Journal of Multivariate Analysis 101 (2010) 2420–2433

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Combining conditional and unconditional moment restrictions with missing responses

Xiaohui Yuan [a], Tianqing Liu [a,*], Nan Lin [b], Baoxue Zhang [a,*]

[a] Key Laboratory for Applied Statistics of MOE and School of Mathematics and Statistics, Northeast Normal University, Changchun, China

[b] Department of Mathematics, Washington University in St. Louis, St. Louis, USA

ARTICLE INFO

Article history:
Received 24 October 2009
Available online 25 June 2010

AMS subject classifications:
62B04
62H11
62J08

Keywords:
Conditional moment restrictions
Combined empirical likelihood
Missing response
Unconditional moment restrictions
Wilks' theorem

ABSTRACT

Many statistical models, e.g. regression models, can be viewed as conditional moment restrictions when distributional assumptions on the error term are not assumed. For such models, several estimators that achieve the semiparametric efficiency bound have been proposed. However, in many studies, auxiliary information is available as unconditional moment restrictions. Meanwhile, we also consider the presence of missing responses. We propose the combined empirical likelihood (CEL) estimator to incorporate such auxiliary information to improve the estimation efficiency of the conditional moment restriction models. We show that, when assuming responses are strongly ignorable missing at random, the CEL estimator achieves better efficiency than the previous estimators due to utilization of the auxiliary information. Based on the asymptotic property of the CEL estimator, we also develop Wilks' type tests and corresponding confidence regions for the model parameter and the mean response. Since kernel smoothing is used, the CEL method may have difficulty for problems with high dimensional covariates. In such situations, we propose an instrumental variable-based empirical likelihood (IVEL) method to handle this problem. The merit of the CEL and IVEL are further illustrated through simulation studies.

1. Introduction

Many statistical models appear as conditional moment restrictions. For example, a linear regression model without assumptions on the error distribution. There have been comprehensive studies on estimation and hypothesis testing for conditional moment restriction models, such as Chamberlain [1], Newey [7,8], Robinson [12], and Tripathi and Kitamura [14]. Recently, Kitamura et al. [6] proposed an optimal empirical likelihood (EL)-based estimator, called the smoothed empirical likelihood (SEL) estimator, that achieves the semiparametric efficiency bound.

Sometimes, in addition to the conditional moment restrictions, auxiliary information is available as unconditional moment restrictions, particularly often in econometric studies. The following is an example in Imbens and Lancaster [5]. Consider regressing the US household expenditure on food, y, on household income, x, and suppose that $E(y|x) = \mu(x, \theta) = \theta_1 + \theta_2 x$. This can be expressed in the general form of a conditional moment restriction model as $E\{g(x, y, \theta)|x\} = 0$ with probability 1, where $g(x, y, \theta) = y - \mu(x, \theta)$. In this example, the national average household expenditure on food in the US is known as μ_y. Then, we can write this auxiliary information as the following unconditional moment restriction on x,

$$0 = \mu_y - E(y) = \mu_y - E\{E(y|x)\} = E\{\mu_y - \mu(x, \theta)\} = E\{\psi(x, \theta)\}.$$

* Corresponding author.
E-mail addresses: yuanxiaohui888@gmail.com (X. Yuan), tianqingliu@gmail.com (T. Liu), nlin@math.wustl.edu (N. Lin), bxzhang@nenu.edu.cn (B. Zhang).

0047-259X/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2010.06.013

In this paper, we consider utilization of the auxiliary information to obtain a more efficient estimator than the SEL estimator in conditional moment restriction models. Existing relevant research includes Chaudhuri et al. [2] and Qin [10]. However, these methods do not apply to conditional restriction models as they both require to fully specify the conditional distribution of the response given covariates. Another goal of our study is to allow our estimator for the case of missing responses.

Let y be a response variable and x be a s-dimensional vector of covariates. The response y can be continuous, discrete, or mixed. And we assume that x is continuously distributed with density f. Suppose that we have a random sample with missing responses, (x_i, y_i, δ_i), $i = 1, 2, \ldots, n$, where all the x_i's are observed, and the indicator δ_i is 0 if y_i is missing, and 1 otherwise. Like in [3,15,16] and others, we assume δ and y are conditionally independent given x, namely the strongly ignorable missing at random (MAR) proposed by Rosenbaum and Rubin [13]. As a result, $P(\delta = 1|y, x) = P(\delta = 1|x) =: P(x)$, where $P(x)$ is the propensity score and prescribes a pattern of selection bias in the missingness. Let $z = (x', y)'$ and $z_i = (x_i', y_i)'$, for $i = 1, 2, \ldots, n$. Let $g(z, \theta)$ be a known $q \times 1$ vector function, and $\psi(x, \theta)$ be a known $r \times 1$ vector function. Suppose that the conditional moment restriction is given by

$$E\{g(z, \theta_0)|x\} = 0 \quad \text{with probability 1}, \tag{1.1}$$

and the unconditional moment restriction is

$$E\{\psi(x, \theta_0)\} = 0, \tag{1.2}$$

where the true parameter value θ_0 is interior to parameter space Θ and Θ is a compact subset of R^p. We are interested in the inference about θ and also the mean response $\mu_y = E(y)$.

We propose an approach called combined empirical likelihood (CEL) to combine the information implied by conditional moment restrictions (1.1) and unconditional moment restrictions (1.2). By combining the information from conditional moment restrictions (1.1) and unconditional moment restrictions (1.2), we obtain a more efficient estimator of θ than the SEL estimator proposed by Kitamura et al. [6]. Under mild regularity conditions, the CEL estimator is asymptotically normally distributed, hence Wilks' type tests for θ and μ_y can then be constructed straightforwardly.

Our approach is partly motivated by Qin [10], but advances in two major aspects. Firstly, our model assumption is weaker than that considered in [10]. Qin required to specify the conditional distribution of $y|x$, whereas our approach only requires conditional moment specification. Secondly, the theory in [10] assumes that missing values are missing completely at random, whereas we only assume strongly ignorable MAR. Meanwhile, we also consider hypothesis testing for model parameter θ, which was not covered in [10].

The organization of the paper is as follows. In Section 2, we extend the SEL method to estimation in conditional moment restriction models with missing responses. In Section 3, we then further consider the situation where unconditional moment restrictions are available and propose our CEL approach and its asymptotic properties. Wilks' type test and confidence region for the model parameter are then given. In Section 4, we apply the CEL approach to the inference of the mean response. In Section 5, we propose an instrumental variable-based empirical likelihood (IVEL) method which uses information from the conditional moments restrictions (1.1) with missing responses and the unconditional moments restrictions (1.2). Simulation studies are presented in Section 6, and we conclude our paper in Section 7. Proofs and technical conditions are presented in the Appendix.

2. Extension of the SEL method to missing responses

In this section, we consider only conditional moment restriction models without any auxiliary information, and extend the SEL method in [6] to the case of missing responses. Suppose that, without considering missing responses, the conditional moment restriction model is given by (1.1). Then in the case of missing responses, simply by replacing $g(z, \theta)$ with $\tilde{g}(z, \delta, \theta) = \delta g(z, \theta)$, the SEL method can be extended straightforwardly. The main driving fact is that

$$E\{\tilde{g}(z, \delta, \theta_0)|x\} = 0 \quad \text{with probability 1}, \tag{2.1}$$

when we assume MAR. We will next briefly describe our extension to the SEL method. Let

$$w_{ij} = \frac{K((x_i - x_j)/b_n)}{\sum_{j=1}^{n} K((x_i - x_j)/b_n)}, \tag{2.2}$$

where $K(\cdot)$ is a kernel function satisfying condition C3 in the Appendix, and b_n is a sequence of positive numbers which satisfies condition C8. The extended SEL estimator is obtained by maximizing $\sum_{i=1}^n \sum_{j=1}^n w_{ij} \log p_{ij}$ subject to

$$p_{ij} \geq 0, \quad \sum_{j=1}^{n} p_{ij} = 1, \qquad \sum_{j=1}^{n} p_{ij}\tilde{g}(z_j, \delta_j, \theta) = 0, \quad \text{for } i, j = 1, 2, \ldots, n.$$

By the Lagrange multiplier method, we have

$$\hat{p}_{ij} = \frac{w_{ij}}{1 + \lambda_i'\tilde{g}(z_j, \delta_j, \theta)}, \tag{2.3}$$

where, for any given $\theta \in \Theta$, λ_i solves

$$\sum_{j=1}^{n} \frac{w_{ij}\tilde{g}(z_j, \delta_j, \theta)}{1 + \lambda_i'\tilde{g}(z_j, \delta_j, \theta)} = 0, \quad \text{for } i = 1, 2, \dots, n. \tag{2.4}$$

According to (2.3), the SEL at θ is defined as

$$\mathrm{SEL}(\theta) = \sum_{i=1}^{n}\sum_{j=1}^{n} T_{in} w_{ij} \log \hat{p}_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{n} T_{in} w_{ij} \log \left\{ \frac{w_{ij}}{1 + \lambda_i'\tilde{g}(z_j, \delta_j, \theta)} \right\},$$

where T_{in} is a sequence of trimming functions needed for technical reasons. Cheng [3], Wang and Rao [16] and Kitamura et al. [6] used the indicator function

$$T_{in} = I\left\{ \frac{1}{nb_n^s} \sum_{j=1}^{n} K\left(\frac{x_i - x_j}{b_n}\right) \geq b_n^\tau \right\}, \tag{2.5}$$

where the trimming parameter $\tau \in (0, 1)$ and s is given in condition C3 in the Appendix. Then, the SEL estimator of θ_0 is defined as

$$\hat{\theta} = \arg\max_{\theta\in\Theta} \mathrm{SEL}(\theta). \tag{2.6}$$

The asymptotic distribution of $\hat{\theta}$ is given by the following theorem.

Theorem 2.1. *Under conditions* C1–C8 *in the Appendix, as $n \to \infty$, we have*

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Sigma_1),$$

where

$$\Sigma_1 = \{E[P(x)D'(x, \theta_0)V^{-1}(x, \theta_0)D(x, \theta_0)]\}^{-1}$$

with $P(x) = P(\delta = 1|x)$, $D(x, \theta) = E\{\partial g(z, \theta)/\partial\theta|x\}$ and $V(x, \theta) = E\{g(z, \theta)g'(z, \theta)|x\}$.

3. The combined empirical likelihood (CEL) method

In this section, we consider the conditional moment restriction model (1.1) with missing responses when the auxiliary unconditional moment restriction (1.2) is available. Based on the basic fact that the joint likelihood can be decomposed into the product of a conditional likelihood and a marginal likelihood, we propose to estimate θ by maximizing the combined empirical log-likelihood

$$\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} \log p_{ij} + \sum_{k=1}^{n} \log q_k \tag{3.1}$$

subject to the restrictions

$$\begin{aligned} &q_k \geq 0, \quad \sum_{k=1}^{n} q_k = 1, \qquad \sum_{k=1}^{n} q_k\psi(x_k, \theta) = 0, \\ &p_{ij} \geq 0, \quad \sum_{j=1}^{n} p_{ij} = 1, \qquad \sum_{j=1}^{n} p_{ij}\tilde{g}(z_j, \delta_j, \theta) = 0, \quad \text{for } i, j, k = 1, 2, \dots, n, \end{aligned} \tag{3.2}$$

where w_{ij} is defined in (2.2). Intuitively, we can interpret $\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} \log p_{ij}$ as the conditional empirical log-likelihood of y given x and $\sum_{k=1}^{n} \log q_k$ as the marginal empirical log-likelihood of x. The maximization of (3.1) subject to (3.2) can be conveniently solved by Lagrange multipliers. The Lagrangian is

$$\begin{aligned} L(\theta) = &\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} \log p_{ij} + \sum_{k=1}^{n} \log q_k - \sum_{i=1}^{n} \eta_i \left(\sum_{j=1}^{n} p_{ij} - 1\right) - \gamma \left(\sum_{k=1}^{n} q_k - 1\right) \\ &- \sum_{i=1}^{n} \lambda_i' \sum_{j=1}^{n} p_{ij}\tilde{g}(z_j, \delta_j, \theta) - t' \sum_{k=1}^{n} q_k\psi(x_k, \theta), \end{aligned}$$

where $\eta_1, \ldots, \eta_n, \gamma, \{\lambda_i \in R^q : i = 1, 2, \ldots, n\}$ and $\{t \in R^r\}$ are the multipliers for the constraints (3.2). It is easily verified that the solution is

$$\hat{p}_{ij} = \frac{w_{ij}}{1 + \lambda_i' \tilde{g}(z_j, \delta_j, \theta)} \quad \text{and} \quad \hat{q}_k = \frac{1}{n} \frac{1}{1 + t'\psi(x_k, \theta)}, \tag{3.3}$$

where, for each $\theta \in \Theta$, t and λ_i solve

$$\frac{1}{n} \sum_{k=1}^{n} \frac{\psi(x_k, \theta)}{1 + t'\psi(x_k, \theta)} = 0, \tag{3.4}$$

and

$$\sum_{j=1}^{n} \frac{w_{ij} \tilde{g}(z_j, \delta_j, \theta)}{1 + \lambda_i' \tilde{g}(z_j, \delta_j, \theta)} = 0, \quad \text{for } i = 1, 2, \ldots, n. \tag{3.5}$$

Plugging (3.3) into (3.1) and adjusting the objective function by the trimming function (2.5), we have

$$\begin{aligned} \mathrm{CEL}(\theta) &= \sum_{i=1}^{n} \sum_{j=1}^{n} T_{in} w_{ij} \log \hat{p}_{ij} + \sum_{k=1}^{n} \log \hat{q}_k \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} T_{in} w_{ij} \log \left\{ \frac{w_{ij}}{1 + \lambda_i' \tilde{g}(z_j, \delta_j, \theta)} \right\} + \sum_{k=1}^{n} \log \left\{ \frac{1}{n} \frac{1}{1 + t'\psi(x_k, \theta)} \right\}. \end{aligned} \tag{3.6}$$

Then, the maximum CEL estimator of θ_0 is defined as

$$\tilde{\theta} = \arg \max_{\theta \in \Theta} \mathrm{CEL}(\theta). \tag{3.7}$$

The asymptotic distribution of $\tilde{\theta}$ is given by the following theorem.

Theorem 3.1. *Suppose that the conditions* C1–C8 *in the Appendix are satisfied. Let* $S_{11}(\theta) = E\{\psi(x, \theta)\psi'(x, \theta)\}$, $S_{12}(\theta) = E\{\partial\psi(x, \theta)/\partial\theta\}$, *and* $S_{21}(\theta) = S_{12}'(\theta)$. *If* $S_{11}(\theta_0)$ *is a finite positive definite matrix, we have, as* $n \to \infty$,

$$\sqrt{n}(\tilde{\theta} - \theta_0) \xrightarrow{d} N(0, \Sigma_2),$$

where $\Sigma_2 = \{\Sigma_1^{-1} + S_{21}(\theta_0) S_{11}^{-1}(\theta_0) S_{12}(\theta_0)\}^{-1}$ *with* Σ_1 *given in Theorem* 2.1.

For two matrices A and B, we write $A \le B$ if $B - A$ is a nonnegative-definite matrix. From Theorem 3.1, we conclude the following corollary, which shows that the CEL estimator $\tilde{\theta}$ (3.7) is more efficient than the SEL estimator $\hat{\theta}$ (2.6).

Corollary 3.1. *If both* Σ_1 *and* $S_{11}(\theta_0)$ *are positive definite, we have* $\Sigma_2 \le \Sigma_1$, *and equality holds if and only if* $S_{12}(\theta_0) = 0$.

Next, we define a CEL ratio test for θ. Let $\mathrm{CELR}(\theta) = \mathrm{CEL}(\theta) - \mathrm{CEL}(\tilde{\theta})$. Then the null asymptotic chi-square distribution of $-2\mathrm{CELR}(\theta_0)$ is given in the following theorem.

Theorem 3.2. *Suppose that conditions* C1–C8 *in the Appendix are satisfied. If* θ_0 *is the true value of the parameter, we have, as* $n \to \infty$, $-2\mathrm{CELR}(\theta_0) \xrightarrow{d} \chi_p^2$, *where the* p *is the dimension of* θ_0.

Let $\chi_{p,\alpha}^2$ be the $(1 - \alpha)$th quantile of χ_p^2 for $0 < \alpha < 1$. Then, Theorem 3.2 implies that an asymptotically $100(1 - \alpha)\%$ confidence region for θ is given by $I_{\theta,\alpha} = \{\theta | -2\mathrm{CELR}(\theta) \le \chi_{p,\alpha}^2\}$.

4. Inference for the mean response

In this section we consider the CEL-based inference for the mean response $\mu_y = E(y)$. Denote the conditional mean by $\mu(x, \theta) = E(y|x)$. Note that we then have $\mu_y = E[\mu(x, \theta)]$. Let $\phi(x, \theta, \mu_y) = \mu_y - \mu(x, \theta)$. Similarly as in (3.6), we define

$$\mathrm{CEL}^+(\mu_y, \theta) = \sum_{i=1}^{n} \sum_{j=1}^{n} T_{in} w_{ij} \log \left\{ \frac{w_{ij}}{1 + \lambda_i' \tilde{g}(z_j, \delta_j, \theta)} \right\} + \sum_{k=1}^{n} \log \left\{ \frac{1}{n} \frac{1}{1 + t\phi(x_k, \theta, \mu_y)} \right\},$$

where, for each $\theta \in \Theta$, t and λ_i solve

$$\frac{1}{n} \sum_{k=1}^{n} \frac{\phi(x_k, \theta, \mu_y)}{1 + t\phi(x_k, \theta, \mu_y)} = 0, \tag{4.1}$$

and

$$\sum_{j=1}^{n} \frac{w_{ij}\tilde{g}(z_j, \delta_j, \theta)}{1 + \lambda_i'\tilde{g}(z_j, \delta_j, \theta)} = 0, \quad \text{for } i = 1, 2, \ldots, n. \tag{4.2}$$

Let

$$\mathrm{CELR}^{+}(\mu_y) = \sup_{\theta}\{\mathrm{CEL}^{+}(\mu_y, \theta)\} - \sup_{\mu,\theta}\{\mathrm{CEL}^{+}(\mu_y, \theta)\},$$

then we have the following results.

Theorem 4.1. *Suppose that the conditions* C1–C8 *in the Appendix are satisfied. At the true value μ_{y0}, we have $-2\mathrm{CELR}^{+}(\mu_{y0}) \xrightarrow{d} \chi_1^2$, as $n \to \infty$.*

Theorem 4.1 implies that an asymptotically $100(1-\alpha)\%$ confidence interval for μ_y is given by $I_{\mu_y,\alpha} = \{\mu_y | -2\mathrm{CELR}^{+}(\mu_y) \le \chi_{1,\alpha}^2\}$.

5. An instrumental variable-based empirical likelihood (IVEL) method

In this section, we discuss an alternative empirical likelihood estimator based on instrumental variables, which we call the IVEL method. Notice that $\tilde{g}(z, \delta, \theta_0)$ is not correlated with any function of x in (2.1) when we assume strongly ignorable MAR. Therefore, for a matrix of instrumental variables $v(x, \theta_0)$, (2.1) implies the unconditional moment restriction

$$E\{\varphi(z, \delta, \theta_0)\} = 0, \tag{5.1}$$

where $\varphi(z, \delta, \theta) = v(x, \theta)\tilde{g}(z, \delta, \theta)$. Based on (1.2) and (5.1), it seems natural to construct an IVEL estimator as follows.

Let $p_i = dF(x_i, y_i, \delta_i)$, $i = 1, 2, \ldots, n$, where F is the joint distribution function of (x, y, δ). Then we can obtain an empirical likelihood estimator by maximizing

$$\sum_{i=1}^{n} \log p_i$$

subject to the constraints

$$\sum_{i=1}^{n} p_i = 1, \qquad \sum_{i=1}^{n} p_i h(z_i, \delta_i, \theta) = 0, \quad p_i \ge 0,\ i = 1, \ldots, n,$$

where $h(z, \delta, \theta) = (\varphi'(z, \delta, \theta), \psi'(x, \theta))'$. Here, to guarantee the existence of the solution, it is necessary to choose $v(x, \theta)$ such that $\dim(\varphi) \ge \dim(\theta)$. By the Lagrange multiplier method, we have

$$\hat{p}_i = \frac{1}{n}\frac{1}{1 + \lambda' h(z_i, \delta_i, \theta)}, \tag{5.2}$$

where, for any given $\theta \in \Theta$, λ solves

$$\frac{1}{n}\sum_{i=1}^{n} \frac{h(z_i, \delta_i, \theta)}{1 + \lambda' h(z_i, \delta_i, \theta)} = 0.$$

According to (5.2), the IVEL at θ is defined as

$$\mathrm{IVEL}(\theta) = \sum_{i=1}^{n} \log(\hat{p}_i) = \sum_{i=1}^{n} \log\left\{\frac{1}{n}\frac{1}{1 + \lambda' h(z_i, \delta_i, \theta)}\right\}.$$

Then, the IVEL estimator of θ_0 is defined as

$$\breve{\theta} = \arg\max_{\theta\in\Theta} \mathrm{IVEL}(\theta). \tag{5.3}$$

Note that, under the strongly ignorable MAR assumption,

$$E\{\varphi(z, \delta, \theta)\psi'(x, \theta)\} = 0.$$

By Theorem 1 of [11], we have

$$\sqrt{n}(\breve{\theta} - \theta_0) \xrightarrow{d} N(0, \breve{\Sigma}),$$

where

$$\breve{\Sigma} = \left\{ E\left(\frac{\partial\varphi}{\partial\theta}\right)' (E\varphi\varphi')^{-1} E\left(\frac{\partial\varphi}{\partial\theta}\right) + E\left(\frac{\partial\psi}{\partial\theta}\right)' (E\psi\psi')^{-1} E\left(\frac{\partial\psi}{\partial\theta}\right) \right\}^{-1}.$$

Let $\ddot{\theta}$ denote the maximum empirical likelihood estimate in [11] based on estimating Eq. (5.1). Similarly, by Theorem 1 of [11], we have

$$\sqrt{n}(\ddot{\theta} - \theta_0) \xrightarrow{d} N(0, \ddot{\Sigma}),$$

where

$$\ddot{\Sigma} = \left\{ E\left(\frac{\partial\varphi}{\partial\theta}\right)' (E\varphi\varphi')^{-1} E\left(\frac{\partial\varphi}{\partial\theta}\right) \right\}^{-1}.$$

As shown in [1], the asymptotic variance of any $n^{1/2}$-consistent regular estimator of θ_0 in (2.1) cannot be smaller than Σ_1, where Σ_1 is defined in Theorem 2.1. Since $\ddot{\theta}$ is a $n^{1/2}$-consistent regular estimator of θ_0 in (2.1), we have $\ddot{\Sigma} \geq \Sigma_1$. Therefore,

$$E\left(\frac{\partial\varphi}{\partial\theta}\right)' (E\varphi\varphi')^{-1} E\left(\frac{\partial\varphi}{\partial\theta}\right) \leq E\{P(x)D'(x, \theta_0)V^{-1}(x, \theta_0)D(x, \theta_0)\}.$$

Furthermore, we have

$$\begin{aligned}
\breve{\Sigma} &= \left\{ E\left(\frac{\partial\varphi}{\partial\theta}\right)' (E\varphi\varphi')^{-1} E\left(\frac{\partial\varphi}{\partial\theta}\right) + E\left(\frac{\partial\psi}{\partial\theta}\right)' (E\psi\psi')^{-1} E\left(\frac{\partial\psi}{\partial\theta}\right) \right\}^{-1} \\
&\geq \left\{ E\{P(x)D'(x, \theta_0)V^{-1}(x, \theta_0)D(x, \theta_0)\} + E\left(\frac{\partial\psi}{\partial\theta}\right)' (E\psi\psi')^{-1} E\left(\frac{\partial\psi}{\partial\theta}\right) \right\}^{-1} \\
&= \Sigma_2.
\end{aligned}$$

An interesting question is to find the instrument υ that yields an IVEL estimator $\breve{\theta}$ as efficient as the CEL estimator. Using the standard generalized method of moments (GMM) theory in [4], one can show that the lower bound Σ_2 is achieved by the "optimal" instrument $\upsilon^*(x, \theta_0) = D'(x, \theta_0)V^{-1}(x, \theta_0)$. But because θ_0 is unknown and the functional form of D and V may not be available, an IVEL estimator using the "optimal" instrument υ^* is infeasible. On the other hand, from Theorem 3.1, we know that the maximum CEL estimator $\tilde{\theta}$ achieves the lower bound Σ_2 and it avoids finding the optimal instrument.

6. Simulation studies

In this section, we conduct simulation studies to compare the CEL estimator (3.7) with the OLS estimator, the SEL estimator (2.6) and the IVEL estimator (5.3). We simulate data from the following three models.

Model 1:

$$y = \theta_1 + \theta_2 x + \epsilon, \quad x \sim N(1, 1), \epsilon \sim N(0, 1), x \perp \epsilon,$$

where the notation '$\perp$' stands for 'independence'.

Model 2:

$$y|x \sim \text{Exp}(\theta_1 + \theta_2 x), \quad x \sim U(0, 2),$$

where $\text{Exp}(\lambda)$ denotes the exponential distribution with mean λ.

Model 3:

$$y|x \sim \text{Exp}(e^{\theta_1 + \theta_2 x}), \quad x \sim N(1, 1).$$

In all three models, the true value of (θ_1, θ_2) is $(1, 1)$. For each model, we consider three different MAR mechanisms.

Case 1: $P_1(\delta = 1|x) = \exp(1.50 + 1.25x)/\{1 + \exp(1.50 + 1.25x)\}$.

Case 2: $P_2(\delta = 1|x) = \exp(0.50 + 0.70x)/\{1 + \exp(0.50 + 0.70x)\}$.

Case 3: $P_3(\delta = 1|x) = \exp(0.15 + 0.25x)/\{1 + \exp(0.15 + 0.25x)\}$.

The average missing rates for Cases 1, 2 and 3 are about 0.10, 0.25, and 0.40, respectively. For each model and each of the above three cases of missingness, we generate 1000 Monte Carlo random samples of size $n = 100$ and 200. For data simulated from Models 1 and 2, we estimate the conditional moment restriction model (1.1) with $g(z, \theta) = y - E(y|x, \theta) = y - \theta_1 - \theta_2 x$, where the auxiliary unconditional moment restriction (1.2) is given by $\psi(x, \theta) = \mu_y - \theta_1 - \theta_2 x$, while for Model 3, we estimate the conditional moment restriction model (1.1) with $g(z, \theta) = y - E(y|x, \theta) = y - e^{\theta_1 + \theta_2 x}$, where the auxiliary unconditional moment restriction (1.2) is given by $\psi(x, \theta) = \mu_y - e^{\theta_1 + \theta_2 x}$. The known constants μ_y for Models 1, 2 and 3 are 2, 2 and $e^{2.5}$, respectively. We consider four estimators of $\theta = (\theta_1, \theta_2)$, the SEL estimator $\hat{\theta}$, the CEL estimator $\tilde{\theta}$,

Table 1
RMSE of the OLS, SEL, IVEL and CEL estimators.

Model	n	Estimator	Missingness					
			Case 1		Case 2		Case 3	
			θ_1	θ_2	θ_1	θ_2	θ_1	θ_2
1	100	OLS	0.1627	0.1141	0.1738	0.1211	0.1875	0.1312
		SEL	0.1669	0.1201	0.1830	0.1302	0.1981	0.1424
		IVEL	0.1386	0.1131	0.1470	0.1205	0.1560	0.1304
		CEL	0.1443	0.1198	0.1567	0.1305	0.1673	0.1419
	200	OLS	0.1123	0.0774	0.1270	0.0853	0.1348	0.0927
		SEL	0.1142	0.0795	0.1305	0.0884	0.1396	0.0970
		IVEL	0.0956	0.0771	0.1044	0.0845	0.1096	0.0926
		CEL	0.0971	0.0793	0.1078	0.0880	0.1141	0.0972
2	100	OLS	0.3536	0.3854	0.4003	0.4230	0.4284	0.4745
		SEL	0.3379	0.3678	0.3850	0.4098	0.4125	0.4532
		IVEL	0.3510	0.3534	0.3924	0.3948	0.4433	0.4480
		CEL	0.3373	0.3373	0.3810	0.3821	0.4231	0.4268
	200	OLS	0.2527	0.2813	0.2858	0.3114	0.3089	0.3473
		SEL	0.2397	0.2657	0.2733	0.2949	0.2939	0.3241
		IVEL	0.2559	0.2587	0.2872	0.2895	0.3093	0.3111
		CEL	0.2384	0.2412	0.2706	0.2728	0.2930	0.2949
3	100	OLS	0.5510	0.2724	0.5760	0.2813	0.5622	0.2915
		SEL	0.2221	0.1483	0.2602	0.1643	0.2774	0.1851
		IVEL	0.2968	0.1711	0.3327	0.1865	0.3660	0.2013
		CEL	0.1492	0.0914	0.1705	0.0988	0.1710	0.1018
	200	OLS	0.5037	0.2346	0.5272	0.2423	0.5533	0.2611
		SEL	0.1700	0.1132	0.1979	0.1228	0.2003	0.1269
		IVEL	0.2017	0.1238	0.2458	0.1364	0.2695	0.1473
		CEL	0.1057	0.0661	0.1210	0.0694	0.1273	0.0750

Table 2
Empirical sizes of the CEL ratio tests (%) (nominal level = 5%).

Model	n	$H_0 : \mu_y = \mu_{y0}$			$H_0 : \theta = \theta_0$		
		Case 1	Case 2	Case 3	Case 1	Case 2	Case 3
1	100	5.2	4.9	4.3	3.3	3.1	3.1
	200	4.6	4.3	4.8	4.9	4.3	4.1
2	100	7.5	8.0	9.2	2.7	3.0	2.6
	200	5.5	4.8	5.7	4.1	3.9	4.4
3	100	4.3	2.9	5.6	5.1	5.6	6.9
	200	3.6	4.0	3.7	3.8	3.9	4.6

the OLS estimator $\bar{\theta} = \arg\min_\theta \sum_{i=1}^n \delta_i g(z_i, \theta)^2$, and the IVEL estimator $\breve{\theta}$. For the IVEL estimator $\breve{\theta}$, we choose $(1, x)'$ as the instrumental variable, which is optimal under Model 1 but not under Models 2 and 3.

Kitamura et al. [6] reported that the performance of SEL is relatively insensitive to the choice of the trimming parameter τ, and therefore set $T_{in} = 1$. Our simulation results (not reported here) show that our CEL approach is also insensitive to τ. Therefore, we set $T_{in} = 1$ for the CEL estimator as well. Moreover, both the SEL and CEL estimators require to choose the bandwidth, b_n. We follow the same cross-validation procedure suggested in [6].

Table 1 shows the root mean squared error (RMSE) of the four estimators under different models and missingness. Under Model 1, the OLS estimator is the maximum likelihood estimate, and it has slightly smaller RMSE than the SEL estimator as expected. However, by utilizing the unconditional moment restriction, the CEL estimator and the IVEL estimator are more efficient than both the OLS and SEL estimator, suggested by the smaller RMSEs. In this case, the IVEL uses the optimal instrument and gives a slightly smaller RMSE than the CEL estimator. Under Model 2, the OLS estimator always produces larger RMSEs than the other three estimators. Compared to the SEL estimator, the CEL estimator has similar RMSEs for θ_1 but consistently smaller RMSEs for θ_2. Under Model 3, the CEL estimator always gives smaller RMSEs for both θ_1 and θ_2 than the SEL estimator. Therefore, the simulation results show that the CEL is more efficient due to utilization of the auxiliary information and agree with our theory that the CEL estimator is more efficient than the SEL estimator. Furthermore, under Models 2 and 3, when the instrument is not optimal, the CEL estimator always gives smaller RMSEs for both θ_1 and θ_2 than the IVEL estimator as expected.

The empirical sizes are given by the proportion of the CEL ratio statistics greater than or equal to the critical value under H_0. The critical value is $\chi^2_{2,0.05} = 5.9915$ for -2CELR, and $\chi^2_{1,0.05} = 3.8415$ for $-2\mathrm{CELR}^+$. In Table 2, tests on the model parameter θ always has empirical sizes below the nominal level 0.05, whereas tests on μ_y can have sizes larger than the nominal level 0.05 when $n = 100$. But when $n = 200$, the empirical size is close to 0.05 for all tests on μ_y, which agrees well with our asymptotic theory.

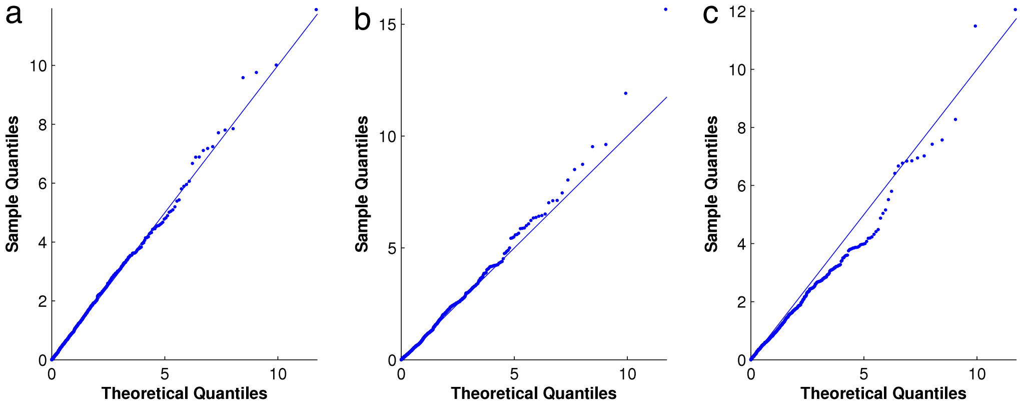

Fig. 1. Q–Q plots of -2CELR^{+} relative to χ_1^2 based on 1000 simulation runs with sample size = 200. (a) Model 1, Case 1, (b) Model 2, Case 1, (c) Model 3, Case 1.

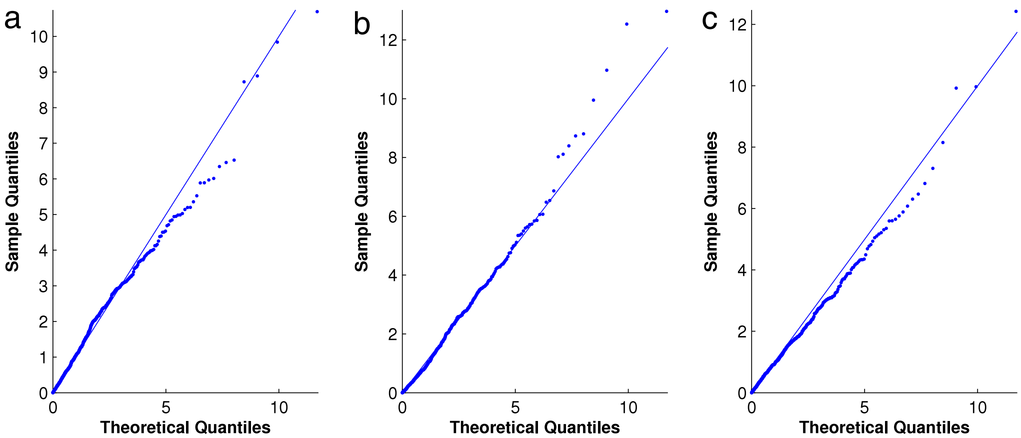

Fig. 2. Q–Q plots of -2CELR^{+} relative to χ_1^2 based on 1000 simulation runs with sample size = 200. (a) Model 1, Case 2, (b) Model 2, Case 2, (c) Model 3, Case 2.

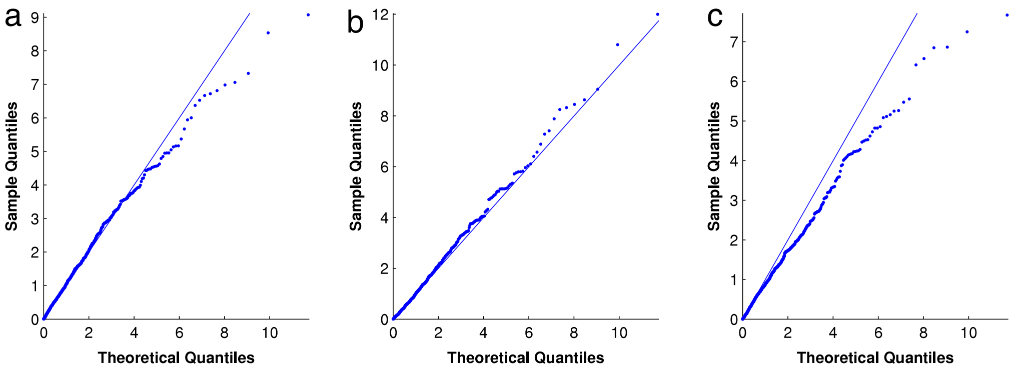

Fig. 3. Q–Q plots of -2CELR^{+} relative to χ_1^2 based on 1000 simulation runs with sample size = 200. (a) Model 1, Case 3, (b) Model 2, Case 3, (c) Model 3, Case 3.

In Figs. 1–6, we give the chi-square Q–Q plots for the CEL ratio statistics, i.e. -2CELR or -2CELR^{+}, to verify their null distributions. The plots clearly show that the test statistics closely follow the asymptotic chi-square distributions.

7. Conclusion

In this paper, we develop a CEL approach to inference for conditional moment restriction models with missing responses and auxiliary unconditional moment restrictions. Under MAR and other mild regularity conditions, the CEL estimator is

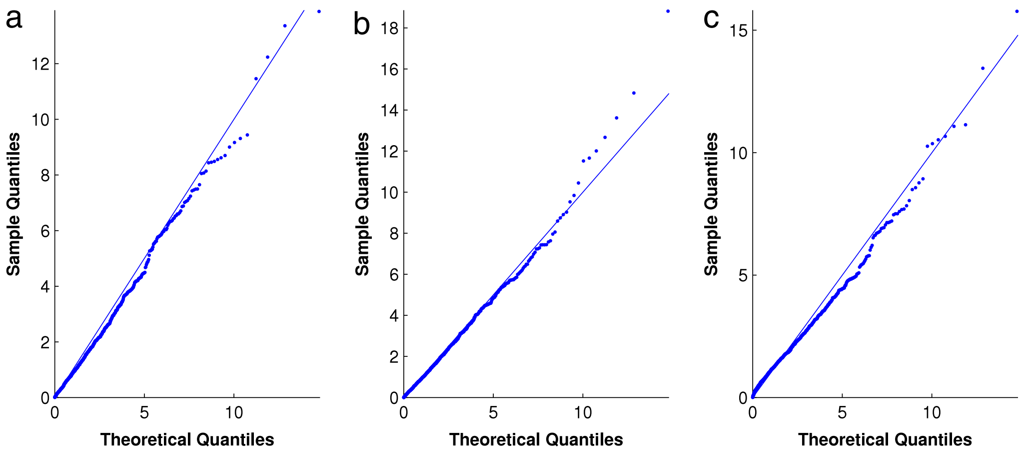

Fig. 4. Q–Q plots of -2CELR relative to χ_2^2 based on 1000 simulation runs with sample size $= 200$. (a) Model 1, Case 1, (b) Model 2, Case 1, (c) Model 3, Case 1.

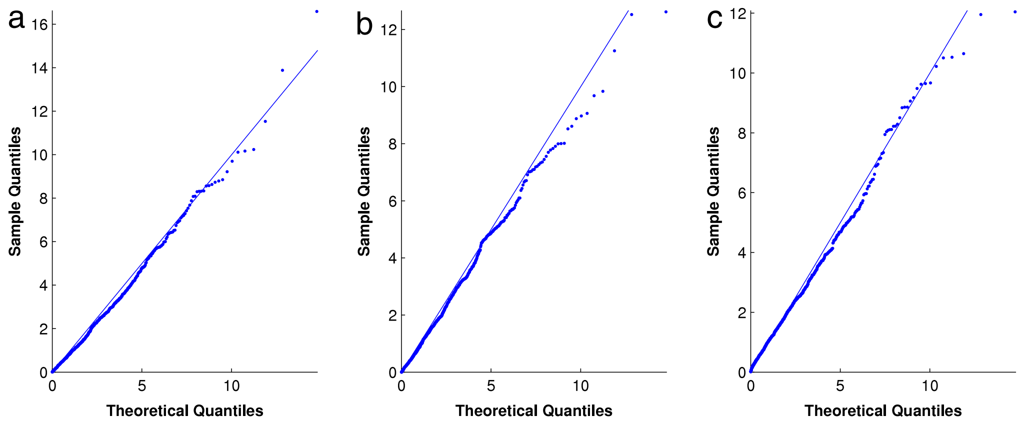

Fig. 5. Q–Q plots of -2CELR relative to χ_2^2 based on 1000 simulation runs with sample size $= 200$. (a) Model 1, Case 2, (b) Model 2, Case 2, (c) Model 3, Case 2.

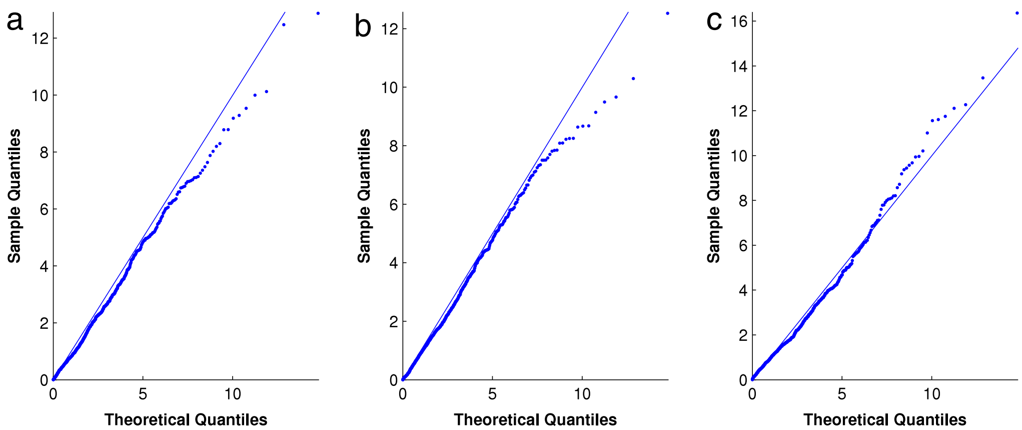

Fig. 6. Q–Q plots of -2CELR relative to χ_2^2 based on 1000 simulation runs with sample size $= 200$. (a) Model 1, Case 3, (b) Model 2, Case 3, (c) Model 3, Case 3.

consistent and asymptotically normal. It is asymptotically more efficient than the SEL estimator due to utilization of the auxiliary unconditional moment restrictions. Wilk's type tests and confidence intervals were also given for the model parameter and mean response. Since kernel smoothing is used, the CEL method may have difficulty for problems with high dimensional covariates. In such situations, we develop an IVEL approach to handle this problem. Simulation studies also show that CEL and IVEL provide more efficient estimates of the model parameter and the CEL-based tests correctly achieve the nominal level.

Acknowledgments

The authors are partly supported by the NSF of China (No. 10871037) and Program for New Century Excellent Talents in University (NCET-09-0284). The authors thank two anonymous referees and the associate editor for their careful reading of the paper and insightful comments.

Appendix

In the following, unless mentioned otherwise, all limits are taken as $n \to \infty$. The following regularity conditions are needed for proving the theorems.

C1: $E\{g(z, \theta_0)|x\} = 0$ with probability 1. For each $\theta \neq \theta_0$, there exists a set $\mathcal{X}_\theta \subseteq R^s$ such that $Pr\{x \in \mathcal{X}_\theta\} > 0$, and $E\{g(z, \theta)|x\} \neq 0$ for every $x \in \mathcal{X}_\theta$.
C2: The probability $P(\delta = 1|x)$ is bounded away from zero, i.e. $\inf_x P(\delta = 1|x) \geq c_0$ for some $c_0 > 0$.
C3: For $x = (x^{(1)}, \ldots, x^{(s)})$, let $K(x) = \prod_{i=1}^{s} \kappa(x^{(i)})$. Here $\kappa : R \to R$ is a continuously differentiable $p.d.f.$ on $[-1, 1]$. Also, κ is symmetric about the origin, and is bounded away from zero on $[-a, a]$ for some $a \in (0, 1)$. It also satisfies $\int |x \cdot \log |x||^{1/2} \kappa(x) dx < \infty$.
C4: (1) $E\{\sup_{\theta\in\Theta} \|g(z, \theta)\|^{m_1}\} < \infty$ for some $m_1 \geq 8$.
(2) $E\{\sup_{\theta\in\Theta} \|\psi(x, \theta)\|^{m_2}\} < \infty$ for some $m_2 \geq 3$.
C5: (1) $f(x)$ is the density function of x, $0 < f(x) \leq \sup_{x\in R^s} f(x) < \infty, f \in C^2(R^s)$, $\sup_{x\in R^s} \|\nabla_x f(x)\| < \infty$, $\sup_{x\in R^s} \|\nabla_{xx} f(x)\| < \infty$.
(2) $E\|x\|^{1+\varrho} < \infty$ for some $\varrho > 0$.
(3) $\theta \mapsto g(z, \theta)$ is continuous on Θ with probability 1, and $E\{\sup_{\theta\in\Theta} \|\partial g(z, \theta)/\partial\theta\|\} < \infty$
(4) $\theta \mapsto \psi(x, \theta)$ is continuous on Θ with probability 1, and $E\{\sup_{\theta\in\Theta} \|\partial\psi(x, \theta)/\partial\theta\|\} < \infty$.
(5) $(\theta, x) \mapsto \|\nabla_{xx}\{E[g^{(l)}(z, \theta)|x]f(x)\}\|$ is uniformly bounded on $\Theta \times R^s$ for $1 \leq l \leq q$.
C6: There exists a closed ball B_0 around θ_0 such that for $1 \leq i, a \leq q$, $1 \leq b \leq r$, and $1 \leq j, k \leq p$,
(1) $\theta \mapsto D(x, \theta), \theta \mapsto V(x, \theta), \theta \mapsto \frac{\partial\psi(x,\theta)}{\partial\theta}$ and $\theta \mapsto S_{11}(\theta)$ are continuous on B_0 with probability 1.
(2) $\inf_{(\xi,x,\theta)\in R^q\times R^s\times B_0} \xi' V(x, \theta)\xi > 0$ and $\sup_{(\xi,x,\theta)\in R^q\times R^s\times B_0} \xi' V(x, \theta)\xi < \infty$, where $\|\xi\| = 1$. $\inf_{(\zeta,\theta)\in R^r\times B_0} \zeta' S_{11}(\theta)\zeta > 0$ and $\sup_{(\zeta,\theta)\in R^r\times B_0} \zeta' S_{11}(\theta)\zeta < \infty$, where $\|\zeta\| = 1$.
(3) With probability 1, $\sup_{\theta\in B_0} \|\partial g^{(i)}(z, \theta)/\partial\theta\| \leq d_1(z)$, $\sup_{\theta\in B_0} \|\partial^2 g^{(i)}(z, \theta)/(\partial\theta^{(j)}\partial\theta^{(k)})\| \leq l_1(z)$, $\sup_{\theta\in B_0} \|\partial\psi^{(b)}(x, \theta)/\partial\theta\| \leq d_2(x)$ and $\sup_{\theta\in B_0} \|\partial^2\psi^{(b)}(x, \theta)/(\partial\theta^{(j)}\partial\theta^{(k)})\| \leq l_2(x)$ hold for some real valued functions $d_1(z)$, $l_1(z)$, $d_2(x)$ and $l_2(x)$ such that $E(d_1^{\eta}(z)) < \infty$ for some $\eta \geq 4$, $E(d_2(x)) < \infty$, $E(l_1(z)) < \infty$ and $E(l_2(x)) < \infty$.
(4) $\sup_{x\in R^s} \|\nabla_x\{D^{(ij)}(x, \theta_0)f(x)\}\| < \infty$ and $\sup_{(x,\theta)\in R^s\times B_0} \|\nabla_{xx}\{D^{(ij)}(x, \theta)f(x)\}\| < \infty$.
(5) $\sup_{x\in R^s} \|\nabla_x\{V^{(ia)}(x, \theta_0)f(x)\}\| < \infty$ and $\sup_{(x,\theta)\in R^s\times B_0} \|\nabla_{xx}\{V^{(ia)}(x, \theta)f(x)\}\| < \infty$.
C7: When solving (3.5) for $\lambda_1, \ldots, \lambda_n$, we only search over the set $\{\gamma \in R^q : \|\gamma\| \leq cn^{-1/m_1}\}$ for some $c > 0$; when solving (3.4) for t, we only search over the set $\{\nu \in R^r : \|\nu\| \leq cn^{-1/m_2}\}$ for some $c > 0$, where m_1 and m_2 are defined in C4.
C8: Let $m = \max(m_1, m_2)$, $\tau \in (0, 1)$, $\varrho \geq \max\{1/\eta + 1/2, 2/m + 1/2\}$, $\eta > 2$, $b_n \to 0$, and $\beta \in (0, 1/2)$ such that $n^{1-2\beta-2/m} b_n^{2s+4\tau} \to \infty$, $n^{\varrho} b_n^{2\tau} \to \infty$, $n^{\varrho-1/\eta} b_n^{\tau} \to \infty$, $n^{\varrho-2/m} b_n^{\tau} \to \infty$, $n^{1-2\beta} b_n^{5s/2+6\tau} \to \infty$, $n^{2\varrho-1/\eta-1/m-1/2} b_n^{2\tau} \to \infty$, and $n^{2\varrho-3/m-1/2} b_n^{3\tau} \to \infty$.

Most of the above conditions were assumed for the SEL estimator in [6]. Additional conditions are on the missing data mechanism and the unconditional moment restrictions. Condition C1 guarantees the identifiability of θ_0. Condition C2 implies that responses cannot be missing with probability 1 anywhere in the domain of the x variable. Condition C3 is a standard assumption for kernel methods. It is also assumed in [6]. Parts (1) and (2) of C4 are needed to prove the consistency of the multiplier t and λ_i. C5 and C6 contain a set of boundedness and moment conditions that are used to show the consistency of the CEL estimator $\tilde{\theta}$. C7 restricts the multipliers λ_i to an n^{-1/m_1}-neighborhood of the origin and t to an n^{-1/m_2}-neighborhood, which is needed to establish the asymptotic normality of $\tilde{\theta}$. C8 collects the conditions on ϱ, η, b_n under which the consistency and asymptotic normality results hold.

Before proving the theorems, we need a few lemmas first.

Lemma A.1. *Suppose that conditions* C1 *and* C4–C6 *hold. Then* $t(\theta_0) = S_{11}^{-1}(\theta_0)\frac{1}{n}\sum_{k=1}^{n} \psi(x_k, \theta_0) + o_p(n^{-1/2})$, *and* $\|t(\theta_0)\| = O_p(n^{-1/2}) = o_p(1)$.

Proof. See the proof of Theorem 1 in [9]. □

Lemma A.2. *Suppose that the conditions* C1–C6 *and* C8 *hold. We have* $\tilde{\theta} \xrightarrow{p} \theta_0$.

Proof. Note that $\tilde{\theta}$ maximizes the objective function

$$G_n(\theta) = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} T_{in}w_{ij}\log\left\{1+\lambda_i'(\theta)\tilde{g}(z_j,\delta_j,\theta)\right\} - \frac{1}{n}\sum_{k=1}^{n}\log\left\{1+t'(\theta)\psi(x_k,\theta)\right\}$$
$$= G_{n1}(\theta) + G_{n2}(\theta),$$

where $\lambda_i(\theta) = \arg\max_\gamma \sum_{j=1}^n w_{ij}\log(1+\gamma'\tilde{g}(z_j,\delta_j,\theta))$ and $t(\theta) = \arg\max_\nu \sum_{k=1}^n \log(1+\nu'\psi(x_k,\theta))$.

By the proof of Theorem 3.1 in [6], for each $\delta > 0$, there exists a strictly positive number $H_1(\delta)$ such that

$$Pr\left\{\sup_{\theta\in\Theta\backslash B(\theta_0,\delta)} G_{n1}(\theta) > -n^{-1/m_1}H_1(\delta)\right\} < \frac{\delta}{4}, \tag{A.1}$$

and

$$Pr\left\{G_{n1}(\theta_0) < -d_n^2 H_1(\delta)\right\} < \frac{\delta}{4}, \tag{A.2}$$

where $n^{1/m_1}d_n^2 \to 0$.

Similarly, we can show that there exists a strictly positive number $H_2(\delta)$ such that

$$Pr\left\{\sup_{\theta\in\Theta\backslash B(\theta_0,\delta)} G_{n2}(\theta) > -n^{-1/m_2}H_2(\delta)\right\} < \frac{\delta}{4}. \tag{A.3}$$

Next, we evaluate $G_{n2}(\theta)$ at the true value θ_0. By Lemma A.1, we have for some $r \in (1/m_2, 1/2)$, $\|t(\theta_0)\| = O_p(n^{-1/2}) = o_p(n^{-r})$. Then, we obtain

$$G_{n2}(\theta_0) \geq -t'(\theta_0)\frac{1}{n}\sum_{k=1}^{n}\psi(x_k,\theta_0) = o_p(n^{-r})O_p(1) = o_p(n^{-r}).$$

Therefore, for all large n, we have

$$Pr\left\{G_{n2}(\theta_0) < -n^{-r}H_2(\delta)\right\} < \frac{\delta}{4}. \tag{A.4}$$

Let $m = \max(m_1, m_2)$ and $H(\delta) = \min\{H_1(\delta), H_2(\delta)\}$. Since $G_n(\theta) = G_{n1}(\theta) + G_{n2}(\theta)$, by (A.1) and (A.3), we have

$$Pr\left\{\sup_{\theta\in\Theta\backslash B(\theta_0,\delta)} G_n(\theta) > -n^{-1/m}H(\delta)\right\}$$
$$\leq Pr\left\{\sup_{\theta\in\Theta\backslash B(\theta_0,\delta)} G_{n1}(\theta) > -n^{-1/m_1}H_1(\delta)\right\} + Pr\left\{\sup_{\theta\in\Theta\backslash B(\theta_0,\delta)} G_{n2}(\theta) > -n^{-1/m_2}H_2(\delta)\right\} \leq \frac{\delta}{2}.$$

On the other hand, by (A.2) and (A.4), we have

$$Pr\left\{G_n(\theta_0) < -a_nH(\delta)\right\} \leq Pr\left\{G_{n1}(\theta_0) < -d_n^2H_1(\delta)\right\} + Pr\left\{G_{n2}(\theta_0) < -n^{-r}H_2(\delta)\right\} \leq \frac{\delta}{2},$$

where $a_n = c(\delta)\cdot\max(d_n^2, n^{-r})$ and $c(\delta) = \{H_1(\delta)+H_2(\delta)\}/H(\delta)$. We can show that $n^{1/m}a_n \to 0$. Thus, for any $\delta > 0$, there exists a positive integer $n_0(\delta)$ such that $Pr\{\tilde{\theta} \in B(\theta_0,\delta)\} \geq 1-\delta$ for all $n > n_0(\delta)$. The proof is then complete. □

Lemma A.3. *Assuming that the conditions* C1 *and* C4–C6 *hold, let*

$$B = \sum_{k=1}^{n}\frac{[\nabla_\theta\psi(x_k,\theta_0)]t(\theta_0)}{1+t'(\theta_0)\psi(x_k,\theta_0)}.$$

Then we have

$$n^{-1/2}B = n^{-1/2}S_{21}(\theta_0)S_{11}^{-1}(\theta_0)\sum_{k=1}^{n}\psi(x_k,\theta_0) + o_p(1).$$

Proof. By Lemma A.1, we have

$$n^{-1/2}B = n^{-1/2}\sum_{k=1}^{n}\frac{[\nabla_\theta\psi(x_k,\theta_0)]t(\theta_0)}{1+t'(\theta_0)\psi(x_k,\theta_0)}$$
$$= n^{1/2}\left(\frac{1}{n}\sum_{k=1}^{n}\frac{\partial\psi(x_k,\theta_0)}{\partial\theta}\right)'\left(S_{11}^{-1}(\theta_0)\frac{1}{n}\sum_{k=1}^{n}\psi(x_k,\theta_0)+o_p(n^{-1/2})\right)$$
$$= n^{-1/2}S_{21}(\theta_0)S_{11}^{-1}(\theta_0)\sum_{k=1}^{n}\psi(x_k,\theta_0)+o_p(1). \quad \square$$

Lemma A.4. *Suppose that the conditions* C1–C8 *hold. Then, we have*

$$\sup_{\theta\in B_0}\|-n^{-1}\nabla_{\theta\theta}\mathrm{CEL}(\theta)-\{E[P(x)D'(x,\theta)V^{-1}(x,\theta)D(x,\theta)]+S_{21}(\theta)S_{11}^{-1}(\theta)S_{12}(\theta)\}\| = o_p(1).$$

Proof. The proof is similar to the proof of Lemma C.1 in [6]. □

Proof of Theorem 2.1. The proof is similar to the proof of Theorem 3.2 in [6]. □

Proof of Theorem 3.1. The first-order condition for (3.7) is $\nabla_\theta\mathrm{CEL}(\tilde\theta)=0$. By the Taylor expansion, we get

$$0 = n^{-1/2}\nabla_\theta\mathrm{CEL}(\theta_0)+\frac{1}{n}\nabla_{\theta\theta}\mathrm{CEL}(\theta^*)n^{1/2}(\tilde\theta-\theta_0),$$

for some θ^* between $\tilde\theta$ and θ_0. Note that

$$-\nabla_\theta\mathrm{CEL}(\theta_0)=\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{T_{in}w_{ij}[\nabla_\theta\tilde g(z_j,\delta_j,\theta_0)]\lambda_i(\theta_0)}{1+\lambda_i'(\theta_0)\tilde g(z_j,\delta_j,\theta_0)}+\sum_{k=1}^{n}\frac{[\nabla_\theta\psi(x_k,\theta_0)]t(\theta_0)}{1+t'(\theta_0)\psi(x_k,\theta_0)}=A+B.$$

By the proof of Theorem 3.2 in [6], we have

$$n^{-1/2}A = n^{-1/2}\sum_{i=1}^{n}\upsilon^*(x_i,\theta_0)\tilde g(z_i,\delta_i,\theta_0)+o_p(1),$$

where $\upsilon^*(x,\theta)=D'(x,\theta)V^{-1}(x,\theta)$. It follows from Lemma A.3 that

$$n^{-1/2}B = n^{-1/2}S_{21}(\theta_0)S_{11}^{-1}(\theta_0)\sum_{k=1}^{n}\psi(x_k,\theta_0)+o_p(1).$$

Furthermore, by Lemmas A.2 and A.4, we have $-n^{-1}\nabla_{\theta\theta}\mathrm{CEL}(\theta^*)\xrightarrow{p}\Sigma_2^{-1}$. Therefore,

$$n^{1/2}(\tilde\theta-\theta_0) = \left\{-n^{-1}\nabla_{\theta\theta}\mathrm{CEL}(\theta^*)\right\}^{-1}n^{-1/2}\nabla_\theta\mathrm{CEL}(\theta_0)$$
$$= -\Sigma_2 n^{-1/2}\sum_{i=1}^{n}\left\{\upsilon^*(x_i,\theta_0)\tilde g(z_i,\delta_i,\theta_0)+S_{21}(\theta_0)S_{11}^{-1}(\theta_0)\psi(x_i,\theta_0)\right\}+o_p(1).$$

Since $-n^{-1/2}\sum_{i=1}^{n}\left\{\upsilon^*(x_i,\theta_0)\tilde g(z_i,\delta_i,\theta_0)+S_{21}(\theta_0)S_{11}^{-1}(\theta_0)\psi(x_i,\theta_0)\right\}\xrightarrow{d}N(0,\Sigma_2^{-1})$, Theorem 3.1 is then proved. □

Proof of Theorem 3.2. Since $\nabla_\theta\mathrm{CEL}(\tilde\theta)=0$, by the Taylor expansion, we get

$$\mathrm{CEL}(\theta_0)=\mathrm{CEL}(\tilde\theta)+\frac{1}{2}(\theta_0-\tilde\theta)'\nabla_{\theta\theta}\mathrm{CEL}(\theta^*)(\theta_0-\tilde\theta)$$

holds for some θ^* between $\tilde\theta$ and θ_0. Then, we can write

$$-2\mathrm{CELR}(\theta_0)=\sqrt{n}(\tilde\theta-\theta_0)'\left\{-n^{-1}\nabla_{\theta\theta}\mathrm{CEL}(\theta^*)\right\}\sqrt{n}(\tilde\theta-\theta_0).$$

Furthermore, from Lemmas A.2, A.4 and Theorem 3.1, we have $-n^{-1}\nabla_{\theta\theta}\mathrm{CEL}(\theta^*)\xrightarrow{p}\Sigma_2^{-1}$ and $\Sigma_2^{-1/2}\sqrt{n}(\tilde\theta-\theta_0)\xrightarrow{d}N(0,I_p)$. Then, Theorem 3.2 can be proved by the continuous mapping theorem. □

Proof of Theorem 4.1. It is easy to see that

$$\sup_{\theta}\{\mathrm{CEL}^{+}(\mu_y,\theta)\}=\sum_{i=1}^{n}\sum_{j=1}^{n}T_{in}w_{ij}\log\left\{\frac{w_{ij}}{1+\lambda_i'(\tilde{\theta})\tilde{g}(z_j,\delta_j,\tilde{\theta})}\right\}+\sum_{k=1}^{n}\log\left\{\frac{1}{n}\frac{1}{1+t(\tilde{\theta},\mu_y)\phi(x_k,\tilde{\theta},\mu_y)}\right\},$$

and

$$\sup_{\mu_y,\theta}\{\mathrm{CEL}^{+}(\mu_y,\theta)\}=\sum_{i=1}^{n}\sum_{j=1}^{n}T_{in}w_{ij}\log\left\{\frac{w_{ij}}{1+\lambda_i'(\hat{\theta})\tilde{g}(z_j,\delta_j,\hat{\theta})}\right\}+\sum_{k=1}^{n}\log\left\{\frac{1}{n}\right\},$$

where $t(\theta,\mu_y)=\arg\max_{\nu}\sum_{k=1}^{n}\log(1+\nu'\phi(x_k,\theta,\mu_y))$ and $\lambda_i(\theta)$ is defined in the proof of Lemma A.2.

Similar to the proof of Lemma A.1, we can show that

$$t(\theta_0,\mu_{y0})=\bar{S}_{11}^{-1}(\theta_0,\mu_{y0})\left[\frac{1}{n}\sum_{i=1}^{n}\phi(x_i,\theta_0,\mu_{y0})\right]+o_p(n^{-1/2}),$$

where $\bar{S}_{11}(\theta,\mu_y)=E\{\phi(x,\theta,\mu_y)\phi'(x,\theta,\mu_y)\}$. By the Taylor expansion, it then follows that

$$\begin{aligned}&\sum_{k=1}^{n}\log\left\{\frac{1}{1+t(\theta_0,\mu_{y0})\phi(x_k,\theta_0,\mu_{y0})}\right\}\\&\quad=-\frac{1}{2}\left[\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\phi(x_k,\theta_0,\mu_{y0})\right]'\bar{S}_{11}^{-1}(\theta_0,\mu_{y0})\left[\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\phi(x_k,\theta_0,\mu_{y0})\right]+o_p(1).\end{aligned}$$

Thus, we get

$$\begin{aligned}\mathrm{CELR}^{+}(\mu_{y0})&=\sup_{\theta}\{\mathrm{CEL}^{+}(\mu_{y0},\theta)\}-\mathrm{CEL}^{+}(\mu_{y0},\theta_0)+\mathrm{CEL}^{+}(\mu_{y0},\theta_0)-\sup_{\mu_y,\theta}\{\mathrm{CEL}^{+}(\mu_y,\theta)\}\\&=\frac{1}{2}\sqrt{n}(\tilde{\theta}-\theta_0)'\left(-\frac{1}{n}\nabla_{\theta\theta}\mathrm{CEL}^{+}(\mu_{y0},\theta^{*})\right)\sqrt{n}(\tilde{\theta}-\theta_0)\\&\quad-\frac{1}{2}\sqrt{n}(\hat{\theta}-\theta_0)'\left(-\frac{1}{n}\nabla_{\theta\theta}\mathrm{SEL}(\theta^{**})\right)\sqrt{n}(\hat{\theta}-\theta_0)\\&\quad-\frac{1}{2}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi(x_i,\theta_0,\mu_{y0})\right]'\bar{S}_{11}^{-1}(\theta_0,\mu_{y0})\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi(x_i,\theta_0,\mu_{y0})\right]+o_p(1)\end{aligned}\tag{A.5}$$

with θ^{*} between $\tilde{\theta}$ and θ_0, and θ^{**} between $\hat{\theta}$ and θ_0. Let

$$\Sigma_3=\{\Sigma_1^{-1}+\bar{S}_{21}(\theta_0,\mu_{y0})\bar{S}_{11}^{-1}(\theta_0,\mu_{y0})\bar{S}_{12}(\theta_0,\mu_{y0})\}^{-1},$$

where $\bar{S}_{12}(\theta,\mu_y)=E\left\{\partial\phi(x,\theta,\mu_y)/\partial\theta\right\}$, and $\bar{S}_{21}(\theta,\mu_y)=\bar{S}_{12}'(\theta,\mu_y)$. Similar to the proof of Lemma A.4, we have

$$-\frac{1}{n}\nabla_{\theta\theta}\mathrm{SEL}(\theta^{**})\xrightarrow{p}\Sigma_1^{-1},\tag{A.6}$$

and

$$-\frac{1}{n}\nabla_{\theta\theta}\mathrm{CEL}^{+}(\mu_{y0},\theta^{*})\xrightarrow{p}\Sigma_3^{-1},\tag{A.7}$$

Furthermore, we can show that

$$n^{1/2}(\tilde{\theta}-\theta_0)=-\Sigma_3\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\upsilon^{*}(x_i,\theta_0)\tilde{g}(z_i,\delta_i,\theta_0)+\bar{S}_{21}(\theta_0,\mu_{y0})\bar{S}_{11}^{-1}(\theta_0,\mu_{y0})\phi(x_i,\theta_0,\mu_{y0})\right\}+o_p(1),\tag{A.8}$$

and

$$n^{1/2}(\hat{\theta}-\theta_0)=-\Sigma_1\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\upsilon^{*}(x_i,\theta_0)\tilde{g}(z_i,\delta_i,\theta_0)+o_p(1).\tag{A.9}$$

Plugging (A.6)–(A.9) into (A.5) and arranging terms, we obtain

$$-2\mathrm{CELR}^{+}(\mu_{y0}) = \begin{pmatrix} \Sigma_1^{1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\upsilon^*(x_i,\theta_0)\tilde{g}(z_i,\delta_i,\theta_0) \\ \bar{S}_{11}^{-1/2}(\theta_0,\mu_{y0})\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi(x_i,\theta_0,\mu_{y0}) \end{pmatrix}' (I_{p+1}-M) \times \begin{pmatrix} \Sigma_1^{1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\upsilon^*(x_i,\theta_0)\tilde{g}(z_i,\delta_i,\theta_0) \\ \bar{S}_{11}^{-1/2}(\theta_0,\mu_{y0})\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi(x_i,\theta_0,\mu_{y0}) \end{pmatrix} + o_p(1),$$

where

$$M = \begin{pmatrix} \Sigma_1^{-1/2} & 0 \\ 0 & \bar{S}_{11}^{1/2}(\theta_0,\mu_{y0}) \end{pmatrix} Q \begin{pmatrix} \Sigma_1^{-1/2} & 0 \\ 0 & \bar{S}_{11}^{1/2}(\theta_0,\mu_{y0}) \end{pmatrix},$$

and

$$Q = \begin{pmatrix} I_p \\ \bar{S}_{11}^{-1}(\theta_0,\mu_{y0})\bar{S}_{12}(\theta_0,\mu_{y0}) \end{pmatrix} \Sigma_3 \begin{pmatrix} I_p \\ \bar{S}_{11}^{-1}(\theta_0,\mu_{y0})\bar{S}_{12}(\theta_0,\mu_{y0}) \end{pmatrix}'.$$

Here, $(I_{p+1}-M)$ is a symmetric idempotent matrix with $rank(I_{p+1}-M) = 1$. Therefore, by the central limit theorem and the continuous mapping theorem, Theorem 4.1 can be easily proved. □

References

[1] G. Chamberlain, Asymptotic effciency in estimation with conditional moment restrictions, J. Econometrics 34 (1987) 305–334.
[2] S. Chaudhuri, M.S. Handcock, M.S. Rendall, Generalized linear models incorporating population level information: an empirical-likelihood-based approach, J. R. Stat. Soc. B 70 (2008) 311–328.
[3] P.E. Cheng, Nonparametric estimation of mean functionals with data missing at random, J. Amer. Statist. Assoc. 89 (1994) 81–87.
[4] L.P. Hansen, Large sample properties of generalized methods of moments estimators, Econometrica 50 (1982) 1029–1054.
[5] G.W. Imbens, T. Lancaster, Combining micro and macro data in microeconometric models, Rev. Econom. Stud. 61 (1994) 655–680.
[6] Y. Kitamura, G. Tripathi, H. Ahn, Empirical likelihood-based inference in conditional moment restriction models, Econometrica 72 (2004) 1667–1714.
[7] W.K. Newey, Efficient instrumental variables estimation of nonlinear models, Econometrica 58 (1990) 809–837.
[8] W.K. Newey, Efficient estimation of models with conditional moment restrictions, in: R. Engle, D. McFadden. (Eds.), Handbook of Statistic Vol. IV, Elsevier, Amsterdam, 1993, pp. 2111–2245.
[9] A. Owen, Empirical likelihood ratio confidence regions, Ann. Statist. 18 (1990) 90–120.
[10] J. Qin, Combining parametric and empirical likelihoods, Biometrika 87 (2000) 484–490.
[11] J. Qin, J. Lawless, Empirical likelihood and general estimating equations, Ann. Statist. 22 (1994) 300–325.
[12] P.M. Robinson, Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form, Econometrica 55 (1987) 875–981.
[13] P.R. Rosenbaum, D.B. Rubin, The central role of the propensity score in observational studies for causal effects, Biometrika 70 (1983) 41–55.
[14] G. Tripathi, Y. Kitamura, Testing conditional moment restrictions, Ann. Statist. 31 (2003) 2059–2095.
[15] D. Wang, S.X. Chen, Empirical likelihood for estimating equations with missing values, Ann. Statist. 37 (2009) 1490–1517.
[16] Q. Wang, J.N.K. Rao, Empirical likelihood-based inference under imputation for missing response data, Ann. Statist. 30 (2002) 896–924.