}_i^2 \{(Y_i - X_i\theta)(Y_i - X_i\theta - Z_i'\hat{\gamma}^{(i)})\}.$$

We define the modified JEL statistic as in (3.3) with the OLS estimator $\hat{\theta}$ and $\tilde{\Gamma}\tilde{\Gamma}' = \hat{\Sigma}$.

[4]Kline, Saggio and Sølvsten (2018) proposed similar estimators for quadratic forms in the parameters of linear models with many regressors and heteroskedasticity.

Theorem MR2. *Consider the setup of this section. Under Assumption MR,* $\ell^m(\theta) \xrightarrow{d} \chi_1^2$ *(for both cases).*

Similar comments to Theorems SB2 and MW2 apply. The modified JEL $\ell^m(\theta)$ follows the χ_1^2 limiting distribution for all cases without estimating the variance components Σ, Ξ, and Ψ. Under the asymptotics $\frac{K}{n} \to \tau \in (0, 1/2)$, Cattaneo, Jansson and Newey (2018a) developed a robust Wald inference method for θ. It is interesting to note that the above theorems on JEL allow $\tau \in (0, 1)$. Therefore, we expect that our JEL inference works better when $\frac{K}{n} \geq \frac{1}{2}$. In the next section, we examine this point by a simulation study.

6. Numerical illustration

This section conducts a simulation study to evaluate the finite sample properties of the JEL inference methods. In particular, we adopt the simulation designs in Cattaneo, Jansson and Newey (2018a).

First, we consider a semiparametric partially linear model (Model 1):

$$\begin{aligned} Y &= \beta X + g(W) + U, \qquad U|X,W \sim N(0, \sigma_U^2), \qquad \sigma_U^2 = c_U\{1 + (t(X) + \iota' W)^2\}^{\vartheta}, \\ X &= h(W) + V, \qquad V|W \sim N(0, \sigma_V^2), \qquad \sigma_V^2 = c_V\{1 + (\iota' W)^2\}^{\vartheta}, \end{aligned}$$

where $\beta = 1$, W is a six-dimensional mutually independent $U[-1, 1]$ random variables, the unknown regression functions are set to $g(w) = \exp(-|w|^{1/2})$ and $h(w) = \exp(|w|^{1/2})$, $\iota = (1, 1, \ldots, 1)'$, and $t(a) = a\mathbb{I}\{-2 \leq a \leq 2\} + 2\text{sgn}(a)(1 - \mathbb{I}\{2 \leq a \leq 2\})$. The constants c_U and c_V are chosen so that $Var(U) = Var(V) = 1$, and we consider two cases: Homoskedastic ($\vartheta = 0$) and Heteroskedastic ($\vartheta = 1$). We observe a random sample $\{Y_i, X_i, W_i\}_{i=1}^n$ form (Y, X, W) of size $n = 250$ for each Monte Carlo replication. To approximate the unknown function g, we employ power series expansions. To be specific, we consider the polynomial basis expansion in Table 1.

K	$p_K(w)$
7	$(1, w_1, w_2, w_3, w_4, w_5, w_6)$
13	$(p_7(w), w_1^2, w_2^2, w_3^2, w_4^2, w_5^2, w_6^2)$
28	$p_{13}(w)$ + first-order interactions
34	$(p_{28}(w), w_1^3, w_2^3, w_3^3, w_4^3, w_5^3, w_6^3)$
84	$p_{34}(w)$ + second-order interactions
90	$(p_{84}(w), w_1^4, w_2^4, w_3^4, w_4^4, w_5^4, w_6^4)$
210	$p_{90}(w)$ + third-order interactions
216	$(p_{210}(w), w_1^4, w_2^4, w_3^4, w_4^4, w_5^4, w_6^4, w_6^5)$

Table 1. Basis functions

Second, we consider a linear model (Model 2):

$$\begin{aligned} Y &= \beta X + \gamma' W + U, \qquad U|X,W \sim N(0, \sigma_U^2), \qquad \sigma_U^2 = c_U\{1 + (t(X) + \iota' W)^2\}^{\vartheta}, \\ X &= V, \qquad V|W \sim N(0, \sigma_V^2), \qquad \sigma_V^2 = c_V\{1 + (\iota' W)^2\}^{\vartheta}, \end{aligned}$$

where $W = (1, W_2, \ldots, W_K)$ with $W_j = \mathbb{I}\{N(0, 1) \geq 1.5\}$ for $j = 2, \ldots, K$, $\beta = 1$, and $\gamma = 0$.

We compare four methods to construct confidence intervals for β: (i) Wald-type confidence interval (Wald-HC0) with the usual version of Eicker-White heteroskedasticity-robust standard error, (ii) Wald-type confidence interval (Wald-CJN) with the heteroskedasticity-robust standard error proposed by Cattaneo, Jansson and Newey (2018a), (iii) jackknife empirical likelihood confidence interval (JEL) in Section 5, and (iv) modified jackknife empirical likelihood confidence interval (mJEL) in Section 5.

Tables 2 and 3 give the empirical coverage rates of all the intervals across 1,000 replications for Model 1 and Model 2, respectively. The nominal rate is 0.95. The main findings from the simulation study are in line with our theoretical results.[5] Wald-HC0 intervals tend to under-cover especially when the dimension K is large. Wald-CJN intervals offer close to correct empirical empirical coverage when $K/n < 1/2$, but tend to under-cover when $K/n \geq 1/2$. JEL intervals are conservative, which verifies our theoretical results. mJEL are most robust to the dimension compared to the other intervals and they offer close-to-correct empirical coverages in all cases.

We also analyze the power properties of the tests for $H_0 : \beta = 1$ under the alternative hypotheses $H_1 : \beta = 1 + \Delta$ for $\Delta = -0.2, -0.1, 0.1, 0.2$. Tables 4-5 (Model 1) and 6-7 (Model 2) give the calibrated powers of all the tests across 1,000 replications (i.e., the rejection frequencies of these tests where the critical values are given by the Monte Carlo 95% percentiles of these test statistics under H_0). The results suggest that the proposed mJEL tests exhibit good calibrated power.

K	Wald-HC0	Wald-CJN	JEL	mJEL	Wald-HC0	Wald-CJN	JEL	mJEL
	Homoskedastic				Heteroskedastic			
7	0.871	0.875	0.883	0.875	0.866	0.875	0.879	0.874
13	0.937	0.941	0.951	0.943	0.925	0.937	0.940	0.936
28	0.931	0.944	0.960	0.946	0.921	0.941	0.953	0.939
34	0.922	0.948	0.964	0.948	0.918	0.945	0.957	0.944
84	0.908	0.949	0.970	0.956	0.878	0.939	0.962	0.940
90	0.894	0.937	0.972	0.951	0.851	0.930	0.967	0.944
210	0.660	0.847	0.965	0.955	0.646	0.842	0.965	0.946
216	0.656	0.858	0.974	0.957	0.636	0.850	0.959	0.951

TABLE 2. Coverage probabilities of 95% confidence intervals (Model 1)

K	Wald-HC0	Wald-CJN	JEL	mJEL	Wald-HC0	Wald-CJN	JEL	mJEL
	Homoskedastic				Heteroskedastic			
5	0.943	0.945	0.948	0.945	0.931	0.935	0.940	0.936
25	0.932	0.944	0.960	0.946	0.900	0.940	0.953	0.945
50	0.905	0.941	0.957	0.944	0.896	0.950	0.970	0.953
100	0.848	0.939	0.971	0.946	0.831	0.927	0.968	0.943
200	0.595	0.868	0.959	0.948	0.585	0.853	0.966	0.951

TABLE 3. Coverage probabilities of 95% confidence intervals (Model 2)

[5]In Table 2, all the intervals do not provide close-to-correct empirical coverage when $K = 7$, because the semi-parametric model clearly exhibits misspecification error when K is small.

Δ	Wald-HC0	Wald-CJN	JEL	mJEL	Wald-HC0	Wald-CJN	JEL	mJEL
	Homoskedastic				Heteroskedastic			
-0.2	0.687	0.673	0.655	0.667	0.616	0.611	0.578	0.596
-0.1	0.252	0.227	0.240	0.232	0.200	0.211	0.203	0.192
0.1	0.235	0.229	0.221	0.223	0217	0.231	0.207	0.212
0.2	0.709	0.687	0.651	0.672	0.633	0.643	0.571	0.571

Table 4. Calibrated power for Model 1 ($n = 250$, $K = 90$)

Δ	Wald-HC0	Wald-CJN	JEL	mJEL	Wald-HC0	Wald-CJN	JEL	mJEL
	Homoskedastic				Heteroskedastic			
-0.2	0.259	0.162	0.212	0.202	0.239	0.121	0.194	0.191
-0.1	0.094	0.075	0.094	0.094	0.086	0.075	0.093	0.082
0.1	0.100	0.061	0.073	0.088	0.095	0.052	0.061	0.069
0.2	0.252	0.159	0.125	0.135	0.236	0.147	0.108	0.114

Table 5. Calibrated power for Model 1 ($n = 250$, $K = 210$)

Δ	Wald-HC0	Wald-CJN	JEL	mJEL	Wald-HC0	Wald-CJN	JEL	mJEL
	Homoskedastic				Heteroskedastic			
-0.2	0.692	0.686	0.709	0.717	0.564	0.538	0.561	0.556
-0.1	0.231	0.225	0.247	0.248	0.211	0.202	0.216	0.209
0.1	0.238	0.236	0.254	0.257	0.181	0.161	0.177	0.169
0.2	0.665	0.653	0.687	0.688	0.574	0.533	0.564	0.544

Table 6. Calibrated power for Model 2 ($n = 250$, $K = 100$)

Δ	Wald-HC0	Wald-CJN	JEL	mJEL	Wald-HC0	Wald-CJN	JEL	mJEL
	Homoskedastic				Heteroskedastic			
-0.2	0.335	0.230	0.321	0.323	0.280	0.154	0. 240	0.249
-0.1	0.123	0.100	0.121	0.122	0.099	0.070	0.087	0.095
0.1	0.130	0.120	0.122	0.118	0.112	0.070	0.104	0.104
0.2	0.295	0.210	0.280	0.283	0.291	0.184	0.255	0.266

Table 7. Calibrated power for Model 2 ($n = 250$, $K = 200$)

Appendix A. Mathematical Appendix

A.1. Assumptions.

Assumption SP.

- **(i):** $\{Y_i, X_i, Z_i\}_{i=1}^n$ *is independent and identically distributed.* X *is compactly supported in* $\mathbb{R}^d$ *and its density* f *is uniformly bounded from above and away from zero.* μ *and* f *are continuously differentiable to order* s. $E[|Y - \mu(X)|^{2+\delta}] < \infty$ *for some* $\delta > 0$, $E[Y^p] < \infty$ *for some* $p \geq 4$, *and* $E[Y^p|X = x]f(x)$ *is bounded.* g *has bounded second derivative in* μ.
- **(ii):** K *is an* s*-th order kernel function that integrates to 1 in its compact support. Also,* $nh^{2d}/(\log n)^2 \to \infty$ *and* $nh^{2s} \to 0$ *as* $n \to \infty$.

Assumption SB.

- **(i):** f *is* $(s+1)$ *times differentiable, and* f *and its first* $(s+1)$ *derivatives are bounded for some* $s \geq 2$. m *is twice differentiable,* $e = mf$ *has the bounded second derivative,* $v(x) =$

$E[Y^2|X=x]$ *is differentiable,* vf *has the bounded first derivative, and* $\lim_{|x|\to\infty}\{m(x)+|e(x)|\}=0$. $E[Y^4]<\infty$, $E[Var(Y|X)f(X)]>0$, *and* $Var\left(\frac{\partial e(X)}{\partial X}-Y\frac{\partial f(X)}{\partial X}\right)$ *is positive definite.*

(ii): K *is even, differentiable with the bounded first derivative* $\dot{K}$, *and* s*-th order kernel. Also,* $\int \dot{K}(u)\dot{K}(u)'du$ *is positive definite and*

$$\int |K(u)|\{1+|u|^s\}du + \int |\dot{K}(u)|\{1+|u|^2\}du < \infty.$$

As $n\to\infty$, *it holds* $\min(nh_n^{d+2},1)nh_n^{2s}\to 0$ *and* $n^2h_n^d\to\infty$.

Assumption MW.

(i): *There are positive constants* C *and* C_1 *such that* $\max_{1\le i\le n}P_{ii}\le C<1$ *and* $C_1^{-1}\le \pi'\left(\frac{1}{n}\sum_{i=1}^n Z_iZ_i'\right)\pi\le C_1$ *for all* n *large enough. Also,* $n^{-2}\sum_{i=1}^n|\pi'Z_i|^4\to 0$ *as* $n\to\infty$.

(ii): $\{U_i,\epsilon_i\}_{i=1}^n$ *are independent with* $E[U_i]=0$ *and* $E[\epsilon_i]=0$. *Also for some positive constant* C_2, *the minimum eigenvalue of* $Var(U_i,\epsilon_i)$ *is larger than* C_2^{-1} *and* $\max_{1\le i\le n}\{E[U_i^2],E[U_i^4],E[\epsilon_i^2],E[\epsilon_i^4]\}<C_2$.

(iii): Σ, Ψ, *and* Ξ *exist. Also* $\sqrt{K}/\mu_n^2\to 0$ *as* $n\to\infty$.

Assumption MR. *Let* $\lambda_{\min}(\cdot)$ *denote the minimum eigenvalue of its argument.*

(i): $\{Y_i,X_i,Z_i\}_{i=1}^n$ *is independent and identically distributed.*

(ii): $Pr(\lambda_{\min}(\sum_{i=1}^n Z_iZ_i')>0)\to 1$, *and*

$$\max_{1\le i\le n}|Z_i'\gamma|^2+\max_{1\le i\le n}\{E[U_i^4|X_i,Z_i]+E[||V_i||^4|Z_i]\}+\max_{1\le i\le n}\{1/E[U_i^2|X_i,Z_i]+1/\lambda_{min}(E[V_iV_i'|Z_i])\}=O_p(1),$$

with $V_i=X_i-E[X_i|Z_i]$.

(iii): $E[||X_i||^2]=O(1)$, $nE[(E[U_i|X_i,Z_i])^2]=o(1)$, *and* $\max_{1\le i\le n}||\tilde{X}_i||/\sqrt{n}=o_p(1)$.

A.2. **Proof of Theorem SP.** To simplify the presentation, we focus on the case where both g and μ are scalar-valued functions. First, by Lemmas 2, 4, and 3 below, the same argument as in the proof of Owen (1990, eq. (2.14)) guarantees $\hat{\lambda}=O_p(n^{-1/2})$.

Next, we obtain an asymptotic approximation for $\hat{\lambda}$. The first-order condition for $\hat{\lambda}$ satisfies

$$0=\frac{1}{n}\sum_{i=1}^n\frac{V_i(\theta)}{1+\hat{\lambda}V_i(\theta)}=\frac{1}{n}\sum_{i=1}^n V_i(\theta)-\frac{1}{n}\sum_{i=1}^n V_i(\theta)^2\hat{\lambda}+\frac{1}{n}\sum_{i=1}^n\frac{V_i(\theta)^3\hat{\lambda}^2}{1+\hat{\lambda}V_i(\theta)},$$

where the second equality follows from the identity $(1+x)^{-1}=1-x+x^2(1+x)^{-1}$. By applying Lemmas 2, 4, and 3, and $\hat{\lambda}=O_p(n^{-1/2})$, we have

$$\hat{\lambda}=\frac{\sum_{i=1}^n V_i(\theta)}{\sum_{i=1}^n V_i(\theta)^2}+o_p(n^{-1/2}).$$

By using this expansion for $\hat{\lambda}$, a Taylor expansion yields

$$2\sum_{i=1}^n\log(1+\hat{\lambda}V_i(\theta))=2\sum_{i=1}^n\left[\hat{\lambda}V_i(\theta)-\frac{1}{2}\{\hat{\lambda}V_i(\theta)\}^2\right]+o_p(1)=\frac{\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n V_i(\theta)\right]^2}{\frac{1}{n}\sum_{i=1}^n V_i(\theta)^2}+o_p(1).$$

The conclusion follows by Lemmas 2 and 3.

A.2.1. *Lemmas for Theorem SP.* Let $\hat{f}_j = \hat{f}(X_j)$, $\hat{\mu}_j = \hat{\mu}(X_j)$, and $\hat{\mu}_j^{(i)} = \hat{\mu}^{(i)}(X_j)$. We use the following identities.

Lemma 1. *It holds*

$$\hat{\mu}_j - \hat{\mu}_j^{(i)} = \frac{1}{n-2}\left\{\frac{1}{\hat{f}_j}\frac{1}{h^d}K\left(\frac{X_j - X_i}{h}\right)Y_i - \hat{\mu}_j\right\}, \tag{A.1}$$

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{n-1}\sum_{j\neq i}a(X_j)\right) = \frac{1}{n}\sum_{j=1}^{n}a(X_j), \tag{A.2}$$

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{n-1}\sum_{j\neq i}a(X_j)\hat{\mu}_j^{(i)}\right) = \frac{1}{n}\sum_{j=1}^{n}a(X_j)\hat{\mu}_j = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{n-1}\sum_{j\neq i}a(X_j)\hat{\mu}_j\right). \tag{A.3}$$

for any function a.

Proof. Let $K_{jk} = K\left(\frac{X_j - X_k}{h}\right)$. For (A.1), note that

$$\begin{aligned}
\hat{\mu}_j - \hat{\mu}_j^{(i)} &= \frac{1}{\hat{f}_j}\frac{1}{n-1}\sum_{k\neq j}\frac{1}{h^d}K_{jk}Y_k - \frac{1}{\hat{f}_j}\frac{1}{n-2}\left\{\sum_{k\neq j}\frac{1}{h^d}K_{jk}Y_k - \frac{1}{h^d}K_{ji}Y_i\right\} \\
&= -\frac{1}{\hat{f}_j}\frac{1}{(n-1)(n-2)}\sum_{k\neq j}\frac{1}{h^d}K_{jk}Y_k + \frac{1}{\hat{f}_j}\frac{1}{n-2}\frac{1}{h^d}K_{ji}Y_i \\
&= \frac{1}{n-2}\left\{\frac{1}{\hat{f}_j}\frac{1}{h^d}K_{ji}Y_i - \hat{\mu}_j\right\}.
\end{aligned}$$

For (A.2), note that

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{n-1}\sum_{j\neq i}a(X_j)\right) &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{n-1}\sum_{j=1}^{n}a(X_j) - \frac{1}{n-1}a(X_i)\right) \\
&= \frac{1}{n-1}\sum_{j=1}^{n}a(X_j) - \frac{1}{n(n-1)}\sum_{i=1}^{n}a(X_i) = \frac{1}{n}\sum_{j=1}^{n}a(X_j).
\end{aligned}$$

For (A.3), note that

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{n-1}\sum_{j\neq i}a(X_j)\hat{\mu}_j^{(i)}\right)\\
= &\quad \frac{1}{n(n-1)(n-2)}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\neq i,j}a(X_j)\frac{1}{\hat{f}_j}\frac{1}{h^d}K_{jk}Y_k\\
= &\quad \frac{1}{n(n-1)(n-2)h^d}\sum_{i=1}^{n}\left\{\sum_{j=1}^{n}\sum_{k\neq j}a(X_j)\frac{1}{\hat{f}_j}K_{jk}Y_k-\sum_{j\neq i}a(X_j)\frac{1}{\hat{f}_j}K_{ji}Y_i-\sum_{k\neq i}a(X_i)\frac{1}{\hat{f}_i}K_{ik}Y_k\right\}\\
= &\quad \frac{1}{n-2}\sum_{j=1}^{n}a(X_j)\frac{1}{\hat{f}_j}\left(\frac{1}{n-1}\sum_{k\neq j}\frac{1}{h^d}K_{jk}Y_k\right)\\
&\quad -\frac{1}{n(n-2)}\left\{\frac{1}{n-1}\sum_{i=1}^{n}\sum_{j\neq i}a(X_j)\frac{1}{h^d}\frac{1}{\hat{f}_j}K_{ji}Y_i+\sum_{i=1}^{n}a(X_i)\frac{1}{\hat{f}_i}\left(\frac{1}{n-1}\sum_{k\neq i}\frac{1}{h^d}K_{ik}Y_k\right)\right\}\\
= &\quad \left(\frac{n}{n-2}-\frac{2}{n-2}\right)\frac{1}{n}\sum_{j=1}^{n}a(X_j)\hat{\mu}_j=\frac{1}{n}\sum_{j=1}^{n}a(X_j)\hat{\mu}_j.
\end{aligned}
$$

Thus the first equality of (A.3) follows. The second equality of (A.3) follows from (A.2).

Hereafter, by suppressing Z_j, θ, and X_j, we denote by $\mu_j=\mu(X_j)$, $g_j(\hat{\mu}_j)=g(Z_j,\theta,\hat{\mu}(X_j))$, $g_j(\hat{\mu}_j^{(i)})=g(Z_j,\theta,\hat{\mu}^{(i)}(X_j))$, $g_{1j}(\mu_j)=\frac{\partial}{\partial\mu}g(Z_j,\theta,\mu(X_j))$, and $g_{2j}(\mu_j)=\frac{\partial^2}{\partial\mu^2}g(Z_j,\theta,\mu(X_j))$.

Lemma 2. *Under Assumption SP,*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i(\theta)\xrightarrow{d}N(0,\Omega),$$

where $\Omega=E[\psi(Z,X)\psi(Z,X)']$ *with*
$\psi(Z,X)=-E\left[\frac{\partial g(Z,\theta,\mu(X))}{\partial\theta'}\right]\left\{g(Z,\theta,\mu(X))+E\left[\frac{\partial g(Z,\theta,\mu(X))}{\partial\mu'}\middle| X\right]\{Y-\mu(X)\}\right\}.$

Proof. We can write

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i(\theta) &= \sqrt{n}S(\theta)-\frac{n-1}{\sqrt{n}}\sum_{i=1}^{n}\{S^{(i)}(\theta)-S(\theta)\}\\
&= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}g_j(\hat{\mu}_j)-\frac{n-1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{1}{n-1}\sum_{j\neq i}g_j(\hat{\mu}_j^{(i)})-\frac{1}{n}\sum_{j=1}^{n}g_j(\hat{\mu}_j)\right\}\\
&\equiv M_1-M_2.
\end{aligned}
$$

By Newey (Theorem 4.2, 1994b), Assumption SP guarantees $M_1 \xrightarrow{d} N(0,\Omega)$. Thus, it is enough to show that $M_2 \xrightarrow{p} 0$. An expansion of $g_j(\hat{\mu}_j^{(i)})$ around $\hat{\mu}_j^{(i)} = \hat{\mu}_j$ yields

$$\begin{aligned} M_2 &= \sqrt{n}(n-1)\left[\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1}{n-1}\sum_{j\neq i} g_j(\hat{\mu}_j^{(i)})\right\} - \frac{1}{n}\sum_{j=1}^{n} g_j(\hat{\mu}_j)\right] \\ &\quad + \sqrt{n}(n-1)\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{n-1}\sum_{j\neq i} g_{1j}(\hat{\mu}_j)(\hat{\mu}_j^{(i)} - \hat{\mu}_j)\right] + \sqrt{n}(n-1)\frac{1}{n}\sum_{i=1}^{n} R_i \\ &\equiv M_{21} + M_{22} + M_{23}, \end{aligned}$$

where

$$R_i = \frac{1}{2(n-1)}\sum_{j\neq i} g_{2j}(\bar{\mu}_j^{(i)})(\hat{\mu}_j^{(i)} - \hat{\mu}_j)^2, \tag{A.4}$$

and $\bar{\mu}_j^{(i)}$ lies between $\hat{\mu}_j$ and $\hat{\mu}_j^{(i)}$. By (A.2) and (A.3), we have $M_{21} = M_{22} = 0$. For $M_{23} = o_p(1)$, it is enough to show that

$$\frac{1}{n}\sum_{i=1}^{n} R_i = o_p(n^{-3/2}). \tag{A.5}$$

By using Lemma 1, decompose

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} R_i &= \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{2(n-1)}\sum_{j\neq i} g_{2j}(\bar{\mu}_j^{(i)})\frac{1}{(n-2)^2}\left\{\frac{1}{\hat{f}_j}\frac{1}{h^d}K\left(\frac{X_j - X_i}{h}\right)Y_i - \hat{\mu}_j\right\}^2\right] \\ &= \frac{1}{2n(n-1)(n-2)^2}\sum_{i=1}^{n}\sum_{j\neq i} g_{2j}(\bar{\mu}_j^{(i)})\frac{1}{\hat{f}_j^2}\frac{1}{h^{2d}}K\left(\frac{X_j - X_i}{h}\right)^2 Y_i^2 \\ &\quad - \frac{1}{n(n-1)(n-2)^2}\sum_{i=1}^{n}\sum_{j\neq i} g_{2j}(\bar{\mu}_j^{(i)})\frac{1}{\hat{f}_j}\frac{1}{h^d}K\left(\frac{X_j - X_i}{h}\right)Y_i\hat{\mu}_j \\ &\quad + \frac{1}{2n(n-1)(n-2)^2}\sum_{i=1}^{n}\sum_{j\neq i} g_{2j}(\bar{\mu}_j^{(i)})\hat{\mu}_j^2 \\ &\equiv A_1 - A_2 + A_3. \end{aligned}$$

Note that by applying Hansen (2008, Theorem 10) under Assumption SP, it holds.

$$\max_{1\le j\le n}|\hat{\mu}_j - \mu_j| = o_p(n^{-1/4}), \qquad \max_{1\le j\le n}|\hat{f}_j - f(X_j)| = o_p(n^{-1/4}). \tag{A.6}$$

For A_1, since g_2 is assumed to be bounded, it holds

$$|A_1| \le \frac{C_1}{n^4 h^{2d}}\sum_{i=1}^{n}\sum_{j\neq i}\frac{1}{\hat{f}_j^2}K\left(\frac{X_j - X_i}{h}\right)^2 Y_i^2,$$

for some $C_1 > 0$. Due to (A.6) and the law of large numbers, the assumption $nh^{2d}/(\log n)^2 \to \infty$ guarantees $A_1 = o_p(n^{-3/2})$. Similarly, for A_2, since g_2 and K are assumed to be bounded, it holds

$$|A_2| \le \frac{C_2}{n^4 h^d}\sum_{i=1}^{n}\sum_{j\neq i}|Y_i||\hat{\mu}_j\hat{f}_j^{-1}|,$$

for some $C_2 > 0$. Due to (A.6) and the law of large numbers, the assumption $nh^{2d}/(\log n)^2 \to \infty$ guarantees $A_2 = o_p(n^{-3/2})$. Finally, for A_3, it holds

$$|A_3| \leq \frac{C_3}{n^4} \sum_{i=1}^{n} \sum_{j \neq i} \hat{\mu}_j^2,$$

for some $C_3 > 0$. Due to (A.6), we have $A_3 = o_p(n^{-3/2})$. Therefore, the conclusion is obtained.

Lemma 3. *Under Assumption SP,*

$$\frac{1}{n} \sum_{i=1}^{n} V_i(\theta)^2 \xrightarrow{p} \Omega.$$

Proof. Note that

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^{n} V_i(\theta)^2 &= S(\theta)^2 - 2(n-1)S(\theta)\frac{1}{n}\sum_{i=1}^{n}\{S^{(i)}(\theta) - S(\theta)\} + (n-1)^2 \frac{1}{n}\sum_{i=1}^{n}\{S^{(i)}(\theta) - S(\theta)\}^2 \\ &\equiv N_1 - 2N_2 + N_3. \end{aligned}$$

First, since $S(\theta) = \frac{1}{n}\sum_{j=1}^{n} g_j(\hat{\mu}_j) = O_p(n^{-1/2})$ (by Newey, 1994b, Theorem 4.2), it holds $N_1 = o_p(1)$. An expansion of $g_j(\hat{\mu}_j^{(i)})$ around $\hat{\mu}_j^{(i)} = \hat{\mu}_j$ yields

$$\begin{aligned} N_2 &= (n-1)\left(\frac{1}{n}\sum_{j=1}^{n} g_j(\hat{\mu}_j)\right)\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{n-1}\sum_{j\neq i} g_j(\hat{\mu}_j) - \frac{1}{n}\sum_{j=1}^{n} g_j(\hat{\mu}_j)\right) \\ &\quad + (n-1)\left(\frac{1}{n}\sum_{j=1}^{n} g_j(\hat{\mu}_j)\right)\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{n-1}\sum_{j\neq i} g_{1j}(\hat{\mu}_j)(\hat{\mu}_j^{(i)} - \hat{\mu}_j)\right) + (n-1)\left(\frac{1}{n}\sum_{j=1}^{n} g_j(\hat{\mu}_j)\right)\frac{1}{n}\sum_{i=1}^{n} R_i, \end{aligned}$$

where R_i is defined in (A.4). Since the first and second terms are zero by Lemma 1 and the third term is $o_p(n^{-1/2})$ by (A.5), we have $N_2 = o_p(1)$.

For N_3, we have

$$\begin{aligned} N_3 &= (n-1)^2 \frac{1}{n}\sum_{i=1}^{n}\left(S^{(i)}(\theta) - \frac{1}{n}\sum_{i=1}^{n} S^{(i)}(\theta)\right)^2 + (n-1)^2\left(\frac{1}{n}\sum_{i=1}^{n} S^{(i)}(\theta) - S(\theta)\right)^2 \\ &= (n-1)^2 \frac{1}{n}\sum_{i=1}^{n}\left(S^{(i)}(\theta) - \frac{1}{n}\sum_{i=1}^{n} S^{(i)}(\theta)\right)^2 + o_p(1) \\ &= (n-1)^2 \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=i+1}^{n}\{S^{(i)}(\theta) - S^{(j)}(\theta)\}^2 + o_p(1), \end{aligned}$$

where the second equality follows from the same argument as in Lemma 2, and the third equality follows from direct calculation (Efron and Stein, 1981, p. 589). Combining these results with $S^{(i)}(\theta) = \frac{1}{n}\sum_{j\neq i}\psi(Z_j, X_j) + o_p(n^{-1/2})$ (by applying Newey, 1994b, Theorem 4.2), we have

$$\frac{1}{n}\sum_{i=1}^{n} V_i(\theta)^2 = N_3 + o_p(1) = \frac{(n-1)^2}{n^4}\sum_{i=1}^{n}\sum_{j=i+1}^{n}\{\psi(Z_j, X_j) - \psi(Z_i, X_i)\}^2 + o_p(1).$$

Thus, the conclusion follows by the law of large numbers.

Lemma 4. *Under Assumption SP, it holds*

$$\max_{1\le i\le n}|V_i(\theta)| = o_p(n^{1/2}).$$

Proof. By an expansion around $\hat{\mu}_j^{(i)} = \hat{\mu}_j$, decompose

$$\begin{aligned}\max_{1\le i\le n}|V_i(\theta)| &\le \max_{1\le i\le n}|g_i(\hat{\mu}_i)| + \max_{1\le i\le n}\left|\sum_{j\ne i} g_{1j}(\hat{\mu}_j)\{\hat{\mu}_j^{(i)} - \hat{\mu}_j\}\right| + \max_{1\le i\le n}\left|\sum_{j\ne i} g_{2j}(\bar{\mu}_j^{(i)})\{\hat{\mu}_j^{(i)} - \hat{\mu}_j\}^2\right| \\ &\equiv T_1 + T_2 + T_3,\end{aligned}$$

where $\bar{\mu}_j^{(i)}$ lies between $\hat{\mu}_j$ and $\hat{\mu}_j^{(i)}$. For T_1, an expansion around $\hat{\mu}_i = \mu_i$ and boundedness of g_2 yield

$$T_1 \le \max_{1\le i\le n}|g_i(\mu_i)| + \max_{1\le i\le n}|g_{1i}(\mu_i)|\max_{1\le i\le n}|\hat{\mu}_i - \mu_i| + C\max_{1\le i\le n}|\hat{\mu}_i - \mu_i|^2,$$

for some $C > 0$. From $E[g_i(\mu_i)^2] < \infty$ and $E[g_{1i}(\mu_i)^2] < \infty$ (guaranteed by Assumption SP), we have $\max_{1\le i\le n}|g_i(\mu_i)| = o_p(n^{1/2})$ and $\max_{1\le i\le n}|g_{1i}(\mu_i)| = o_p(n^{1/2})$. Thus, (A.6) implies $T_1 = o_p(n^{1/2})$.

For T_2, an expansion around $\hat{\mu}_i = \mu_i$ and boundedness of g_2 yield

$$\begin{aligned}T_2 &\le \max_{1\le i\le n}\left|\sum_{j\ne i} g_{1j}(\mu_j)(\hat{\mu}_j^{(i)} - \hat{\mu}_j)\right| + C\max_{1\le i\le n}\left|\sum_{j\ne i}(\hat{\mu}_j - \mu_j)(\hat{\mu}_j^{(i)} - \hat{\mu}_j)\right| \\ &\equiv T_{21} + T_{22}.\end{aligned}$$

For T_{21}, Lemma 1 yields

$$\begin{aligned}T_{21} &\le \frac{1}{nh^d}\max_{1\le i\le n}\left|\sum_{j\ne i} g_{1j}(\mu_j)\frac{1}{\hat{f}_j}K\left(\frac{X_j - X_i}{h}\right)Y_i\right| + \frac{1}{nh^d}\max_{1\le i\le n}\left|\sum_{j\ne i} g_{1j}(\mu_j)\hat{\mu}_j\right| \\ &\equiv T_{211} + T_{212}.\end{aligned}$$

For T_{211}, due to boundedness of K,

$$\begin{aligned}T_{211} &\le \frac{C}{h^d}\max_{1\le i\le n}|Y_i|\cdot\max_{1\le j\le n}\left|\frac{1}{\hat{f}_j} - \frac{1}{f_j}\right|\cdot\frac{1}{n}\sum_{j=1}^n|g_{1j}(\mu_j)| \\ &\quad + \max_{1\le i\le n}|Y_i|\cdot\max_{1\le i\le n}\left|\frac{1}{nh^d}\sum_{j\ne i} g_{1j}(\mu_j)\frac{1}{f_j}K\left(\frac{X_j - X_i}{h}\right)\right| \\ &= o_p(n^{1/2}), \qquad \text{(A.7)}\end{aligned}$$

where the equality follows from the assumption $nh^{2d}/(\log n)^2 \to \infty$, $\max_{1\le i\le n}|Y_i| = o_p(n^{1/4})$ (by the assumption $E|Y|^p < \infty$ for $p \ge 4$), (A.6), the law of large numbers, and the uniform convergence of $\frac{1}{nh^d}\sum_{j\ne i} g_{1j}(\mu_j)\frac{1}{f_j}K\left(\frac{X_j - x}{h}\right)$ over x (as in Hansen, 2008, Theorem 10). Similarly, for T_{212}, the assumption $nh^{2d}/(\log n)^2 \to \infty$, (A.6), and the law of large numbers imply $T_{212} = o_p(n^{1/2})$. Combining these results, $T_{21} = o_p(n^{1/2})$. Also, a similar argument guarantees $T_{22} = o_p(n^{1/2})$.

For T_3, boundedness of g and Lemma 1 imply

$$T_3 \leq \frac{C}{n^2} \max_{1\leq i\leq n} \sum_{j\neq i} \left(\frac{1}{\hat{f}_j} \frac{1}{h^d} K\left(\frac{X_j - X_i}{h}\right) Y_i - \hat{\mu}_j \right)^2 = o_p(n^{1/2}),$$

where the equality follows from a similar argument to the proof of (A.7). Therefore, the conclusion is obtained.

A.3. **Proof of Theorem SB1.** To simplify the presentation, suppose θ is scalar. We only prove the case of $nh^{d+2} \to \kappa$. Other cases are shown in similar ways. As in the proof of Theorem SP, we can prove the asymptotic equivalence

$$\ell(\theta) = \left[\frac{1}{n}\sum_{i=1}^n V_i(\theta)^2\right]^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n V_i(\theta)\right)^2 + o_p(1).$$

Thus, it is enough to show

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n V_i(\theta) \quad \stackrel{d}{\to} \quad N(0, \Sigma + 2\kappa^{-1}\Delta), \tag{A.8}$$

$$\frac{1}{n}\sum_{i=1}^n V_i(\theta)^2 \quad \stackrel{p}{\to} \quad \Sigma + 4\kappa^{-1}\Delta. \tag{A.9}$$

For (A.8), since $\sum_{i=1}^n \{S^{(i)}(\theta) - S(\theta)\} = 0$, it holds

$$\begin{aligned} \frac{1}{\sqrt{n}}\sum_{i=1}^n V_i(\theta) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \{nS(\theta) - (n-1)S^{(i)}(\theta)\} \\ &= \sqrt{n}S(\theta) - (n-1)\frac{1}{\sqrt{n}}\sum_{i=1}^n \{S^{(i)}(\theta) - S(\theta)\} = \sqrt{n}S(\theta). \end{aligned}$$

Thus, (A.8) follows by Cattaneo, Crump and Jansson (2014, Theorem 1).

For (A.9), note that

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n V_i(\theta)^2 &= \frac{1}{n}\sum_{i=1}^n [S(\theta) - (n-1)\{S^{(i)}(\theta) - S(\theta)\}]^2 = S(\theta)^2 + \frac{(n-1)^2}{n}\sum_{i=1}^n \{S^{(i)}(\theta) - S(\theta)\}^2 \\ &= \frac{(n-1)^2}{n}\sum_{i=1}^n \{S^{(i)}(\theta) - S(\theta)\}^2 + o_p(1) \\ &= \frac{(n-1)^2}{n^2}\sum_{i=1}^n \sum_{j=i+1}^n \{S^{(i)}(\theta) - S^{(j)}(\theta)\}^2 + o_p(1), \end{aligned}$$

where the second equality follows from $\sum_{i=1}^n \{S^{(i)}(\theta) - S(\theta)\} = 0$, the third equality follows from $S(\theta) = O_p(n^{-1/2})$ (by (A.8)), and the last equality follows from Efron and Stein (1981, p. 589). Now, decompose

$$S^{(i)}(\theta) = B^{(i)} + \bar{L}^{(i)} + \bar{W}^{(i)},$$

where

$$B^{(i)} = E[\hat{\theta}^{(i)}] - \theta, \qquad \bar{L}^{(i)} = \frac{1}{n-1}\sum_{j\neq i} L_j, \qquad \bar{W}^{(i)} = \binom{n-1}{2}^{-1} \sum_{j\neq i}\sum_{k>j,k\neq i} W_{jk}.$$

By plugging this into the above equation combined with Efron and Stein (1981, eq. (2.3)) and Cattaneo Crump and Jansson (2014, eq. (9)),

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} V_i(\theta)^2 &= \sum_{i=1}^{n}\sum_{j=i+1}^{n}\left[\frac{1}{n-1}(L_j - L_i) + \frac{2}{(n-1)^2}\sum_{k\neq i,j}(W_{jk} - W_{ik})\right]^2 + o_p(1) \\ &= \Sigma + 4\kappa^{-1}\Delta + o_p(1).\end{aligned}$$

Therefore, (A.9) is obtained.

A.4. **Proof of Theorem SB2.** Again, we only prove the case of $nh^{d+2} \to \kappa$ with scalar θ. Other cases are shown in similar ways. As in the proof of Theorem SP, we can prove the asymptotic equivalence

$$\ell^m(\theta) = \left[\frac{1}{n}\sum_{i=1}^{n} V_i^m(\theta)^2\right]^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i^m(\theta)\right)^2 + o_p(1).$$

Thus, it is enough to show

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i^m(\theta) \overset{d}{\to} N(0, \Sigma + 4\kappa^{-1}\Delta), \tag{A.10}$$

$$\frac{1}{n}\sum_{i=1}^{n} V_i^m(\theta)^2 \overset{p}{\to} \Sigma + 4\kappa^{-1}\Delta. \tag{A.11}$$

For (A.10), since $\sum_{i=1}^n V_i(\hat{\theta}) = 0$, we have

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i^m(\theta) &= \hat{\Gamma}\tilde{\Gamma}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i(\theta) \\ &= \sqrt{\frac{n^{-1}\sum_{i=1}^n V_i(\hat{\theta})^2}{n^{-1}\sum_{i=1}^n V_i(\hat{\theta})^2 - n^{-1}\sum_{i=1}^n\sum_{j=i+1}^n Q_{ij}^2}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i(\theta).\end{aligned}$$

By (A.8), it holds $\frac{1}{\sqrt{n}}\sum_{i=1}^n V_i(\theta) \overset{d}{\to} N(0, \Sigma + 2\kappa^{-1}\Delta)$. Also a similar argument to (A.9) yields $\frac{1}{n}\sum_{i=1}^n V_i(\hat{\theta})^2 \overset{p}{\to} \Sigma + 4\kappa^{-1}\Delta$. Thus, for (A.10), it remains to show that

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=i+1}^{n} Q_{ij}^2 \overset{p}{\to} 2\kappa^{-1}\Delta. \tag{A.12}$$

By the same argument as in the proof of Cattaneo, Crump and Jansson (2014, Theorem 2), we have

$$Q_{ij} = \frac{2}{n-2}\left\{W_{ij} - \frac{1}{n-1}\sum_{k\neq j} W_{kj} - \frac{1}{n-1}\sum_{l\neq i} W_{il} + \frac{2}{n(n-1)}\sum_{k=1}^{n}\sum_{l\neq k} W_{kl}\right\} + o_p(n^{-1}). \tag{A.13}$$

Thus by using $E[W_{ij}] = E[W_{ij}W_{kj}] = E[W_{ij}W_{kl}] = 0$, we obtain

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=i+1}^{n} Q_{ij}^2 = \frac{4\kappa^{-1}}{(n-2)^2}\sum_{i=1}^{n}\sum_{j=i+1}^{n} h^{d+2}W_{ij}^2 + o_p(1) \overset{p}{\to} 2\kappa^{-1}\Delta.$$

A.5. **Proof of Theorem MW1.** We only prove for Case (ii). Other cases are shown in similar ways. As in the proof of Theorem SP, we can prove the asymptotic equivalence

$$\ell(\theta) = \left[\frac{(n-1)^2}{\mu_n^2}\sum_{i=1}^n V_i(\theta)^2\right]^{-1}\left(\frac{n-1}{\mu_n}\sum_{i=1}^n V_i(\theta)\right)^2 + o_p(1).$$

Thus, it is enough to show

$$\frac{n-1}{\mu_n}\sum_{i=1}^n V_i(\theta) \quad \xrightarrow{d} \quad N(0, \Sigma + \alpha\Psi), \tag{A.14}$$

$$\frac{(n-1)^2}{\mu_n^2}\sum_{i=1}^n V_i(\theta)^2 \quad \xrightarrow{p} \quad \Sigma + \Xi + 2\alpha\Psi. \tag{A.15}$$

For (A.14), note that

$$\begin{aligned}\sum_{i=1}^n S^{(i)}(\theta) &= \frac{1}{(n-1)(n-2)}\sum_{i=1}^n\left(\sum_{k=1}^n\sum_{l\neq k} X_k P_{kl} U_l - \sum_{k\neq i} X_k P_{ki} U_i - \sum_{l\neq i} X_i P_{il} U_l\right)\\ &= \frac{n}{(n-1)(n-2)}\sum_{k=1}^n\sum_{l\neq k} X_k P_{kl} U_l - \frac{2}{(n-1)(n-2)}\sum_{i=1}^n\sum_{k\neq i} X_k P_{ki} U_i\\ &= \frac{1}{n-1}\sum_{k=1}^n\sum_{l\neq k} X_k P_{kl} U_l = nS(\theta).\end{aligned}$$

Thus, $\frac{n-1}{\mu_n}\sum_{i=1}^n V_i(\theta) = \frac{1}{\mu_n}\sum_{k=1}^n\sum_{l\neq k} X_k P_{kl} U_l$, and (A.14) follows from Chao *et al.* (2012, Lemma A2).

We now prove (A.15). Observe that

$$\begin{aligned}\frac{(n-1)^2}{\mu_n^2}\sum_{i=1}^n V_i(\theta)^2 &= \frac{(n-1)^2}{\mu_n^2}\sum_{i=1}^n\left\{\frac{1}{n-1}\sum_{k=1}^n\sum_{l\neq k} X_k P_{kl} U_l - \frac{1}{n-2}\sum_{k\neq i}\sum_{l\neq i,k} X_k P_{kl} U_l\right\}^2\\ &= \frac{1}{\mu_n^2}\sum_{i=1}^n\left\{\sum_{k\neq i} X_k P_{ki} U_i + \sum_{l\neq i} X_i P_{il} U_l - \frac{1}{n-2}\sum_{k\neq i}\sum_{l\neq i,k} X_k P_{kl} U_l\right\}^2\end{aligned} \tag{A.16}$$

For the last term in (A.16), we have

$$\frac{1}{\mu_n^2}\sum_{i=1}^n\left\{\frac{1}{n-2}\sum_{k\neq i}\sum_{l\neq i,k} X_k P_{kl} U_l\right\}^2 = \frac{1}{\mu_n^2(n-1)}\sum_{k=1}^n\sum_{l\neq k} P_{kl}^2 E[X_k^2]E[U_l^2] + o_p(1) = O_p\left(\frac{K}{\mu_n^2 n}\right),$$

where the first equality follows from Chao *et al.* (2012, Lemmas A2 and A3) and the second equality follows from $\sum_{k=1}^n\sum_{l\neq k} P_{kl}^2 \leq \sum_{k=1}^n P_{kk} = K$. By using (4.1) and $\sum_{k=1}^n Z_k P_{ki} = Z_i$, the first two terms in (A.16) are written as

$$\begin{aligned}\sum_{k\neq i} X_k P_{ki} U_i &= \gamma_n' Z_i(1-P_{ii})U_i + \sum_{k\neq i}\epsilon_k P_{ki} U_i,\\ \sum_{l\neq i} X_i P_{il} U_l &= \sum_{l\neq i}\gamma_n' Z_i P_{il} U_l + \sum_{l\neq i}\epsilon_i P_{il} U_l.\end{aligned}$$

Combining these results,

$$\begin{aligned}\frac{(n-1)^2}{\mu_n^2}\sum_{i=1}^n V_i(\theta)^2 &= \frac{1}{\mu_n^2}\sum_{i=1}^n \left\{\gamma_n' Z_i(1-P_{ii})U_i + \sum_{k\neq i}\epsilon_k P_{ki}U_i + \sum_{l\neq i}\gamma_n' Z_i P_{il}U_l + \sum_{l\neq i}\epsilon_i P_{il}U_l\right\}^2 + o_p(1) \\ &= \frac{1}{\mu_n^2}\sum_{i=1}^n \left[\begin{array}{c}(\gamma_n' Z_i(1-P_{ii})U_i)^2 + \sum_{l\neq i}(\gamma_n' Z_i P_{il}U_l)^2 \\ +\sum_{k\neq i}(\epsilon_k P_{ki}U_i)^2 + \sum_{l\neq i}(\epsilon_i P_{il}U_l)^2 + 2\sum_{k\neq i}\epsilon_k U_k P_{ki}^2\epsilon_i U_i\end{array}\right] + o_p(1),\end{aligned}$$

where the second equality follows from a similar argument in the proof of Chao *et al.* (2012, Lemma A2). Therefore, Chao *et al.* (2012, Lemma A3) implies

$$\frac{1}{\mu_n^2}\sum_{i=1}^n(\gamma_n' Z_i(1-P_{ii})U_i)^2 \xrightarrow{p} \Sigma, \qquad \frac{1}{\mu_n^2}\sum_{i=1}^n\sum_{l\neq i}(\gamma_n' Z_i P_{il}U_l)^2 \xrightarrow{p} \Xi,$$

$$\frac{1}{\mu_n^2}\sum_{i=1}^n\sum_{k\neq i}(\epsilon_k P_{ki}U_i)^2 \xrightarrow{p} \alpha\lim_{n\to\infty}\frac{1}{K}\sum_{i=1}^n\sum_{k\neq i}E[\epsilon_k^2]P_{ik}^2\sigma_i^2,$$

$$\frac{1}{\mu_n^2}\sum_{i=1}^n\sum_{k\neq i}\epsilon_k U_k P_{ki}^2\epsilon_i U_i \xrightarrow{p} \alpha\lim_{n\to\infty}\frac{1}{K}\sum_{i=1}^n\sum_{k\neq i}E[\epsilon_k U_k]P_{ki}^2 E[\epsilon_i U_i],$$

and we obtain (A.15). Therefore, the conclusion follows.

A.6. **Proof of Theorem MW2.** Again, we only prove for Case (ii). Other cases are shown in similar ways. As in the proof of Theorem SP, we can prove the asymptotic equivalence

$$\ell^m(\theta) = \left[\frac{(n-1)^2}{\mu_n^2}\sum_{i=1}^n V_i^m(\theta)^2\right]^{-1}\left(\frac{n-1}{\mu_n}\sum_{i=1}^n V_i^m(\theta)\right)^2 + o_p(1).$$

Thus, it is enough to show

$$\frac{n-1}{\mu_n}\sum_{i=1}^n V_i^m(\theta) \xrightarrow{d} N(0, \Sigma+\Xi+2\alpha\Psi), \tag{A.17}$$

$$\frac{(n-1)^2}{\mu_n^2}\sum_{i=1}^n V_i^m(\theta)^2 \xrightarrow{p} \Sigma+\Xi+2\alpha\Psi. \tag{A.18}$$

For (A.17), since $\sum_{i=1}^n V_i(\hat{\theta}) = 0$, we have

$$\begin{aligned}\frac{n-1}{\mu_n}\sum_{i=1}^n V_i^m(\theta) &= \hat{\Gamma}\tilde{\Gamma}^{-1}\frac{n-1}{\mu_n}\sum_{i=1}^n V_i(\theta) \\ &= \sqrt{\frac{\frac{(n-1)^2}{\mu_n^2}\sum_{i=1}^n V_i(\hat{\theta})^2}{\frac{(n-1)^2}{\mu_n^2}\sum_{i=1}^n V_i(\hat{\theta})^2 - \frac{(n-1)^2}{\mu_n^2}\sum_{i=1}^n\sum_{j=i+1}^n Q_{ij}^2}}\,\frac{n-1}{\mu_n}\sum_{i=1}^n V_i(\theta).\end{aligned}$$

By (A.14), it holds $\frac{n-1}{\mu_n}\sum_{i=1}^n V_i(\theta) \xrightarrow{d} N(0,\Sigma+\alpha\Psi)$. Also a similar argument to (A.15) yields $\frac{(n-1)^2}{\mu_n^2}\sum_{i=1}^n V_i(\hat{\theta})^2 \xrightarrow{p} \Sigma+\Xi+2\alpha\Psi$. Thus, for (A.10), it remains to show that

$$\frac{(n-1)^2}{\mu_n^2}\sum_{i=1}^n\sum_{j=i+1}^n Q_{ij}^2 \xrightarrow{p} \Xi+\alpha\Psi. \tag{A.19}$$

Note that

$$\begin{aligned}
&\frac{(n-1)^2}{\mu_n^2}\sum_{i=1}^{n}\sum_{j=i+1}^{n} Q_{ij}^2 \\
= &\ \frac{1}{2\mu_n^2}\sum_{i=1}^{n}\sum_{j\neq i}\left[\begin{array}{c}\sum_{k=1}^{n}\sum_{l\neq k} X_k P_{kl} U_l - \sum_{k\neq i}\sum_{l\neq i,k} X_k P_{kl} U_l \\ -\sum_{k\neq j}\sum_{l\neq j,k} X_k P_{kl} U_l + \sum_{k\neq i,j}\sum_{l\neq i,j,k} X_k P_{kl} U_l\end{array}\right]^2 + o_p(1) \\
= &\ \frac{1}{2\mu_n^2}\sum_{i=1}^{n}\sum_{j\neq i}(\gamma_n' Z_j P_{ji} U_i + \epsilon_j P_{ji} U_i + \gamma_n' Z_i P_{ij} U_j + \epsilon_i P_{ij} U_j)^2 + o_p(1) \\
= &\ \frac{1}{\mu_n^2}\sum_{i=1}^{n}\sum_{j\neq i}\{(\gamma_n' Z_j P_{ji} U_i)^2 + (\epsilon_j P_{ji} U_i)^2 + \epsilon_i U_i P_{ij}^2 \epsilon_j U_j\} + o_p(1),
\end{aligned}$$

where the first and third equalities follow from Chao *et al.* (2012, Lemmas A2 and A3), the second equality follows from direct calculation and (4.1). Therefore, we have (A.19) due to the following results (Chao *et al.*, 2012, Lemmas A3):

$$\begin{aligned}
\frac{1}{\mu_n^2}\sum_{i=1}^{n}\sum_{j\neq i}(\gamma_n' Z_i P_{ij} U_j)^2 &\overset{p}{\to} \Xi, \\
\frac{1}{\mu_n^2}\sum_{i=1}^{n}\sum_{j\neq i}(\epsilon_j P_{ji} U_i)^2 &\overset{p}{\to} \alpha \lim_{n\to\infty}\frac{1}{K}\sum_{i=1}^{n}\sum_{j\neq i} P_{ij}^2 \sigma_i^2 E[\epsilon_j^2], \\
\frac{1}{\mu_n^2}\sum_{i=1}^{n}\sum_{j\neq i}\epsilon_j U_j P_{ji}^2 \epsilon_i U_i &\overset{p}{\to} \alpha \lim_{n\to\infty}\frac{1}{K}\sum_{i=1}^{n}\sum_{j\neq i} P_{ji}^2 E[\epsilon_j U_j] E[\epsilon_i U_i].
\end{aligned}$$

A.7. **Proof of Theorem MR1.** We only prove for Case (ii). Case (i) can be shown in the same manner. Let

$$\begin{aligned}
\Sigma &= p\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\tilde{X}_i^2 U_i^2, \qquad \Psi = p\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{l\neq i}\tilde{X}_i M_{il} U_l\right)^2, \\
\Xi_1 &= p\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(\tilde{X}_i P_{ii} Z_i'\gamma), \qquad \Xi_2 = p\lim_{n\to\infty}\frac{1}{n^3}\sum_{i=1}^{n}\left(\sum_{k\neq i}\tilde{X}_k M_{kk} Z_k'\gamma\right)^2.
\end{aligned}$$

As in the proof of Theorem SP, we can prove the asymptotic equivalence

$$\ell(\theta) = \left[\frac{1}{n}\sum_{i=1}^{n} V_i(\theta)^2\right]^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i(\theta)\right)^2 + o_p(1).$$

Thus, it is enough to show

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i(\theta) \overset{d}{\to} N(0,\Sigma), \tag{A.20}$$

$$\frac{1}{n}\sum_{i=1}^{n} V_i(\theta)^2 \overset{p}{\to} \Sigma + \Psi + \Xi_1 + \Xi_2, \tag{A.21}$$

For (A.20), a direct calculation yields

$$\sum_{i=1}^{n} S^{(i)}(\theta) = \sum_{k=1}^{n} \tilde{X}_k M_{kk}(Y_k - X_k\theta) + \sum_{k=1}^{n}\sum_{l\neq k} \tilde{X}_k M_{kl}(Y_l - X_l\theta) = nS(\theta).$$

Thus,

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i(\theta) &= \sqrt{n}S(\theta) = \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\sum_{l=1}^{n} \tilde{X}_k M_{kl} U_l = \frac{1}{\sqrt{n}}\sum_{l=1}^{n} \tilde{X}_l U_l \\ &\xrightarrow{d} N(0,\Sigma),\end{aligned}$$

where the second equality follows from (5.1) and $\sum_{l=1}^{n} M_{kl} Z_l' = 0$, the third equality follows from $\sum_{k=1}^{n} \tilde{X}_k M_{kl} = \tilde{X}_l$, and the convergence follows from Cattaneo, Jansson and Newey (2018a, Lemma SA-2).

We now prove (A.21). Decompose

$$\begin{aligned} V_i(\theta) &= \sum_{k\neq i} \tilde{X}_k M_{ki}(Y_i - X_i\theta) + \sum_{l=1}^{n} \tilde{X}_i M_{il}(Y_l - X_l\theta) - \frac{1}{n-2}\sum_{k\neq i}\sum_{l\neq i,k} \tilde{X}_k M_{kl}(Y_l - X_l\theta) \\ &= \sum_{k\neq i} \tilde{X}_k M_{ki} Z_i'\gamma + \sum_{k=1}^{n} \tilde{X}_k M_{ki} U_i + \sum_{l\neq i} \tilde{X}_i M_{il} U_l \\ &\quad + \frac{1}{n-2}\sum_{k\neq i} \tilde{X}_k M_{kk}(Y_k - X_k\theta) - \frac{1}{n-2}\sum_{k\neq i}\sum_{l\neq i} \tilde{X}_k M_{kl}(Y_l - X_l\theta) \\ &\equiv T_{1i} + T_{2i} + T_{3i} + T_{4i} - T_{5i}, \end{aligned}$$

where the second equality follows from (5.1) and $\sum_{l=1}^{n} M_{il} Z_l' = 0$.

For T_{5i}, note that

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} T_{5i}^2 &= \frac{1}{n(n-2)^2}\sum_{i=1}^{n}\left(\begin{array}{c} \sum_{k=1}^{n}\sum_{l=1}^{n} \tilde{X}_k M_{kl}(Y_l - X_l\theta) - \sum_{k=1}^{n} \tilde{X}_k M_{ki}(Y_i - X_i\theta) \\ - \sum_{l=1}^{n} \tilde{X}_i M_{il}(Y_l - X_l\theta) + \tilde{X}_i M_{ii}(Y_i - X_i\theta) \end{array}\right)^2 \\ &= \frac{1}{n(n-2)^2}\sum_{i=1}^{n}\left\{\sum_{l=1}^{n} \tilde{X}_l U_l - \sum_{l=1}^{n} \tilde{X}_i M_{il} U_l - \tilde{X}_i P_{ii}(Y_i - X_i\theta)\right\}^2 \xrightarrow{p} 0, \end{aligned}$$

where the second equality follows from $\sum_{l=1}^{n} M_{il} Z_l' = 0$ and $\sum_{k=1}^{n} \tilde{X}_k M_{kl} = \tilde{X}_l$, and the convergence follows from the law of large numbers.

For T_{4i},

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} T_{4i}^2 &= \frac{1}{n(n-2)^2}\sum_{i=1}^{n}\left(\sum_{k\neq i} \tilde{X}_k M_{kk} Z_k'\gamma\right)^2 + \frac{1}{n(n-2)^2}\sum_{i=1}^{n}\left(\sum_{k\neq i} \tilde{X}_k M_{kk} U_k\right)^2 \\ &\quad + \frac{2}{n(n-2)^2}\sum_{i=1}^{n}\left(\sum_{k\neq i} \tilde{X}_k M_{kk} Z_k'\gamma\right)\left(\sum_{k\neq i} \tilde{X}_k M_{kk} U_k\right) \\ &\xrightarrow{p} \Xi_2, \end{aligned}$$

where the first equality follows from the direct calculation, and the convergence follows from the law of large numbers.

By applying similar arguments to the cross terms, we obtain

$$\frac{1}{n}\sum_{i=1}^{n} V_i(\theta)^2 = \frac{1}{n}\sum_{i=1}^{n}(T_{1i}^2 + T_{2i}^2 + T_{3i}^2) + o_p(1).$$

For T_{1i},

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} T_{1i}^2 &= \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{k=1} \tilde{X}_k M_{ki} Z_i'\gamma - \tilde{X}_i M_{ii} Z_i'\gamma\right)^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}(\tilde{X}_i Z_i'\gamma - \tilde{X}_i M_{ii} Z_i'\gamma)^2 \\ &\xrightarrow{p} \Xi_1,\end{aligned}$$

where the second equality follows from $\sum_{k=1}^{n} \tilde{X}_k M_{kl} = \tilde{X}_l$, and the convergence follows from the law of large numbers and the definition of M_{ii}. Similarly for T_{2i} and T_{3i}, the law of large numbers and $\sum_{k=1}^{n} \tilde{X}_k M_{kl} = \tilde{X}_l$ imply

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} T_{2i}^2 &= \frac{1}{n}\sum_{i=1}^{n}(\tilde{X}_i U_i) \xrightarrow{p} \Sigma, \\ \frac{1}{n}\sum_{i=1}^{n} T_{3i}^2 &= \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{l\neq i} \tilde{X}_i M_{il} U_l\right)^2 \xrightarrow{p} \Psi.\end{aligned}$$

Combining these results, we obtain (A.21).

A.8. **Proof of Theorem MR2.** By using the definition of $\hat{\gamma}^{(i)}$, we can decompose

$$\begin{aligned} &= \frac{1}{n}\sum_{i=1}^{n} \tilde{X}_i^2\{(Y_i - X_i\theta)(Y_i - X_i\theta - Z_i'\hat{\gamma}^{(i)})\} \\ &= \frac{1}{n}\sum_{i=1}^{n} \tilde{X}_i^2\{(Z_i'\gamma + U_i)M_{ii}^{-1}\sum_{j=1}^{n} M_{ij}U_j\} \\ &= \frac{1}{n}\sum_{i=1}^{n} \tilde{X}_i^2 U_i^2 + \frac{1}{n}\sum_{i=1}^{n} \tilde{X}_i^2 (Z_i'\gamma) M_{ii}^{-1}\sum_{j=1}^{n} M_{ij}U_j + \frac{1}{n}\sum_{i=1}^{n} \tilde{X}_i^2 M_{ii}^{-1}\sum_{j\neq i} M_{ij}U_iU_j \\ &\equiv B_1 + B_2 + B_3,\end{aligned}$$

where the first equality is the definition of $\hat{\Sigma}$, and the second equality follows from the relation $M_{ii}(Y_i - X_i\theta - Z_i'\hat{\gamma}^{(i)}) = \sum_{j=1}^{n} M_{ij}U_j$. Thus, it is enough for the conclusion to show that

$B_2 = o_p(1)$ and $B_3 = o_p(1)$. Letting $\sigma_i^2 = E[U_i^2|X_i, Z_i]$, the conditional variance of B_2 is

$$\begin{aligned}
& Var(B_2|Z_1,\ldots,Z_n) \\
= \ & \frac{1}{n^2}\sum_{j=1}^{n}\sigma_j^2\left(\sum_{i=1}^{n}\sum_{k=1}^{n}M_{ij}M_{kj}\tilde{X}_i^2\tilde{X}_k^2M_{ii}^{-1}M_{kk}^{-1}(Z_i'\gamma)(Z_k'\gamma)\right) \\
\leq \ & \frac{1}{n^2}\max_{1\leq j\leq n}\sigma_j^2\left\{\sum_{i=1}^{n}M_{ii}^{-1}\tilde{X}_i^4(Z_i'\gamma)^2 - \sum_{i=1}^{n}\sum_{k\neq i}P_{ik}\tilde{X}_i^2\tilde{X}_k^2M_{ii}^{-1}M_{kk}^{-1}(Z_i'\gamma)(Z_j'\gamma)\right\} \\
\leq \ & \frac{1}{n^2}\max_{1\leq j\leq n}\sigma_j^2\sum_{i=1}^{n}M_{ii}^{-1}\tilde{X}_i^4(Z_i'\gamma)^2 \\
\leq \ & \max_{1\leq j\leq n}\sigma_j^2\cdot\max_{1\leq i\leq n}M_{ii}^{-1}\cdot\max_{1\leq i\leq n}(Z_i'\gamma)^2\left(\max_{1\leq i\leq n}\frac{|\tilde{X}_i|}{\sqrt{n}}\right)^2\frac{1}{n}\sum_{i=1}^{n}\tilde{X}_i^2 \\
= \ & o_p(1),
\end{aligned}$$

where the last equality follows from Assumption MR (ii)-(iii). The conditional variance of B_3 is

$$\begin{aligned}
& Var(B_3|Z_1,\ldots,Z_n) \\
= \ & \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sigma_i^2\sigma_j^2\tilde{X}_i^4M_{ii}^{-2}M_{ij}^2 + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sigma_i^2\sigma_j^2\tilde{X}_i^2\tilde{X}_j^2M_{ii}^{-1}M_{jj}^{-1}M_{ij}^2 \\
\leq \ & 2\left(\max_{1\leq i\leq n}\sigma_i^2\right)^2\max_{1\leq i\leq n}M_{ii}^{-1}\left(\max_{1\leq i\leq n}\frac{|\tilde{X}_i|}{\sqrt{n}}\right)^2\frac{1}{n}\sum_{i=1}^{n}\tilde{X}_i^2(1-M_{ii}) \\
= \ & o_p(1),
\end{aligned}$$

where we used $\sum_{j\neq i}^{n} M_{ij}^2 = \sum_{j=1}^{n} M_{ij}^2 - M_{ii}^2 = M_{ii}(1-M_{ii})$ in the inequality, and the second equality follows from Assumption MR (ii)-(iii). Since $E[B_2|Z_1,\ldots,Z_n] = E[B_3|Z_1,\ldots,Z_n] = 0$, we obtain the conclusion.

References

[1] Angrist, J. D., Imbens, G. W. and A. B. Krueger (1999) Jackknife instrumental variables estimation, *Journal of Applied Econometrics*, 14, 57-67.

[2] Bekker, P. (1994) Alternative approximations to the distributions of instrumental variable estimators, *Econometrica*, 62, 657-681.

[3] Bertail P. (2006) Empirical likelihood in some semi-parametric models, *Bernoulli*, 12, 299-331.

[4] Bravo, F., Escanciano, J. C. and I. van Keilegom (2018) Two-step semiparametric empirical likelihood inference, forthcoming in *Annals of Statistics*.

[5] Cattaneo, M. D., Crump, R. K. and M. Jansson (2010) Robust data-driven inference for density-weighted average derivatives, *Journal of the American Statistical Association*, 105, 1070-1083.

[6] Cattaneo, M. D., Crump, R. K. and M. Jansson (2013) Generalized jackknife estimators of weighted average derivatives, *Journal of the American Statistical Association*, 108, 1243-1268.

[7] Cattaneo, M. D., Crump, R. K. and M. Jansson (2014) Small bandwidth asymptotics for density-weighted average derivatives, *Econometric Theory*, 30, 176-200.

[8] Cattaneo, M. D. and M. Jansson (2018) Kernel-based semiparametric estimators: small bandwidth asymptotics and bootstrap consistency, *Econometrica*, 86, 955-995.

[9] Cattaneo, M. D., Jansson, M. and W. K. Newey (2018a) Inference in linear regression models with many covariates and heteroskedasticity, *Journal of the American Statistical Association*, 113, 1350-1361.

[10] Cattaneo, M. D., Jansson, M., and W. K. Newey (2018b) Alternative asymptotics and the partially linear model with many regressors, *Econometric Theory*, 34, 277-301.

[11] Chao, J. and N. R. Swanson (2005) Consistent estimation with a large number of weak instruments, *Econometrica*, 73, 1673-1692.

[12] Chao, J., Swanson, N. R., Hausman, J. A., Newey, W. K. and T. Woutersen (2012) Asymptotic distribution of JIVE in a heteroskedastic IV regression with many instruments, *Econometric Theory*, 28, 42-86.

[13] Efron, B. and Stein, C. (1981) The jackknife estimate of variance, *Annals of Statistics*, 9, 586-596.

[14] El Karoui, N., Bean, D., Bickel, P. J., Lim, C. and B. Yu (2013) On robust regression with high-dimensional predictors, *Proceedings of the National Academy of Sciences*, 110, 14557-14562.

[15] Gong, Y., Peng, L. and Y. Qi (2010) Smoothed jackknife empirical likelihood method for ROC curve, *Journal of Multivariate Analysis*, 101, 1520-1531.

[16] Hansen, B. E. (2008) Uniform convergence rates for kernel estimation with dependent data, *Econometric Theory*, 24, 726-748.

[17] Hausman, J. A., Newey. W. K., Woutersen, T., Chao, J. and N. R. Swanson (2012) Instrumental variable estimation with heteroskedasticity and many instruments, *Quantitative Economics*, 3, 211-255.

[18] Hinkley, D. V. (1978) Improving the jackknife with special reference to correlation estimation, *Biometrika*, 65, 13-21.

[19] Hjort, N. L., McKeague, I. W. and I. van Keilegom (2009). Extending the scope of empirical likelihood, *Annals of Statistics*, 37, 1079-1111.

[20] Huber, P. J. (1973) Robust regression: asymptotics, conjectures, and Monte Carlo, *Annals of Stastistics*, 1, 799-821.

[21] Kline, P., Saggio, R. and M. Sølvsten (2018) Leave-out estimation of variance components, Working Paper.

[22] Kunitomo, N. (1980) Asymptotic expansions of the distributions of estimators in a linear functional relationship and simultaneous equations, *Journal of the American Statistical Association*, 75, 693-700.

[23] Jing, B. Y., Yuan, J. and W. Zhou (2009) Jackknife empirical likelihood, *Journal of the American Statistical Association*, 104, 1224-1232.

[24] Mammen, E. (1993) Bootstrap and wild bootstrap for high dimensional linear models, *Annals of Statistics*, 21, 255-285.

[25] Newey, W. K. (1994a) The asymptotic variance of semiparametric estimators, *Econometrica*, 62, 1349-1382.

[26] Newey, W. K. (1994b) Kernel estimation of partial means and a general variance, *Econometric Theory*, 10, 233-253.

[27] Owen, A. (1990) Empirical likelihood ratio confidence regions, *Annals of Statistics*, 18, 90-120.

[28] Owen, A. B. (2001) *Empirical Likelihood*, Chapman & Hall/CRC.

[29] Peng, L. Qi, Y. and I. van Keilegom (2012) Jackknife empirical likelihood method for copulas, *Test*, 21, 74-92.

[30] Powell, J. L., Stock, J. H. and T. M. Stoker (1989) Semiparametric estimation of index coefficients, *Econometrica*, 57, 1403-1430.

[31] Rosenbaum, P. and D. Rubin (1983). The central role of the propensity score in observational studies for causal effects, *Biometrika*, 70, 41-55.

[32] Wang, R. Peng, L. and Y. Qi (2013) Jackknife empirical likelihood test for equality of two high dimensional means, *Statistica Sinica*, 23, 667-690.

[33] Zhang, Z. and Y. Zhao (2013) Empirical likelihood for linear transformation models with interval-censored failure time data, *Journal of Multivariate Analysis*, 116, 398-409.

[34] Zheng, S., Jiang, D., Bai, Z. and X. He (2014) Inference on multiple correlation coefficients with moderately high dimensional data, *Biometrika*, 101, 748-754.

[35] Zhong, P.-S. and S. Chen (2014) Jackknife empirical likelihood inference with regression imputation and survey data, *Journal of Multivariate Analysis*, 129, 193-205.

[36] Zhu, L. and L. Xue (2006) Empirical likelihood confidence regions in a partially linear single-index model, *Journal of the Royal Statistical Society*, B, 68, 549-570.

Graduate School of Economics, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186-8601, Japan.

Email address: matsushita.y@r.hit-u.ac.jp

Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, UK.

Email address: t.otsu@lse.ac.uk