JACKKNIFE EMPIRICAL LIKELIHOOD: SMALL BANDWIDTH, SPARSE NETWORK AND HIGH-DIMENSION ASYMPTOTICS

YUKITOSHI MATSUSHITA AND TAISUKE OTSU

Abstract. This paper sheds light on inference problems for statistical models under alternative or nonstandard asymptotic frameworks from the perspective of jackknife empirical likelihood. Examples include small bandwidth asymptotics for semiparametric inference and goodness-of-fit testing, sparse network asymptotics, many covariates asymptotics for regression models, and many-weak instruments asymptotics for instrumental variable regression. We first establish Wilks' theorem for the jackknife empirical likelihood statistic on a general semiparametric inference problem under the conventional asymptotics. We then show that the jackknife empirical likelihood statistic may 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 the jackknife empirical likelihood 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 inference problems for statistical models under alternative or nonstandard asymptotic frameworks from the perspective of jackknife empirical likelihood, initially proposed by Jing, Yuan and Zhou (2009) for one- and two-sample U-statistics. Examples of nonstandard asymptotics include (i) small bandwidth asymptotics for semiparametric inference using average derivatives by Cattaneo, Crump and Jansson (2010, 2014a, b), and for goodness-of-fit testing by a quadratic functional of the density by Bickel and Rosenblatt (1973), (ii) sparse network asymptotics by Bickel, Chen and Levina, (2011), (iii) many-weak instruments asymptotics for instrumental variable regression by Chao et al. (2012), and (iv) many covariates asymptotics for regression models by Cattaneo, Jansson and Newey (2018a, b). These nonstandard asymptotic frameworks, which cover the conventional asymptotics as a special case, are developed to provide better approximations for finite sample properties of statistics and more reliable inference methods. We investigate the behavior of the jackknife empirical likelihood statistics under such nonstandard asymptotics and develop a unified inference approach that has good statistical properties under both the conventional and nonstandard asymptotics. In the main text, we discuss the small bandwidth and sparse network asymptotics, and the results on the many-weak instruments and many covariates asymptotics are presented in Supplementary Material.

In particular, we first consider a general semiparametric inference problem under the conventional asymptotics, and establish Wilks' theorem for the jackknife empirical likelihood 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 jackknife empirical likelihood statistics may lose asymptotic pivotalness under the above nonstandard asymptotic frameworks, and typically converge to quadratic forms of normal vectors with unknown weights. A crucial point, made by Cattaneo, Crump and Jansson (2014b) for the small bandwidth asymptotics, 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) 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 can be of the same order, and Efron and Stein's (1981) bias violates asymptotic pivotalness of the jackknife empirical likelihood statistic. Finally, based on this point, we propose a modification of the jackknife empirical likelihood 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), Efron and Stein (1981), and Cattaneo, Crump and Jansson (2014b) into the estimating equations to construct the jackknife empirical likelihood 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, 2014a, b) advocated the small bandwidth asymptotics to conduct robust statistical inference for semiparametric average derivative estimators. See also Cattaneo and Jansson (2018) for further developments on bootstrap inference. Cattaneo, Crump and Jansson (2014b) is particularly important for this paper since they first pointed out the emergence of Efron and Stein's (1981) bias in the first order. This paper puts forward Cattaneo, Crump and Jansson's (2014b) view toward the jackknife empirical likelihood inference. We also consider goodness-of-fit testing based on a quadratic functional of the density by Bickel and Rosenblatt (1973). In this case, we observe analogous robustness for the bandwidth choices for our jackknife empirical likelihood statistic, cf. Hall (1984). For the network asymptotics, our analysis is considered as robustification of the network method of moments by Bickel, Chen and Levina (2011) and Bhattacharyya and Bickel (2015) for sparse networks. See Supplementary Material for literature on the many-weak instruments and many covariates asymptotics. Cattaneo, Jansson and Ma (2019) employed a jackknife method under nonstandard asymptotics where the first stage of semiparametric generalized method of moments estimation involves many covariates. They used the jackknife to remove the bias term due to many covariates and to estimate their standard error, and then proposed to conduct bootstrap inference. See also Cattaneo, Crump and Jansson (2013) for the jackknife bias correction for weighted average derivatives under weaker bandwidth conditions. This paper focuses on the setups where the bias term is negligible typically because the nonparametric components enter the estimating equations in linear ways, and the asymptotic variance changes under the nonstandard asymptotics.

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), jackknife empirical likelihood has been extended to various contexts, such as Wang, Peng and Qi (2013) for high dimensional means, Gong, Peng and Qi (2010) for receiver operating characteristic 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 Bertail (2006), Zhu and Xue (2006), Hjort, McKeague and Van Keilegom (2009), Bravo, Escanciano and Van Keilegom (2020), among others.

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

$$E[g\{Z,\theta,\mu(X)\}]=0, \tag{1}$$

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 by 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 (1) by setting $g\{Z,\theta,\mu(X)\}=\mu_1(X)-\mu_0(X)-\theta$.

Example 2. Weighted average derivative. Let $m(X)=E(Y|X)$ and w be a known weight function or density function of X. 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 as in Powell, Stock and Stoker (1989), and some nonseparable models. This setup can be considered as a special case of (1) by setting $g\{Z,\theta,\mu(X)\}=w(X)\mu(X)-\theta$ with $\mu(x)=\partial m(x)/\partial x$. For the standard asymptotic analysis in this section, w can be either a known weight function or density function of X. However, the small bandwidth asymptotic analyses are very different for these cases, see Cattaneo, Crump and Jansson (2013) for a known weight case, and Cattaneo, Crump and Jansson (2014a) for the density weighted case. We focus on the density weighted case for the small bandwidth asymptotic analysis in Section 3.

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

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

As shown in Newey (1994), under certain regularity conditions the influence function of $\hat{\theta}$ is given by

$$\psi(Z,X) = -E\left[\frac{\partial g\{Z,\theta,\mu(X)\}}{\partial \theta'}\right]\left(g\{Z,\theta,\mu(X)\} + E\left[\left.\frac{\partial g\{Z,\theta,\mu(X)\}}{\partial \mu'}\right| X\right]\{Y-\mu(X)\}\right), \quad (2)$$

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

2.2. **Jackknife empirical likelihood.** We now introduce the jackknife empirical likelihood approach for the setup in (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. Similar results can be established for local polynomial estimators. For given θ, we construct the jackknife pseudo-values as

$$V_i(\theta) = nS(\theta) - (n-1)S^{(i)}(\theta), \quad (3)$$

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 $\hat{\mu}^{(i)}(X_j) = \frac{1}{\hat{f}(X_j)}\frac{1}{n-2}\sum_{k\neq i,j}\frac{1}{h^d}K\left(\frac{X_j-X_k}{h}\right)Y_k$ 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 jackknife empirical likelihood as

$$\ell(\theta) = -2\sup_{p_1,\ldots,p_n}\sum_{i=1}^{n}\log(np_i), \quad \text{s.t. } p_i \geq 0, \ \sum_{i=1}^{n} p_i = 1, \ \sum_{i=1}^{n} p_i V_i(\theta) = 0.$$

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

$$\ell(\theta) = 2\sup_{\lambda}\sum_{i=1}^{n}\log\{1+\lambda' V_i(\theta)\}. \quad (4)$$

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

Theorem 1. 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 jackknife empirical likelihood statistic $\ell(\theta)$ is asymptotically pivotal and converges to the χ^2_p distribution. Thus, the jackknife empirical likelihood 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), the jackknife empirical likelihood 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 jackknife empirical likelihood is particularly attractive when computation of the estimator $\hat{\theta}$ is expensive. Indeed the jackknife empirical likelihood 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

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

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

$$\hat{\theta} = -\frac{2}{n} \sum_{j=1}^{n} Y_j \frac{\partial \hat{f}(X_j)}{\partial X}, \tag{5}$$

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:

$$\hat{\theta} = \binom{n}{2}^{-1} \sum_{j=1}^{n} \sum_{k=j+1}^{n} U_{jk} = E(\hat{\theta}) + \frac{1}{n} \sum_{j=1}^{n} L_j + \binom{n}{2}^{-1} \sum_{j=1}^{n} \sum_{k=j+1}^{n} W_{jk}, \tag{6}$$

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 (6) as in 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 (2014a) relaxed the requirement $nh^{d+2} \to \infty$, called the small bandwidth asymptotics, so that both the linear and quadratic terms in (6) 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 $\Delta = \lim_{n\to\infty} h^{d+2}E(W_{jk}W_{jk}') = 2E\{\text{var}(Y|X)f(X)\}\int \dot{K}(u)\dot{K}(u)'du$ is the variance of the quadratic term in the Hoeffding decomposition (6). Cattaneo, Crump and Jansson (2014a) 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 jackknife empirical likelihood method to the density weighted average derivative estimator $\hat{\theta}$ in (5). Based on the estimator, we construct the jackknife pseudo-values as in (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 (5). The asymptotic property of the jackknife empirical likelihood statistic in (4) is obtained as follows.

Theorem 2. Consider the setup of this section and suppose Assumption SB in Appendix holds true. Then

$$\ell(\theta) \xrightarrow{d} \begin{cases} \chi_d^2 & \text{under } nh^{d+2}\to\infty, \\ \xi'(\Sigma+4\kappa^{-1}\Delta)\xi & \text{under } nh^{d+2}\to\kappa\in(0,\infty), \\ \frac{1}{2}\chi_d^2 & \text{under } nh^{d+2}\to 0, \end{cases} \tag{7}$$

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

Similar to the estimator $\hat{\theta}$, the limiting distribution of the jackknife empirical likelihood statistic $\ell(\theta)$ depends on the condition on nh^{d+2}. If $nh^{d+2}\to 0$ or ∞, then the jackknife empirical likelihood 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 jackknife empirical likelihood 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 in Efron and Stein (1981). As pointed out by Cattaneo, Crump and Jansson (2014b), 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 jackknife empirical likelihood 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 Cattaneo, Crump and Jansson (2014b) and modify the jackknife empirical likelihood 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. 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 $\bar{Q}=\frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j=i+1}^{n}Q_{ij}$.

Since Q_{ij} is asymptotically expressed as a function of W_{ij}'s but not L_i's, see, eq. (C.13) in Supplementary Material, it can be used to estimate the variance component Δ. We utilize this term to modify the jackknife empirical likelihood statistic as follows

$$\ell^m(\theta)=2\sup_{\lambda}\sum_{i=1}^{n}\log\{1+\lambda' V_i^m(\theta)\}, \tag{8}$$

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

Therefore, the modified jackknife empirical likelihood $\ell^m(\theta)$ is asymptotically pivotal and follows the χ_d^2 limiting distribution for all cases of nh^{d+2}. Note that the modified jackknife empirical likelihood 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. Goodness-of-fit testing

In this section, we consider goodness-of-fit testing for a d-dimensional random vector X with the density function f. In particular, for a specified density function f_0, we wish to test

$$H_0: f=f_0 \quad \text{vs.} \quad H_1: f\neq f_0.$$

Let $\tilde{f}(x)=\frac{1}{nh^d}\sum_{j=1}^{n}K\left(\frac{x-X_j}{h}\right)$ be the kernel density estimator for some kernel function $K:\mathbb{R}^d\to\mathbb{R}$ and bandwidth h. As a test statistic, we consider a modified version of a quadratic functional proposed by Bickel and Rosenblatt (1973):

$$J=\int\{\tilde{f}(x)-K_h*f_0(x)\}^2dx,$$

where $K_h*f_0(x)=\frac{1}{h^d}\int K\left(\frac{x-u}{h}\right)f_0(u)du$. The idea of using the convolution with K_h is first used in Härdle and Mammen (1993). The asymptotic distribution of J is the same for over, optimally, or under-smoothed bandwidths, and it is the same as that of $\int\{\tilde{f}(x)-f_0(x)\}^2dx$ for undersmoothed bandwidths, see, p. 332 of Fan (1994). The main reason of such robustness of J for the bandwidth is due to the fact that the dominant term of J is given by a degenerated U-statistic. Here we show that the jackknife empirical likelihood statistic applied on J enjoys analogous robustness for the bandwidth choices.

Let $\tilde{f}^{(i)}(x) = \frac{1}{(n-1)h^d}\sum_{j\neq i}^n K\left(\frac{x-X_j}{h}\right)$ be the leave-i-out kernel density estimator and define the leave-i-out counterpart of J as

$$J^{(i)} = \int \{\tilde{f}^{(i)}(x) - K_h * f_0(x)\}^2 dx.$$

In this case, we construct the jackknife pseudo-values $V_i = nS - (n-1)S^{(i)}$ by setting

$$S = J - B, \qquad S^{(i)} = J^{(i)} - B,$$

where $B = \frac{1}{nh^d}\int K(z)^2 dz$ is a constant for centering. Then the jackknife empirical likelihood statistic is obtained as

$$\ell_0 = -2 \sup_{\{p_i\}_{i=1}^n} \sum_{i=1}^n \log(np_i), \quad \text{s.t. } p_i \geq 0, \ \sum_{i=1}^n p_i = 1, \ \sum_{i=1}^n p_i V_i = 0.$$

The asymptotic property of the jackknife empirical likelihood statistic ℓ_0 is obtained as follows.

Theorem 4. Consider the setup of this section and suppose Assumption GoF in Appendix holds true. Then under $h \to 0$, $nh^d \to \infty$, and the null hypothesis H_0, it holds

$$\ell_0 \xrightarrow{d} \frac{1}{2}\chi^2(1).$$

Note that the jackknife empirical likelihood statistic ℓ_0 is asymptotically pivotal regardless the over, optimally, or under-smoothed bandwidths. In this example, the limiting distribution is always $\frac{1}{2}\chi^2(1)$, which corresponds to the third case in (7). This is because of the fact that the dominant term of S is given by a degenerated U-statistic. Therefore, in this example, there is no need for modification on the jackknife empirical likelihood statistic as in the previous section.

Since $\sum_{i=1}^n V_i$ converges to a positive constant under the alternative hypothesis, we propose a one-sided version of the signed root jackknife empirical likelihood statistic $S_{EL} = \text{sgn}(\sum_{i=1}^n V_i)\sqrt{2\ell_0}$. Based on the above theorem, we reject H_0 if $S_{EL} > z_{1-\alpha}$, where $z_{1-\alpha}$ is the $(1-\alpha)$-th quantile of the standard normal distribution.

5. Sparse network asymptotics

Consider a random graph on vertices $(1, \ldots, n)$ represented by an $n \times n$ adjacency matrix A, where $A_{kl} = 1$ if there is an edge from node k to l and 0 otherwise. We assume that the graph is undirected and contains no self-loops, which means A is symmetric and diagonals of A are all zero. In this section, we focus on inference for the probability of an edge in the network, $\theta_n = P(A_{kl} = 1)$, which can be estimated by $\hat{\theta} = \binom{n}{2}^{-1}\sum_{k=1}^n\sum_{l=k+1}^n A_{kl}$. In this setup, the parameter θ_n typically depends on n, and note that $d_n = (n-1)\theta_n$ is the expected degree. The case of $d_n = 1$ is called the phase transition, and the case of $d_n \to \infty$ is often considered as a dense graph.

To study the asymptotic properties of $\hat{\theta}$, we employ the nonparametric latent variable model in Bickel, Chen and Levina (2011) and Bhattacharyya and Bickel (2015):

$$P(A_{ij} = 1|\xi_i, \xi_j) = E(A_{ij}|\xi_i, \xi_j) = \theta_n w(\xi_i, \xi_j)\mathbb{I}\{w(\xi_i, \xi_j) \leq \theta_n^{-1}\}, \tag{9}$$

for $i, j \in (1, \ldots, n)$, where $(\xi_1, \ldots, \xi_n)$ are iid $U(0,1)$, and $w(\cdot, \cdot)$ is positive, symmetric, and $\int_0^1 \int_0^1 w(s,t) ds dt = 1$. This model is derived from a general representation theorem of the adjacency matrix A by Bickel and Chen (2009) and is flexible to cover popular network formation models, such as stochastic block models, latent variable models, and preferential attachment models. See Kolaczyk (2009) for a review.

By using the latent variables in (9), the estimation error $\hat{\theta} - \theta_n$ can be decomposed as

$$\hat{\theta} - \theta_n = \frac{1}{n} \sum_{k=1}^{n} L_k + \binom{n}{2}^{-1} \sum_{k=1}^{n} \sum_{l=k+1}^{n} (W_{kl} + R_{kl}), \tag{10}$$

where

$$\begin{aligned} L_k &= 2\{E(A_{kl}|\xi_k) - E(A_{kl})\}, \\ W_{kl} &= E(A_{kl}|\xi_k, \xi_l) - \{E(A_{kl}|\xi_k) - E(A_{kl})\} - \{E(A_{kl}|\xi_l) - E(A_{kl})\} - E(A_{kl}), \\ R_{kl} &= A_{kl} - E(A_{kl}|\xi_k, \xi_l). \end{aligned}$$

The terms by L_k's and W_{kl}'s are analogous to the ones in the Hoeffding decomposition in (6), but the conditioning variables $(\xi_1, \ldots, \xi_n)$ are latent. The third term by R_{kl}'s is composed of projection errors. In Section C.5 in Supplementary Material, we show that

$$\frac{1}{n} \sum_{k=1}^{n} L_k = O_p\left(\frac{d_n}{n\sqrt{n}}\right), \qquad \binom{n}{2}^{-1} \sum_{k=1}^{n} \sum_{l=k+1}^{n} W_{kl} = O_p\left(\frac{d_n}{n^2}\right), \tag{11}$$
$$\binom{n}{2}^{-1} \sum_{k=1}^{n} \sum_{l=k+1}^{n} R_{kl} = O_p\left(\frac{\sqrt{d_n}}{n\sqrt{n}}\right).$$

Thus, as far as $E(A_{ij}|\xi_i)$ does not degenerate to a constant, the limiting distribution of $\hat{\theta}$ is determined by the first linear term in (10) in the dense case with $d_n \to \infty$. On the other hand, in the sparse case with $d_n = O(1)$, the limiting distribution of $\hat{\theta}$ is determined by the first and third terms in (10). Finally, when $E(A_{ij}|\xi_i)$ degenerates to a constant, the third term dominates as far as $d_n = o(n)$. Bhattacharyya and Bickel (2015) proposed a variance estimator that is consistent only in the dense case with non-degenerate $E(A_{ij}|\xi_i)$. Our modified jackknife empirical likelihood inference presented below will be valid for all these cases.

Based on the estimator $\hat{\theta} = \binom{n}{2}^{-1} \sum_{k=1}^{n} \sum_{l=k+1}^{n} A_{kl}$, we construct the jackknife pseudo-values as in (3) with

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

where $\hat{\theta}^{(i)} = \binom{n-1}{2}^{-1} \sum_{k=1, k \neq i}^{n} \sum_{l=k+1, l \neq i}^{n} A_{kl}$ is the leave-i counterpart of $\hat{\theta}$. The limiting distribution of the jackknife empirical likelihood statistic is obtained as follows. Let $\Sigma_n = \mathrm{var}(L_k)/n$ and $\Upsilon_n = \binom{n}{2}^{-1} \mathrm{var}(R_{kl})$.

Theorem 5. Consider the setup of this section under the model (9). Suppose $\int_0^1 \int_0^1 w(s,t)^2 ds dt < \infty$ and $d_n = o(n)$. Then

$$\ell(\theta_n) \xrightarrow{d} \begin{cases} \chi_1^2 & \text{under } d_n \to \infty \text{ and } E(A_{ij}|\xi_i) \text{ is random}, \\ \sigma^{-2} \chi_1^2 & \text{otherwise}, \end{cases}$$

where $\sigma^2 = \lim_{n\to\infty}(\Sigma_n + 2\Upsilon_n)/(\Sigma_n + \Upsilon_n)$.

Similar to the results so far, the limiting distribution of the jackknife empirical likelihood statistic $\ell(\theta_n)$ depends on the behavior of d_n. If the network is dense in the sense that $d_n \to \infty$ and $E(A_{ij}|\xi_i)$ is random, then the jackknife empirical likelihood statistic is asymptotically pivotal. However, for sparse networks with $d_n \nrightarrow \infty$ and possibly degenerate $E(A_{ij}|\xi_i)$, the jackknife empirical likelihood statistic is no longer asymptotically pivotal and its limiting distribution depends on σ^2. It is interesting to note that the discrepancy between $2\Upsilon_n$ and Υ_n in the expression of σ^2 can be understood as the Efron-Stein bias in this context.

It is desirable to modify the jackknife empirical likelihood statistic to have the same χ_1^2 limiting distribution for both cases. Let

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

where $S^{(i,j)}(\theta_n) = \binom{n-2}{2}^{-1} \sum_{k=1,k\neq i,j}^{n} \sum_{l=k+1,l\neq i,j}^{n} A_{kl} - \theta_n$ is the leave-(i,j)-out version of $S(\theta_n)$. Then we define the modified jackknife empirical likelihood statistic $\ell^m(\theta_n)$ as in (8).

Theorem 6. Consider the setup of this section under the model (9). Suppose $\int_0^1 \int_0^1 w(s,t)^2 ds dt < \infty$ and $d_n = o(n)$. Then $\ell^m(\theta_n) \xrightarrow{d} \chi_1^2$ (for both cases).

This theorem shows that the modified jackknife empirical likelihood statistic using the χ^2 critical value is asymptotically valid for both dense and sparse networks as far as $d_n = o(n)$. Note that the currently available inference method by Bhattacharyya and Bickel's (2015) variance estimator is valid only in the dense case with non-degenerate $E(A_{ij}|\xi_i)$.

6. Simulation

This section conducts a simulation study to evaluate the finite sample properties of the jackknife empirical likelihood inference methods. In particular, we focus on the jackknife empirical likelihood inference under the sparse network asymptotics in Section 5, and consider a stochastic block model with $K = 2$ equal-sized communities and the following edge probabilities

$$F_{ab} = P(A_{ij} = 1 | i \in a, j \in b) = s_n S_{ab}, \quad \text{for } 1 \leq a, b \leq K.$$

We set $S = \begin{pmatrix} 0.6 & 0.4 \\ 0.4 & 0.4 \end{pmatrix}$ and vary s_n such that $\theta_n = \pi' F \pi \in (0.5, 0.1, 0.05)$ with $\pi = (0.5, 0.5)'$. The network size is $n = 100$.

We compare four methods to construct confidence intervals for θ_n: (i) Wald-type confidence interval (Wald), which is defined as $[\hat{\theta} \pm 1.96\hat{\sigma}]$ with $\hat{\sigma}^2 = \frac{n-1}{n} \sum_{i=1}^{n} (\hat{\theta}^{(i)} - \hat{\theta})^2$, (ii) bootstrap confidence interval (Boot), which is defined as $[\hat{\theta} - c_{97.5}^* \hat{\sigma},\ \hat{\theta} - c_{2.5}^* \hat{\sigma}]$ with the α-th percentile of the bootstrap approximation c_α^* based on the node resampling network bootstrap by Green and Shalizi (2017) with 999 bootstrap replications, (iii) jackknife empirical likelihood confidence interval (JEL) in Section 5, and (iv) modified jackknife empirical likelihood confidence interval (mJEL) in Section 5.

Table 1 gives the empirical coverage rates and average lengths of the confidence intervals above across 1,000 Monte Carlo replications. The nominal rate is 0.95. The main findings from the simulation study are in line with our theoretical results. The Wald and jackknife empirical likelihood confidence intervals tend to over-cover especially when the network is sparse, which verifies our theoretical results. The bootstrap-based intervals are more accurate than the Wald and jackknife empirical likelihood, but still tend to over-cover for sparse network. The modified jackknife empirical likelihood confidence intervals are most robust to the sparsity of the network compared to the other intervals, and offer close-to-correct empirical coverages in all cases. Furthermore, in terms of the average lengths of the confidence intervals, the modified jackknife empirical likelihood outperforms other methods for all cases.

	Coverage rates				Average interval lengths			
θ_n	Wald	Boot	JEL	mJEL	Wald	Boot	JEL	mJEL
0.5	0.971	0.958	0.972	0.952	0.0583	0.0544	0.0581	0.0514
0.1	0.990	0.974	0.990	0.949	0.0253	0.0225	0.0254	0.0190
0.05	0.995	0.979	0.996	0.949	0.0178	0.0158	0.0179	0.0130

TABLE 1. Coverage rates and average lengths of 95% confidence intervals

We also analyze the power properties of the tests for the null $H_0 : \theta_n = \theta_0$ against the alternative hypotheses $H_1 : \theta_n = \theta_0 + \Delta$ for $\Delta \in (-0.02, -0.01, 0.01, 0.02)$. Table 2 gives the calibrated powers of all the tests across 1,000 Monte Carlo replications, i.e., the rejection frequencies of these tests, where the critical values are given by the Monte Carlo 95th percentiles of these test statistics under H_0. The results suggest that the proposed modified jackknife empirical likelihood test exhibits good calibrated power.

θ_0	Δ	Wald	Boot	JEL	mJEL
0.5	-0.02	0.338	0.288	0.345	0.350
	-0.01	0.131	0.105	0.135	0.142
	0.01	0.900	0.107	0.089	0.085
	0.02	0.258	0.301	0.263	0.253
0.1	-0.02	0.984	0.976	0.981	0.977
	-0.01	0.527	0.455	0.518	0.468
	0.01	0.398	0.456	0.428	0.394
	0.02	0.953	0.958	0.956	0.946
0.05	-0.02	1.000	1.000	1.000	1.000
	-0.01	0.878	0.864	0.871	0.842
	0.01	0.767	0.814	0.798	0.760
	0.02	0.997	0.998	0.998	0.997

TABLE 2. Calibrated powers

7. REAL DATA EXAMPLE

To assess the practical utility of our method, we consider the automobile collision data analyzed by Härdle and Stoker (1989). There are $n = 56$ observations in the data set and the response variable Y indicates whether the accidents are judged to result in fatality, where $Y = 1$ for fatal and $Y = 0$ for not fatal. We focus on three important covariates: X_1 =age of the subject, X_2 =velocity of the automobile, and X_3 =the maximal acceleration. The variables are standardized so that each of them has zero mean and unit variance.

Table 3 presents the density weighted average derivative estimates $\hat{\theta}$ with the standard errors calculated by the Powell, Stock and Stoker's (1989) estimator, and the results for testing significance of each covariate. We employ the data-driven bandwidth selector compatible with the small bandwidth asymptotics proposed by Cattaneo, Crump and Jansson (2010) to implement the modified jackknife empirical likelihood tests. On the other hand we employ the plug-in bandwidth selector proposed by Powell, Stock and Stoker (1989), which is compatible with the standard asymptotics, to implement the point estimators and the Wald tests. The Gaussian kernel is used for all the results.

From Table 3, both the Wald and modified jackknife empirical likelihood (mJEL) methods indicate that X_1 with the estimated slope $\hat{\theta}_1 = .0062$ is statistically significant, and X_3 with the estimated slope $\hat{\theta}_3 = .0016$ is insignificant at the 5% level. On the other hand, for X_2, Wald gives p-value of 0.087 and hence suggests that X_2 with the slope estimate $\hat{\theta}_2 = .0025$ is not statistically distinguishable from zero, while our modified jackknife empirical likelihood gives p-value of 0.049 and hence delivers marginal significance at the 5% level.

	Predictor variables		
$\hat{\theta}$	X_1	X_2	X_3
estimate	.0062	.0025	.0016
s.e.	.0015	.0014	.0015

	Significance tests	
H_0	Wald statistic	mJEL statistic
$\theta_1 = 0$	17.14	11.68
$\theta_2 = 0$	2.93	3.88
$\theta_3 = 0$	1.08	0.31

Table 3. Density weighted average derivative estimates and tests for Collision data

Acknowledgement

The authors would like to thank Matias Cattaneo, Whitney Newey, and seminar participants at Aarhus, Cambridge, LSE, Oxford, Tokyo, York, and European Meeting of the Econometric Society at Manchester for helpful comments. Our research is supported by the JSPS KAKENHI (16KK0074, 18K01541) for Matsushita and ERC Consolidator Grant (SNP 615882) for Otsu.

Appendix A. 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 $\lim_{|x|\to\infty}\{m(x)+|e(x)|\} = 0$. $E(Y^4) < \infty$, $E\{\mathrm{var}(Y|X)f(X)\} > 0$, and $\mathrm{var}\left\{\frac{\partial e(X)}{\partial X} - Y\frac{\partial f(X)}{\partial X}\right\}$ 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 $n^2 h_n^d \to \infty$.

Assumption GoF.

(i): f and its second order derivatives are bounded and uniformly continuous on $\mathbb{R}^d$.

(ii): K is bounded and nonnegative function on $\mathbb{R}^d$ satisfying

$$\int K(u)du = 1, \quad \int uK(u)du = 0, \quad \int u_j u_l K(u)du = 2k\mathbb{I}(j=l) < \infty,$$

for each $j, l = 1, \ldots, d$, where k is a constant that does not depend on j or l.

References

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

[2] Bhattacharyya, S. and P. J. Bickel (2015) Subsampling bootstrap of count features of networks, *Annals of Statistics*, 42, 2384-2411.

[3] Bickel, P. J. and A. Chen (2009) A nonparametric view of network models and Newman-Girvan and other modularities, *Proceedings of the National Academy of Sciences*, 106, 21068-21073.

[4] Bickel, P. J., Chen, A. and E. Levina (2011) The method of moments and degree distributions for network models, *Annals of Statistics*, 39, 2280-2301.

[5] Bickel, P. J. and M. Rosenblatt (1973) On some global measures of the deviations of density function estimates, *Annals of Statistics*, 1, 1071-1095.

[6] Bravo, F., Escanciano, J. C. and I. Van Keilegom (2020) Two-step semiparametric empirical likelihood inference, *Annals of Statistics*, 48, 1-26.

[7] 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.

[8] 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.

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

[10] Cattaneo, M. D., Crump, R. K. and M. Jansson (2014b) Bootstrapping density-weighted average derivatives, *Econometric Theory*, 30, 1135-1164.

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

[12] Cattaneo, M. D., Jansson, M. and X. Ma (2019) Two-step estimation and inference with possibly many included covariates, *Review of Economic Studies*, 86, 1095-1122.

[13] 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.

[14] 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.

[15] 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.

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

[17] Fan, Y. (1994) Testing the goodness-of-fit of a parametric density function by kernel method, *Econometric Theory*, 10, 316-356.

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

[19] Green, A. and C. R. Shalizi (2017) Bootstrapping exchangeable random graphs, Working paper.

[20] Hall, P. (1984) Integrated square error properties of kernel estimators of regression functions, *Annals of Statistics*, 12, 241-260.

[21] Härdle, W. and E. Mammen (1993) Comparing nonparametric versus parametric regression fits, *Annals of Statistics*, 21, 1926-1947.

[22] Härdle, W. and T. M. Stoker (1989) Investigating smooth multiple regression by the method of average derivatives, *Journal of the American Statistical Association*, 84, 986-995.

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

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

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

[26] Kolaczyk, E. D. (2009) *Statistical Analysis of Network Data*, Springer.

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

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

[35] 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

SUPPLEMENTARY MATERIAL FOR "JACKKNIFE EMPIRICAL LIKELIHOOD: SMALL BANDWIDTH, SPARSE NETWORK AND HIGH-DIMENSION ASYMPTOTICS"

YUKITOSHI MATSUSHITA AND TAISUKE OTSU

Abstract. In this file, we provide additional results on the many weak instruments asymptotics (Section A) and many covariates asymptotics (Section B). Section C contains the proofs for all the theorems.

Appendix A. Many-weak instruments asymptotics

A.1. **Instrumental variable regression.** In this section, we consider the instrumental variable regression model

$$\begin{aligned} Y &= X\theta + U, \\ X &= Z'\gamma_n + \epsilon, \end{aligned} \tag{A.1}$$

where Y and X are scalar observables, U and ϵ are scalar error terms, and Z is a K-dimensional vector of instrumental variables. To simplify the presentation, we consider the case where X is scalar, but an extension to the vector case is relatively straightforward.

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.

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 by 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 jackknife instrumental variable estimator is robust to heteroskedasticity and many instruments in contrast to the limited information maximum likelihood and two-stage least squares estimators. Let $\sigma_k^2 = E(U_k^2)$.

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 $C_1^{-1} \le \pi' \left(\frac{1}{n}\sum_{i=1}^n Z_i Z_i'\right) \pi \le C_1$ 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 $\mathrm{var}(U_i, \epsilon_i)$ 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 $n \to \infty$.

Under Assumption MW in the appendix, the limiting distribution of the jackknife instrumental variable estimator is derived as follows by Chao et al. (2012):

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

where

$$\begin{aligned}
H &= \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n (1 - P_{kk})\pi' Z_k Z_k' \pi, \qquad \Sigma = \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^n \sigma_k^2 (1 - P_{kk})^2 \pi' Z_k Z_k' \pi,\\
\Psi &= \lim_{n\to\infty}\frac{1}{K}\sum_{k=1}^n \sum_{l\ne k} P_{kl}^2\{\sigma_k^2 E(\epsilon_l^2) + E(\epsilon_k U_k)E(\epsilon_l U_l)\}.
\end{aligned}$$

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

A.2. **Jackknife empirical likelihood.** In this case, based on the first-order condition of the jackknife instrumental variable estimator, we construct the jackknife pseudo-values as in eq. (3) in the main text with

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

The asymptotic property of the jackknife empirical likelihood statistic in eq. (4) in the main text is obtained as follows. Let $\Xi = \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^n \sum_{l\ne k}\sigma_l^2 P_{lk}^2 \pi' Z_k Z_k' \pi$.

Theorem 7. Consider the setup of this section. Under Assumption MW,

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

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

Similar to Theorem 2, the jackknife empirical likelihood 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 jackknife empirical likelihood 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 with elements P_{kl} is not exactly the projection matrix for the leave-i-out counterpart $S^{(i)}(\theta)$.

It is desirable to modify jackknife empirical likelihood 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 jackknife empirical likelihood statistic $\ell^m(\theta)$ as in eq. (8) in the main text.

Theorem 8. Consider the setup of this section. Under Assumption MW, $\ell^m(\theta) \xrightarrow{d} \chi_1^2$ (for all cases).

Similar comments to Theorem 3 apply. The modified jackknife empirical likelihood $\ell^m(\theta)$ follows the χ_1^2 limiting distribution for all cases without estimating the variance components Σ, Ξ, and Ψ.

Appendix B. Many regressors asymptotics

B.1. **Jackknife empirical likelihood.** We consider the regression model

$$Y = X\theta + Z'\gamma_n + U, \tag{B.1}$$

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 $n \to \infty$.

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 7 on many covariates asymptotics are closely related to Cattaneo, Jansson and Newey (2018a, b).

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 eq. (3) in the main text with

$$S(\theta) = \frac{1}{n} \sum_{k=1}^n \sum_{l=1}^n \tilde{X}_k M_{kl}(Y_l - X_l\theta),$$

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

Let

$$\Sigma = \operatorname*{plim}_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \tilde{X}_i^2 U_i^2, \qquad \Psi = \operatorname*{plim}_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{l\neq i} \tilde{X}_i M_{il} U_l\right)^2,$$

$$\Xi_1 = \operatorname*{plim}_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} (\tilde{X}_i P_{ii} Z_i'\gamma)^2, \qquad \Xi_2 = \operatorname*{plim}_{n\to\infty} \frac{1}{n^3}\sum_{i=1}^{n}\left(\sum_{k\neq i} \tilde{X}_k M_{kk} Z_k'\gamma\right)^2.$$

The limiting distribution of the jackknife empirical likelihood statistic is obtained as follows.

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): $P\{\lambda_{\min}(\sum_{i=1}^n Z_i Z_i') > 0\} \to 1$, and

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

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 $\max_{1\leq i\leq n} |\tilde{X}_i|/\sqrt{n} = o_p(1)$.

Theorem 9. 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 2 and 7, the jackknife empirical likelihood 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 with elements P_{kl} is not exactly the projection matrix for the leave-i-out counterpart $S^{(i)}(\theta)$.

It is desirable to modify jackknife empirical likelihood to have same χ_1^2 limiting distribution for all cases. Let $\hat{\gamma}^{(i)}$ be the leave-i-out ordinary least squares estimator for γ from the regression of $Y_i - X_i\theta$ on Z_i, and

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

Kline, Saggio and Sølvsten (2018) proposed similar estimators for quadratic forms in the parameters of linear models with many regressors and heteroskedasticity. We define the modified jackknife empirical likelihood statistic as in eq. (8) in the main text with the ordinary least squares estimator $\hat{\theta}$ and $\tilde{\Gamma}\tilde{\Gamma}' = \hat{\Sigma}$.

Theorem 10. Consider the setup of this section. Under Assumption MR, $\ell^m(\theta) \xrightarrow{d} \chi_1^2$ (for both cases).

Similar comments to Theorems 3 and 8 apply. The modified jackknife empirical likelihood $\ell^m(\theta)$ 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 jackknife empirical likelihood allow $\tau \in (0, 1)$. Therefore, we expect that our jackknife empirical likelihood inference works better when $\frac{K}{n} \geq \frac{1}{2}$. In the next section, we examine this point by a simulation study.

B.2. **Simulation.** This section conducts a simulation study to evaluate the finite sample properties of the jackknife empirical likelihood 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):

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

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 $\text{var}(U) = \text{var}(V) = 1$, 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. 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 $K/n \geq 1/2$. The jackknife empirical likelihood confidence intervals are conservative, which verifies our theoretical results. The modified jackknife empirical likelihood confidence intervals are most robust to the dimension compared to the other intervals and they offer close-to-correct empirical coverages in all cases. In Table 2, all the intervals do not provide close-to-correct empirical coverage when $K = 7$, because the semiparametric model clearly exhibits misspecification error when K is small.

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 modified jackknife empirical likelihood 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)

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 C. Proofs

C.1. **Proof of Theorem 1.** 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.

C.1.1. *Lemmas for Theorem 1.* 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

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

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

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

for any function a.

Proof. Let $K_{jk} = K\left(\frac{X_j - X_k}{h}\right)$. For (C.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 (C.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 (C.3), note that

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

Thus the first equality of (C.3) follows. The second equality of (C.3) follows from (C.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
$\psi(Z, X) = -E\left[\frac{\partial g\{Z,\theta,\mu(X)\}}{\partial\theta'}\right]\left(g\{Z, \theta, \mu(X)\} + E\left[\left.\frac{\partial g\{Z,\theta,\mu(X)\}}{\partial\mu'}\right| X\right]\{Y - \mu(X)\}\right)$.

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, 1994), 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

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

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

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

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.

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

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

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

for some $C_1 > 0$. Due to (C.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

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

for some $C_2 > 0$. Due to (C.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 (C.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 (1994, 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 (C.4). Since the first and second terms are zero by Lemma 1 and the third term is $o_p(n^{-1/2})$ by (C.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 by 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 (1994, 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

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

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, (C.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

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

For T_{21}, Lemma 1 yields

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

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

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

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 $p\ge 4$, (C.6), the law of large numbers, and the uniform convergence of $\frac{1}{nh^d}\sum_{j\neq i} g_{1j}(\mu_j)\frac{1}{f_j}K\left(\frac{X_j-x}{h}\right)$ over x as in Hansen (2008, Theorem 10). Similarly, for T_{212}, the assumption $nh^{2d}/(\log n)^2\to\infty$, (C.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 (C.7). Therefore, the conclusion is obtained.

C.2. **Proof of Theorem 2.** 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 1, 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

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

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

For (C.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, (C.8) follows by Cattaneo, Crump and Jansson (2014, Theorem 1).

For (C.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 (C.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

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

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, (C.9) is obtained.

C.3. **Proof of Theorem 3.** 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 1, 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

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

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

A similar argument to (C.9) combined with the consistency of $\hat\theta$ yields $\frac{1}{n}\sum_{i=1}^{n} V_i(\hat\theta)^2 \overset{p}{\to} \Sigma + 4\kappa^{-1}\Delta$. Thus, the consistency $\hat\theta$ implies (C.11). It remains to show (C.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 (C.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 (C.9) yields $\frac{1}{n}\sum_{i=1}^{n} V_i(\hat\theta)^2 \overset{p}{\to} \Sigma + 4\kappa^{-1}\Delta$. Thus, for (C.10), it remains to show that

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

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

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

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

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

C.4. **Proof of Theorem 4.** As in the proof of Theorem 1, we can prove the asymptotic equivalence

$$\ell_0 = \left(h^d \sum_{i=1}^n V_i^2\right)^{-1} \left(h^{d/2} \sum_{i=1}^n V_i\right)^2 + o_p(1).$$

Thus, it is enough to show that

$$\frac{h^{d/2}}{\sqrt{2}\sigma} \sum_{i=1}^n V_i \quad \overset{d}{\to} \quad N(0,1), \tag{C.14}$$

$$\frac{h^d}{2\sigma^2} \sum_{i=1}^n V_i^2 \quad \overset{p}{\to} \quad 2, \tag{C.15}$$

where $\sigma^2 = \left\{\int f^2(x)dx\right\} \int \left\{\int K(u)K(u+v)du\right\}^2 dv$. Noting that $\sum_{i=1}^n V_i = nS$, we have $\frac{nh^{d/2}}{\sqrt{2}\sigma} S \overset{d}{\to} N(0,1)$ by following the proof of Hall (1984, Theorem 1). Thus, (C.14) follows.

For (C.15), note that $S = \frac{1}{n^2 h^{2d}} \left(\sum_{i=1}^n W_{ii} + 2\sum_{i=1}^n \sum_{j=i+1}^n W_{ij}\right)$, where

$$W_{ij} = \int \left[K\left(\frac{x-X_i}{h}\right) - E\left\{K\left(\frac{x-X_i}{h}\right)\right\}\right] \left[K\left(\frac{x-X_j}{h}\right) - E\left\{K\left(\frac{x-X_j}{h}\right)\right\}\right] dx.$$

By using this expression, we have

$$\begin{aligned}
\frac{h^d}{2\sigma^2} \sum_{i=1}^n V_i^2 &= \frac{h^d}{2\sigma^2} \sum_{i=1}^n \{S - (n-1)(S^{(i)} - S)\}^2 = \frac{h^d}{2\sigma^2} \left\{nS^2 + \frac{(n-1)^2}{n} \sum_{i=1}^n \sum_{i'=i+1}^n (S^{(i)} - S^{(i')})^2\right\} \\
&= \frac{(n-1)^2 h^d}{2\sigma^2 n} \sum_{i=1}^n \sum_{i'=i+1}^n (S^{(i)} - S^{(i')})^2 + o_p(1) \\
&= \frac{(n-1)^2 h^d}{2\sigma^2 n} \sum_{i=1}^n \sum_{i'=i+1}^n \left[\frac{1}{n^2 h^{2d}} \left\{(W_{i'i'} - W_{ii}) + 2 \sum_{j=1, j\neq i,i'}^n (W_{i'j} - W_{ij})\right\}\right]^2 + o_p(1) \\
&= \frac{(n-1)^2 h^d}{2\sigma^2} \left\{\frac{4}{n^2 h^{4d}} Var(W_{12})\right\} + o_p(1) \\
&= 2 + o_p(1),
\end{aligned}$$

where the second equality follows from the fact that $\sum_{i=1}^n S^{(i)} = nS$, the third equality follows from the fact that $S = O_p(n^{-1}h^{-d/2})$, the fifth equality follows from the law of large numbers, and the last equality follows from the definition of σ^2. Thus, (C.15) is obtained, and the conclusion follows.

C.5. **Proof of Theorem 5.** We first show the orders in eq. (11) in the main text. For L_k, note that

$$\text{var}\left(\frac{1}{n} \sum_{k=1}^n L_k\right) = \frac{4\theta_n^2}{n} \text{var}[E\{w(\xi_i, \xi_j)|\xi_i\}] = O\left(\frac{d_n^2}{n^3}\right),$$

where the first equality follows from eq. (9) of the main text, and the second equality follows from $\text{var}[E\{w(\xi_i, \xi_j)|\xi_i\}] = O(1)$ and the definition $d_n = (n-1)\theta_n$. Similarly, we have

$$\text{var}\left\{\binom{n}{2}^{-1}\sum_{k=1}^{n}\sum_{l=k+1}^{n} W_{kl}\right\} = \binom{n}{2}^{-1}\text{var}(W_{kl})$$
$$= \binom{n}{2}^{-1}\theta_n^2\text{var}\left(w(\xi_k, \xi_l) - [E\{w(\xi_k, \xi_l)|\xi_k\} - 1] - [E\{w(\xi_k, \xi_l)|\xi_l\} - 1] - 1\right) = O\left(\frac{d_n^2}{n^4}\right),$$

and

$$\text{var}\left\{\binom{n}{2}^{-1}\sum_{k=1}^{n}\sum_{l=k+1}^{n} R_{kl}\right\} = \binom{n}{2}^{-1}\text{var}(R_{kl})$$
$$= \binom{n}{2}^{-1}E\{\text{var}(A_{kl}|\xi_k, \xi_l)\} = \binom{n}{2}^{-1}E[\theta_n w(\xi_k, \xi_l)\{1 - \theta_n w(\xi_i, \xi_j\}] = O\left(\frac{d_n}{n^3}\right).$$

We now prove for the case where the network is dense with $d_n \to \infty$ and $E(A_{ij}|\xi_i)$ is random. As in the proof of Theorem 1, we can prove the asymptotic equivalence

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

Thus, it is enough to show that

$$\frac{1}{\sqrt{\omega_n}n}\sum_{i=1}^{n} V_i(\theta_n) \xrightarrow{d} N(0,1), \tag{C.16}$$
$$\frac{1}{\omega_n n^2}\sum_{i=1}^{n} V_i(\theta_n)^2 \xrightarrow{p} 1, \tag{C.17}$$

where $\omega_n = \text{var}(\hat{\theta})$. Note that $\frac{1}{n}\sum_{i=1}^{n} V_i(\theta_n) = \hat{\theta} - \theta_n$. By eq. (11) in the main text, we get

$$\text{var}\left\{\frac{1}{n}\sum_{i=1}^{n} V_i(\theta)\right\} = \omega_n = \Sigma_n,$$

Thus, (C.16) follows from the central limit theorem for U-statistics under the assumptions.

For (C.17), we first note that

$$\sum_{l=1}^{n} V_i(\theta_n)^2 = \sum_{i=1}^{n}\left\{\hat{\theta} - \theta_n + (n-1)(\hat{\theta} - \hat{\theta}^{(i)})\right\}^2 = n(\hat{\theta} - \theta_n)^2 + (n-1)^2\sum_{i=1}^{n}(\hat{\theta} - \hat{\theta}^{(i)})^2$$
$$= n(\hat{\theta} - \theta_n)^2 + (n-1)^2\frac{1}{n}\sum_{i<i'}(\hat{\theta}^{(i)} - \hat{\theta}^{(i')})^2, \tag{C.18}$$

where the second equality follows from $\sum_{i=1}^{n}(\hat{\theta}-\hat{\theta}^{(i)})=0$, and the third equality follows from a direct calculation. Thus, we have

$$\begin{aligned}
\frac{1}{\omega_n n^2}\sum_{i=1}^{n}V_i(\theta_n)^2 &= \frac{1}{\omega_n}\left\{\frac{1}{n}(\hat{\theta}-\theta_n)^2+\frac{(n-1)^2}{n^3}\sum_{i=1}^{n}\sum_{i'=i+1}^{n}(\hat{\theta}^{(i)}-\hat{\theta}^{(i')})^2\right\}\\
&= \frac{(n-1)^2}{\omega_n n^3}\sum_{i=1}^{n}\sum_{i'=i+1}^{n}(\hat{\theta}^{(i)}-\hat{\theta}^{(i')})^2+o_p(1)\\
&= \frac{(n-1)^2}{\omega_n n^3}\sum_{i=1}^{n}\sum_{i'=i+1}^{n}\left\{\frac{1}{n-1}(L_{i'}-L_i)\right\}^2+o_p(1)=\frac{n-1}{\omega_n n^2}\mathrm{var}(L_1)+o_p(1)\\
&\xrightarrow{p} 1,
\end{aligned}$$

where the second equality follows from $\frac{1}{\omega_n n}(\hat{\theta}-\theta_n)^2\xrightarrow{p}0$ by the consistency $\hat{\theta}\xrightarrow{p}\theta_n$, and the fourth equality follows from the law of large numbers.

Finally, we consider the case where the network is sparse with $d_n=O(1)$, or $E(A_{ij}|\xi_i)$ degenerates to a constant. For this case, it is enough to show (C.16) and

$$\frac{1}{\omega_n n^2}\sum_{i=1}^{n}V_i(\theta_n)^2\xrightarrow{p}\sigma^2. \tag{C.19}$$

Using the fact that the terms in eq. (10) in the main text are uncorrelated, we get

$$\mathrm{var}\left\{\frac{1}{n}\sum_{i=1}^{n}V_i(\theta)\right\}=\omega_n=\Sigma_n+\Delta_n+\Upsilon_n,$$

where $\Delta_n=\binom{n}{2}^{-2}\sum_{k=1}^{n}\sum_{l=k+1}^{n}\mathrm{var}(W_{kl})=\binom{n}{2}^{-1}\mathrm{var}(W_{kl})$. Thus, (C.16) follows from the central limit theorem for U-statistics under the assumptions.

For (C.19), we have

$$\begin{aligned}
&\frac{1}{\omega_n n^2}\sum_{i=1}^{n}V_i(\theta_n)^2\\
=\;& \frac{1}{\omega_n}\left\{\frac{1}{n}(\hat{\theta}-\theta_n)^2+\frac{(n-1)^2}{n^3}\sum_{i=1}^{n}\sum_{i'=i+1}^{n}(\hat{\theta}^{(i)}-\hat{\theta}^{(i')})^2\right\}=\frac{(n-1)^2}{\omega_n n^3}\sum_{i=1}^{n}\sum_{i'=i+1}^{n}(\hat{\theta}^{(i)}-\hat{\theta}^{(i')})^2+o_p(1)\\
=\;& \frac{(n-1)^2}{\omega_n n^3}\sum_{i=1}^{n}\sum_{i'=i+1}^{n}\left[\frac{1}{n-1}(L_{i'}-L_i)+\binom{n-1}{2}^{-1}\sum_{j=1,j\neq i,i'}^{n}\{(W_{i'j}-W_{ij})+(R_{i'j}-R_{ij})\}\right]^2+o_p(1)\\
=\;& \frac{(n-1)^2}{\omega_n n^2}\left[\frac{\mathrm{var}(L_1)}{n-1}+2\binom{n-1}{2}^{-1}\{\mathrm{var}(W_{12})+\mathrm{var}(R_{12})\}\right]+o_p(1)\\
\xrightarrow{p}\;& \sigma^2,
\end{aligned}$$

where the first equality follows from (C.18), the second equality follows from $\frac{1}{\omega_n n}(\hat{\theta}-\theta_n)^2\xrightarrow{p}0$ (by the consistency $\hat{\theta}\xrightarrow{p}\theta_n$), the fourth equality follows from the law of large numbers, and the convergence follows from eq. (11) in the main text.

C.6. **Proof of Theorem 6.** As in the proof of Theorem 5, we can prove the asymptotic equivalence

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

Thus, it is enough to show

$$\frac{1}{\sqrt{\omega_n}n}\sum_{i=1}^n V_i^m(\theta_n) \ \xrightarrow{d}\ N(0,\sigma^2), \tag{C.20}$$

$$\frac{1}{\omega_n n^2}\sum_{i=1}^n V_i^m(\theta_n)^2 \ \xrightarrow{p}\ \sigma^2, \tag{C.21}$$

where $\omega_n = \mathrm{var}(\hat\theta)$. A similar argument to (C.19) yields

$$\frac{1}{\omega_n n^2}\sum_{i=1}^n V_i(\theta_n)^2 \xrightarrow{p} \sigma^2. \tag{C.22}$$

Thus, the consistency $|\hat\theta - \theta_n| \xrightarrow{p} 0$ implies (C.21).

It remains to show (C.20). Since $\sum_{i=1}^n V_i(\hat\theta) = 0$, we have

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

By (C.16), it holds $\frac{1}{\sqrt{\omega_n}n}\sum_{i=1}^n V_i(\theta_n) \xrightarrow{d} N(0,1)$. Also a similar argument to (C.19) yields $\frac{1}{\omega_n n^2}\sum_{i=1}^n V_i(\hat\theta)^2 \xrightarrow{p} \sigma^2$. Thus, for (C.20), it remains to show that

$$\frac{1}{\omega_n n^2}\sum_{i=1}^n V_i(\hat\theta)^2 - \frac{1}{\omega_n n^2}\sum_{i=1}^n\sum_{j=i+1}^n Q_{ij}^2 \xrightarrow{p} 1 \tag{C.23}$$

By a direct calculation, we have

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

Thus by using eq. (11) in the main text, we obtain (C.23).

C.7. **Proof of Theorem 7.** We only prove for Case (ii). Other cases are shown in similar ways. As in the proof of Theorem 1, 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

$$\frac{n-1}{\mu_n}\sum_{i=1}^n V_i(\theta) \quad \overset{d}{\to} \quad N(0,\Sigma+\alpha\Psi), \tag{C.24}$$

$$\frac{(n-1)^2}{\mu_n^2}\sum_{i=1}^n V_i(\theta)^2 \quad \overset{p}{\to} \quad \Sigma+\Xi+2\alpha\Psi. \tag{C.25}$$

For (C.24), note that

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

Thus, $\frac{n-1}{\mu_n}\sum_{i=1}^n V_i(\theta) = \frac{1}{\mu_n}\sum_{k=1}^n\sum_{l\neq k} X_kP_{kl}U_l$, and (C.24) follows from Chao et al. (2012, Lemma A2).

We now prove (C.25). Observe that

$$\begin{aligned}\frac{(n-1)^2}{\mu_n^2}\sum_{i=1}^n V_i(\theta)^2 &= \frac{(n-1)^2}{\mu_n^2}\sum_{i=1}^n\left(\frac{1}{n-1}\sum_{k=1}^n\sum_{l\neq k} X_kP_{kl}U_l - \frac{1}{n-2}\sum_{k\neq i}\sum_{l\neq i,k} X_kP_{kl}U_l\right)^2\\ &= \frac{1}{\mu_n^2}\sum_{i=1}^n\left(\sum_{k\neq i} X_kP_{ki}U_i + \sum_{l\neq i} X_iP_{il}U_l - \frac{1}{n-2}\sum_{k\neq i}\sum_{l\neq i,k} X_kP_{kl}U_l\right)^2 \end{aligned}\tag{C.26}$$

For the last term in (C.26), we have

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

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 (A.1) and $\sum_{k=1}^n Z_kP_{ki} = Z_i$, the first two terms in (C.26) are written as

$$\begin{aligned}\sum_{k\neq i} X_kP_{ki}U_i &= \gamma_n' Z_i(1-P_{ii})U_i + \sum_{k\neq i}\epsilon_kP_{ki}U_i,\\ \sum_{l\neq i} X_iP_{il}U_l &= \sum_{l\neq i}\gamma_n' Z_iP_{il}U_l + \sum_{l\neq i}\epsilon_iP_{il}U_l.\end{aligned}$$

Combining these results,

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

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

$$
\begin{aligned}
\frac{1}{\mu_n^2}\sum_{i=1}^n\{\gamma_n' Z_i(1-P_{ii})U_i\}^2 &\xrightarrow{p} \Sigma, \qquad \frac{1}{\mu_n^2}\sum_{i=1}^n\sum_{l\neq i}(\gamma_n' Z_i P_{il} U_l)^2 \xrightarrow{p} \Xi,\\
\frac{1}{\mu_n^2}\sum_{i=1}^n\sum_{k\neq i}(\epsilon_k P_{ki} U_i)^2 &\xrightarrow{p} \alpha\lim_{n\to\infty}\frac{1}{K}\sum_{i=1}^n\sum_{k\neq i}E(\epsilon_k^2)P_{ik}^2\sigma_i^2,\\
\frac{1}{\mu_n^2}\sum_{i=1}^n\sum_{k\neq i}\epsilon_k U_k P_{ki}^2\epsilon_i U_i &\xrightarrow{p} \alpha\lim_{n\to\infty}\frac{1}{K}\sum_{i=1}^n\sum_{k\neq i}E(\epsilon_k U_k)P_{ki}^2 E(\epsilon_i U_i),
\end{aligned}
$$

and we obtain (C.25). Therefore, the conclusion follows.

C.8. **Proof of Theorem 8.** Again, we only prove for Case (ii). Other cases are shown in similar ways. As in the proof of Theorem 1, 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

$$
\frac{n-1}{\mu_n}\sum_{i=1}^n V_i^m(\theta) \xrightarrow{d} N(0,\Sigma+\Xi+2\alpha\Psi), \tag{C.27}
$$

$$
\frac{(n-1)^2}{\mu_n^2}\sum_{i=1}^n V_i^m(\theta)^2 \xrightarrow{p} \Sigma+\Xi+2\alpha\Psi. \tag{C.28}
$$

A similar argument to (C.25) combined with the consistency of $\hat\theta$ 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, the consistency $\hat\theta$ implies (C.28). It remains to show (C.27). 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 (C.24), 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 (C.25) 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 (C.10), it remains to show that

$$
\frac{(n-1)^2}{\mu_n^2}\sum_{i=1}^n\sum_{j=i+1}^n Q_{ij}^2 \xrightarrow{p} \Xi+\alpha\Psi. \tag{C.29}
$$

Note that

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

where the first and third equalities follow from Chao et al. (2012, Lemmas A2 and A3), the second equality follows from direct calculation and (A.1). Therefore, we have (C.29) due to the following results in Chao et al. (2012, Lemmas A3):

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

C.9. **Proof of Theorem 9.** 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 1, 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

$$
\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i(\theta) \xrightarrow{d} N(0,\Sigma), \tag{C.30}
$$

$$
\frac{1}{n}\sum_{i=1}^{n} V_i(\theta)^2 \xrightarrow{p} \Sigma + \Psi + \Xi_1 + \Xi_2, \tag{C.31}
$$

For (C.30), 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 (B.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 (C.31). 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 (B.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 (C.31).

C.10. **Proof of Theorem 10.** 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

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

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

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

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 $E(B_2|Z_1,\dots,Z_n) = E(B_3|Z_1,\dots,Z_n) = 0$, 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] Cattaneo, M. D., Crump, R. K. and M. Jansson (2014) Small bandwidth asymptotics for density-weighted average derivatives, *Econometric Theory*, 30, 176-200.

[4] 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.

[5] 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.

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

[7] 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.

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

[9] 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.

[10] Hall, P. (1984) Central limit theorem for integrated square error of multivariate nonparametric density estimators, *Journal of Multivariate Analysis*, 14, 1-16.

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

[12] 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.

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

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

[15] 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.

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

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

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

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

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