# 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 $\theta$ satisfying

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

where $X$ and $Z$ are observables, $g$ is a known function up to $\theta$ and $\mu$, and $\mu$ 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 $\mu$ is available. Then the parameters $\theta$ 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

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

and the asymptotic variance of $\hat{\theta}$ is obtained by $\mathrm{var}\{\psi(Z,X)\}$. To obtain the Wald-type confidence set for $\theta$, we need to estimate the asymptotic variance $\mathrm{var}\{\psi(Z,X)\}$ that involves analytical or often numerical derivatives of $g$ and estimation of the conditional mean $E\left[\frac{\partial g\{Z,\theta,\mu(X)\}}{\partial \mu'}\middle| X\right]$ and average derivatives $E\left[\frac{\partial g\{Z,\theta,\mu(X)\}}{\partial \theta'}\right]$. We provide an alternative inference approach based on the jackknife empirical likelihood, which does not require estimation of nonparametric components in $\mathrm{var}\{\psi(Z,X)\}$ 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 $\theta$, 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 $\theta$, 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 $\theta$.

This theorem says that the jackknife empirical likelihood statistic $\ell(\theta)$ is asymptotically pivotal and converges to the $\chi^2_p$ distribution. Thus, the jackknife empirical likelihood confidence set of $\theta$ 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 $\chi^2_p$ 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 $\theta$ 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 $\theta$ 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 $\infty$, then the jackknife empirical likelihood statistic is asymptotically pivotal but obeys different limiting distributions. In particular, if we use the conventional $\chi_d^2$ 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 $\kappa$. It is interesting to note that discrepancy of the constants multiplied to $\kappa^{-1}\Delta$ in the variance of $\xi$ 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 $\theta$ 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 $\Delta$. 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, $\ell^m(\theta)\xrightarrow{d}\chi_d^2$ regardless of the condition on $nh^{d+2}$.

Therefore, the modified jackknife empirical likelihood $\ell^m(\theta)$ is asymptotically pivotal and follows the $\chi_d^2$ 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 $\Sigma$ and $\Delta$, 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 $\ell_0$ 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 $\ell_0$ 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 $\theta_n$ 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 $\int_0^1\int_0^1 w(s,t)dsdt = 1$. 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

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

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 $\int_0^1\int_0^1 w(s,t)^2 dsdt < \infty$ 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 $\sigma^2$. It is interesting to note that the discrepancy between $2\Upsilon_n$ and $\Upsilon_n$ in the expression of $\sigma^2$ 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 $\chi_1^2$ 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 $\int_0^1\int_0^1 w(s,t)^2 dsdt < \infty$ 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 $\chi^2$ 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 $\theta_n$: (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 $\alpha$-th percentile of the bootstrap approximation $c_\alpha^*$ 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 | | | |
|---|---|---|---|---|---|---|---|---|
| $\theta_n$ | 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.

| $\theta_0$ | $\Delta$ | 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. $\mu$ 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 $\mu$.
- **(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 $\lim_{|x|\to\infty}\{m(x) + |e(x)|\} = 0$. $E(Y^4) < \infty$, $E\{\text{var}(Y|X)f(X)\} > 0$, and var $\left\{\frac{\partial e(X)}{\partial X} - Y\frac{\partial f(X)}{\partial X}\right\}$ is positive definite.

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

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

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

**Assumption 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`