LIKELIHOOD INFERENCE ON SEMIPARAMETRIC MODELS WITH GENERATED REGRESSORS

YUKITOSHI MATSUSHITA AND TAISUKE OTSU

Abstract. Hahn and Ridder (2013) formulated influence functions of semiparametric three step estimators where generated regressors are computed in the first step. This class of estimators covers several important examples for empirical analysis, such as production function estimators by Olley and Pakes (1996), and propensity score matching estimators for treatment effects by Heckman, Ichimura and Todd (1998). This paper develops a nonparametric likelihood-based inference method for the parameters in such three step estimation problems. By modifying the moment functions to account for influences from the first and second step estimation, the resulting likelihood ratio statistic becomes asymptotically pivotal not only without estimating the asymptotic variance but also without undersmoothing.

1. Introduction

There is a class of econometric problems, where the parameter of interest is estimated by certain three (or more) steps. In the first step, generated regressors (say, $\hat{V}_i$) are computed by some parametric or nonparametric estimation. In the second step, a certain nonparametric regression (say, from Y_i on $(X_i, \hat{V}_i)$) is implemented to obtain an estimator $\hat{\gamma}(X_i, \hat{V}_i)$. In the third step, the parameter of interest β_* is estimated by the sample average or more generally by the method of moments, $n^{-1}\sum_{i=1}^{n} g(\hat{\gamma}(X_i, \hat{V}_i), \hat{\beta}) = 0$. Indeed several important econometric estimators are formulated in this three step manner or interpreted as a special case. Examples include production function estimators by Olley and Pakes (1996), propensity score matching estimators for treatment effects by Heckman, Ichimura and Todd (1998), and various semiparametric estimators that involve generated regressors or control variables.

This three step approach provides an intuitive way to construct a point estimator for the main parameter β_*. On the other hand, the three step formulation complicates inference methods on β_* that properly take into account the sampling variations contained in $\hat{\beta}$. In particular, it is known that for regression models, the estimation errors in generated regressors should be incorporated to compute the standard errors (Pagan, 1984), and it is not trivial to characterize how the estimation errors of the generated regressors $\hat{V}_i$ contribute to the standard error of $\hat{\beta}$. By applying Newey's (1994) path derivative method, Hahn and Ridder (2013) settled this problem and derived the influence function of $\hat{\beta}$. As shown in Hahn and Ridder (2013), the influence function consists of three components: the main term due to the third step, adjustment for the second step estimation of $\hat{\gamma}$, and adjustment due to the first step estimation of $\hat{V}_i$. The third component is the most challenging one and is further decomposed into two terms associated with the two roles of $\hat{V}_i$'s played in the second step nonparametric regression as a conditioning variable and argument.

Based on the influence function by Hahn and Ridder (2013), we can derive the asymptotic variance of $\hat{\beta}$ and compute the standard error to conduct statistical inference on β_*. However, due to the inherently complicated form of the influence function, it is not a trivial task to estimate precisely the asymptotic variance, which typically involves several terms calling for additional nonparametric fits. Therefore, it is practically relevant and theoretically intriguing to ask whether we can circumvent such asymptotic variance estimation to conduct inference on β_*.

For conventional parametric models, the likelihood ratio test can be employed for this purpose. The likelihood ratio statistic is asymptotically pivotal and chi-squared distributed so that we can form the confidence interval by inverting the likelihood ratio. In this paper, we develop a nonparametric likelihood function for the parameter β_* defined in the three step estimation problems by using the method of empirical likelihood (Owen, 1988). In particular, we show that by modifying the moment functions to account for influences from the first and second step estimation, the resulting empirical likelihood ratio statistic becomes asymptotically pivotal and chi-squared distributed. Our statistic allows to conduct inference on β_* without estimating the asymptotic variance. Additionally, in contrast to inference based on the t-ratio, another desirable feature of our empirical likelihood inference is that it does not require undersmoothing for the bandwidths in the first and second step estimation.

When there is no generated regressor (i.e., the target variable, say V_{*i}, is directly observable), estimation of β_* becomes two step, and inference on β_* has been considered in the literature of empirical likelihood. Hjort, McKeague and van Keilegom (2009) studied the properties of the 'plug-in' empirical likelihood using the moment function $g(\hat{\gamma}(X_i, V_{*i}), \beta)$, and showed that the empirical likelihood ratio statistic is not asymptotically pivotal in general. Bravo, Escanciano and van Keilegom (2015) introduced a correction term to the moment function to recover asymptotic pivotalness of the empirical likelihood statistic. In line with this literature, this paper extends the empirical likelihood approach toward three step estimation problems. As shown in Hahn and Ridder (2013), the analysis of three step estimators is significantly different from the two step ones, and semiparametric estimators that involve generated or control variables (e.g., Olley and Pakes' (1996) estimator) cannot be handled by the two step methods.

This paper is organized as follows. In Section 2, we present the basic setup and main results. Sections 2.1 and 2.2 consider the cases of parametric and nonparametric first step, respectively. In Section 3, we provide some extensions of our approach to the cases of additional variables (Section 3.1) and partial means (Section 3.2). In Section 4, our method is illustrated using two examples; a simplified version of Olley and Pakes' (1996) estimator (Section 4.1) and propensity score matching estimators (Section 4.2).

2. Main results

Our notation follows closely that of Hahn and Ridder (2013). Suppose we observe a random sample $\{Y_i, X_i, Z_i\}_{i=1}^n$ of $(Y, X, Z) \in \mathbb{R} \times \mathbb{R}^{d_x} \times \mathbb{R}^{d_z}$. We wish to conduct inference on the scalar parameter β_* satisfying the moment condition

$$E[g(\mu(X, V_*), \beta_*)] = 0, \tag{1}$$

where g is a known function up to $\mu(\cdot,\cdot)$ and β, $\mu(X,V_*) = E[Y|X,V_*]$ is the conditional mean, and V_* is a scalar unobservable regressor expressed as $V_* = \varphi(X,Z)$ by some unknown function φ. When φ is known up to finite dimensional parameters α_*, we denote it by $V_* = \varphi(X,Z,\alpha_*)$. We can estimate β_* in three-steps. First, evaluate the unobservable regressor V_{*i} by its sample counterpart $\hat{V}_i = \varphi(X_i,Z_i,\hat{\alpha})$ based on some parameter estimator $\hat{\alpha}$ of α_* (called a parametric first step) or $\hat{V}_i = \hat{\varphi}(X_i,Z_i)$ based on a nonparametric estimator (called a nonparametric first step). The sample counterpart $\hat{V}_i$ is often called the generated regressor. Second, estimate the conditional mean function $\mu(X_i,V_{*i})$ by nonparametric regression of Y_i on $(X_i,\hat{V}_i)$. We denote the estimated function (evaluated at $(X_i,\hat{V}_i)$) by $\hat{\gamma}(X_i,\hat{V}_i)$.[1] Third, compute the estimator $\hat{\beta}$ for the parameter of interest β_* by solving $n^{-1}\sum_{i=1}^n g(\hat{\gamma}(X_i,\hat{V}_i),\hat{\beta}) = 0$.

Several estimators in econometrics and statistics are formulated in this three-step manner. Examples include semiparametric estimators with generated regressors, and some average treatment effect estimators. See Section 4 below for some specific examples. Hahn and Ridder (2013) derived the influence function for $\hat{\beta}$ by analyzing carefully the effect of the first step estimation. This paper focuses on (nonparametric) likelihood-based inference on β_* without estimating the asymptotic variance of $\hat{\beta}$ and also without undersmoothing the bandwidth to compute $\hat{\gamma}(\cdot,\cdot)$ in the second step (and $\hat{V}_i$ in the nonparametric first step).

2.1. **Case of parametric first step.** We first consider the case where the unobservable regressor $V_* = \varphi(X,Z,\alpha_*)$ is generated from a parametric model indexed by α_*. Suppose we have an estimator $\hat{\alpha}$ for α_*, which admits the asymptotic linear form

$$\sqrt{n}(\hat{\alpha}-\alpha_*) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi(X_i,Z_i,\alpha_*) + o_p(1), \tag{2}$$

and satisfies $n^{-1}\sum_{i=1}^n \psi(X_i,Z_i,\hat{\alpha}) = o_p(n^{-1/2})$. The function ψ is called the influence function for $\hat{\alpha}$. In this case, we evaluate the unobservable regressor V_{*i} by the generated regressor $\hat{V}_i = \varphi(X_i,Z_i,\hat{\alpha})$.

To proceed, we fix the nonparametric estimators for the conditional mean function $\mu(x,v) = E[Y|X=x,V_*=v]$ and its partial derivative $\mu_v(x,v) = \partial\mu(x,v)/\partial v$ with respect to the second argument. To be specific, we hereafter consider the local linear regression from Y_i on $(X_i,\hat{V}_i)$ (i.e., the weighted least square estimation using the kernel weights). We employ the intercept and slope coefficient of $\hat{V}_i$ as estimators for $\mu(x,v)$ and $\mu_v(x,v)$, respectively. Denote these estimators by $\hat{\gamma}(x,v)$ and $\hat{\gamma}_v(x,v)$, respectively.[2]

Let $g_1(\mu,\beta)$ and $g_2(\mu,\beta)$ be the first and second derivatives of $g(\cdot,\cdot)$ with respect to its first argument evaluated at (μ,β), and $\varphi_\alpha(x,z,\alpha)$ be the partial derivative of $\varphi(\cdot,\cdot,\cdot)$ with respect to its third argument evaluated at (x,z,α). Based on the above notation, we propose the following empirical likelihood function

$$\ell(\beta) = -2\sup_{\{p_i\}_{i=1}^n}\left\{\sum_{i=1}^n \log(np_i) : p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i\tilde{g}_i(\beta) = 0\right\},$$

[1]Here we follow the notation of Hahn and Ridder (2013). They reserve the notation $\hat{\mu}(X_i,V_{*i})$ for (infeasible) nonparametric regression of Y_i on (X_i,V_{*i}).

[2]Similar results can be derived for other estimators, such as the kernel and local polynomial regression estimators.

where

$$\begin{aligned}\tilde{g}_i(\beta) &= g(\hat{\gamma}(X_i, \hat{V}_i), \beta) + \hat{\Delta}'\psi(X_i, Z_i, \hat{\alpha}) + g_1(\hat{\gamma}(X_i, \hat{V}_i), \beta)\{Y_i - \hat{\gamma}(X_i, \hat{V}_i)\}, \qquad (3)\\ \hat{\Delta} &= \frac{1}{n}\sum_{i=1}^{n}\{Y_i - \hat{\gamma}(X_i, \hat{V}_i)\}g_2(\hat{\gamma}(X_i, \hat{V}_i), \beta)\hat{\gamma}_v(X_i, \hat{V}_i)\varphi_\alpha(X_i, Z_i, \hat{\alpha}).\end{aligned}$$

Note that our moment function $\tilde{g}_i(\beta)$ is composed of three terms. The first term in (3) is a plug-in version of the original moment function in (1), and the others are correction terms to achieve asymptotic pivotalness. The second term is an adjustment due to the first step estimation of $\hat{V}_i$, and the third term is another adjustment due to the second step estimation of $\hat{\gamma}(\cdot, \cdot)$. By applying the Lagrange multiplier method, the dual form of $\ell(\beta)$ is written as

$$\ell(\beta) = 2\sup_{\lambda}\sum_{i=1}^{n}\log(1 + \lambda\tilde{g}_i(\beta)). \qquad (4)$$

In practice, we compute $\ell(\beta)$ using this dual form. The main result of this paper, the asymptotic distribution of the empirical likelihood ratio, is presented as follows. The proof is given in Appendix A.

Theorem 1. *Consider the setup of this subsection and impose Assumption P in Appendix A. Then*

$$\ell(\beta_*) \xrightarrow{d} \chi^2(1).$$

Remark 1. This theorem says that the empirical likelihood statistic $\ell(\beta_*)$ is asymptotically pivotal and converges to the $\chi^2(1)$ distribution. Based on this theorem, the $100(1-\alpha)\%$ asymptotic confidence interval is constructed as $ELCI_\alpha = \{\beta : \ell(\beta) \leq \chi^2_{1-\alpha}(1)\}$, where $\chi^2_{1-\alpha}(1)$ is the $(1-\alpha)$-th quantile of the $\chi^2(1)$ distribution. This feature of asymptotic pivotalness is particularly attractive in the present setup because the asymptotic variance of the three step estimator $\hat{\beta}$ takes a somewhat complicated form due to three components in the influence function from each step. Although we can express the asymptotic variance of $\hat{\beta}$ based on the influence function derived by Hahn and Ridder (2013), whether we can precisely estimate the asymptotic variance so that the resulting t-ratio is reliable for inference on β_* is another problem entirely. In contrast, our empirical likelihood statistic $\ell(\beta_*)$ is internally studentized and circumvents such asymptotic variance estimation. A drawback of the empirical likelihood confidence interval $ELCI_\alpha$ (compared to the one based on the t-ratio) is that it requires a numerical search to find the endpoints of the interval. However, in most examples of three-step estimation the parameter of interest β_* is scalar and a grid search can be applied to compute $ELCI_\alpha$.

Remark 2. We emphasize that it is crucial to incorporate the last two terms in (3) to achieve the asymptotic pivotalness. Without these terms, the corresponding empirical likelihood ratio $2\sup_\lambda\sum_{i=1}^{n}\log(1 + \lambda g(\hat{\gamma}(X_i, \hat{V}_i), \beta_*))$ converges to the $\chi^2(1)$ distribution multiplied by a constant that depends on some nuisance parameters, and is not asymptotically pivotal. Due to the correction terms in $\tilde{g}_i(\beta_*)$, the Bartlett identity (see, Mykland, 1999) is asymptotically recovered and the resulting empirical likelihood ratio is properly studentized. When there is no first step estimation for the generated regressor (i.e., $\hat{V}_i = V_{*i}$ is directly observable), the properties of the empirical likelihood ratio $2\sup_\lambda \sum_{i=1}^n \log(1+\lambda g(\hat{\gamma}(X_i, V_{*i}), \beta_*))$ with a plug-in estimator $\hat{\gamma}(\cdot,\cdot)$ were investigated by Hjort, McKeague and van Keilegom (2009). They showed that the empirical likelihood ratio with plug-in estimators converges to a weighted sum of chi-squared random variables, and is not asymptotically pivotal. For this problem, in the case of no generated regressor, an insightful paper by Bravo, Escanciano and van Keilegom (2015) introduced a correction term to the moment function to recover asymptotic pivotalness. Our result in Theorem 1 is considered a generalization of Bravo, Escanciano and van Keilegom (2015) to three step estimation problems studied in Hahn and Ridder (2013). Indeed, if there is no first step estimation for generated regressors (i.e., $\hat{V}_i = V_{*i}$), our moment function $\tilde{g}_i(\beta_*)$ in (3) does not have the second term and reduces to the one in Bravo, Escanciano and van Keilegom (2015). It should be noted that this generalization to three step estimation problems is not trivial due to the non-trivial influence of the first step estimation for generated regressors as shown by Hahn and Ridder (2013).

Remark 3. We note that the condition on the bandwidth h to compute $\hat{\gamma}(\cdot,\cdot)$ (Assumption P (iv)) does not require undersmoothing, i.e., we only require $nh^{4s} \to 0$ instead of $nh^{2s} \to 0$. Thus, for example, the MSE optimal bandwidth is allowed. This desirable property is known in the empirical likelihood literature for several setups (e.g., Zhu and Xue, 2006, Bravo, Escanciano and van Keilegom, 2015). Theorem 1 shows that a similar result holds for three step estimation problems. Intuitively, the first and third terms in $\tilde{g}_i(\beta_*)$ share the same form as the smoothing bias and these bias terms are automatically cancelled out. We emphasize that in contrast to the empirical likelihood confidence interval $ELCI_\alpha$, the Wald-type confidence interval using the asymptotic variance estimator based on Hahn and Ridder's (2013) formula requires undersmoothing for the bandwidth.

Remark 4. If the parameter of interest is explicitly defined as $\beta_* = h(\mu(X, V_*))$ for some known function h, then we can apply Theorem 1 by setting $g(\mu(X, V_*), \beta_*) = h(\mu(X, V_*)) - \beta_*$. Also, if g is linear in μ, then the second term in (3) vanishes (by $g_2(\cdot) = 0$), and the moment function simplifies to

$$\tilde{g}_i(\beta) = g(\hat{\gamma}(X_i, \hat{V}_i), \beta) + g_1(\hat{\gamma}(X_i, \hat{V}_i), \beta)\{Y_i - \hat{\gamma}(X_i, \hat{V}_i)\}.$$

Remark 5. Theorem 1 can be generalized to the case of multidimensional μ, where $\mu(X_i, V_{*i}) = (\mu_1(X_i, V_{*i}), \ldots, \mu_K(X_i, V_{*i}))'$ and $\mu_k(X_i, V_{*i}) = E[Y_{k,i}|X_i, V_{*i}]$ for $k = 1, \ldots, K$. In this case, the statistic in (4) is modified by replacing $\tilde{g}_i(\beta)$ with

$$\begin{aligned}
\tilde{g}_i(\beta) &= g(\hat{\gamma}(X_i, \hat{V}_i), \beta) + \hat{\Delta}'\psi(X_i, Z_i, \hat{\alpha}) + \sum_{k=1}^K g_{1k}(\hat{\gamma}(X_i, \hat{V}_i), \beta)\{Y_{k,i} - \hat{\gamma}_k(X_i, \hat{V}_i)\}, \\
\hat{\Delta} &= \sum_{k=1}^K \left[\frac{1}{n} \sum_{i=1}^n \{Y_{k,i} - \hat{\gamma}_k(X_i, \hat{V}_i)\} g_{2k}(\hat{\gamma}(X_i, \hat{V}_i), \beta) \hat{\gamma}_{k,v}(X_i, \hat{V}_i) \varphi_\alpha(X_i, Z_i, \hat{\alpha}) \right],
\end{aligned}$$

where $g_{1k}(\mu, \beta)$ and $g_{2k}(\mu, \beta)$ are the first and second derivatives of $g(\cdot,\cdot)$ with respect to its k-th argument evaluated at (μ, β), respectively, and $\hat{\gamma}_{k,v}(X_i, \hat{V}_i)$ is the slope coefficient of $\hat{V}_i$ in a local polynomial regression of $Y_{k,i}$ on $(X_i, \hat{V}_i)$.

2.2. **Case of nonparametric first step.** We next consider the case where $V_* = \varphi(X, Z)$ is written as an unknown function φ and needs to be estimated by some nonparametric method. In particular, we focus on the situation where V_* is written as the conditional mean (i.e., $V_* = \varphi(X, Z) = E[U|X, Z]$ for some observable U) and $\varphi(X, Z)$ is estimated by the nonparametric kernel estimator $\hat{\varphi}(X, Z)$ using the bandwidth b. We conjecture that analogous results can be derived for other estimators such as series estimators. Let us redefine the generated regressor as $\hat{V}_i = \hat{\varphi}(X_i, Z_i)$.

In this case, we modify the empirical likelihood statistic in (4) by replacing $\tilde{g}_i(\beta)$ with

$$\tilde{g}_i(\beta) = g(\hat{\gamma}(X_i, \hat{V}_i), \beta) + \hat{\Delta}_{1i}(U_i - \hat{V}_i) + g_1(\hat{\gamma}(X_i, \hat{V}_i), \beta)\{Y_i - \hat{\gamma}(X_i, \hat{V}_i)\},$$

where $\hat{\Delta}_{1i}$ is the nonparametric regression fit of $\{Y_i - \hat{\gamma}(X_i, \hat{V}_i)\}g_2(\hat{\gamma}(X_i, \hat{V}_i), \beta)\hat{\gamma}_v(X_i, \hat{V}_i)$ on (X_i, Z_i). Similar to the case of a parametric first step, the empirical likelihood ratio converges to the $\chi^2(1)$ distribution without undersmoothing.

Theorem 2. *Consider the setup of this subsection and impose Assumption NP in Appendix B. Then*

$$\ell(\beta_*) \xrightarrow{d} \chi^2(1).$$

The proof is presented in Appendix B. Similar comments to Theorem 1 apply here. The last two terms in $\tilde{g}_i(\beta)$ recover internal studentization and asymptotic pivotalness. Again, if there is no first step estimation for the generated regressor (i.e., $\hat{V}_i = V_{*i}$ is directly observable), the moment function $\tilde{g}_i(\beta)$ reduces to the one in Bravo, Escanciano and van Keilegom (2015). Similar to the bandwidth h for the second step estimator $\hat{\gamma}(\cdot, \cdot)$, the bandwidth b for the first step estimator $\hat{\varphi}(\cdot, \cdot)$ does not require undersmoothing (see, Assumption NP (i) in Appendix B).

3. Extensions

3.1. **Additional variables in third step.** We now consider an extension to the moment condition

$$E[g(W, \mu(X, V_*), \beta_*)] = 0,$$

where $W \in \mathbb{R}^{d_w}$ is a vector of additional variables. The vector W may contain X and Z as subvectors. This extension is useful to accommodate, for example, partially linear models with generated regressors (see, Section 4.1 below).

Our empirical likelihood approach can be modified to accommodate additional variables W as follows. Let $g_1(w, \mu, \beta)$ be the partial derivative of $g(\cdot, \cdot, \cdot)$ with respect to its $(d_w + 1)$-th argument evaluated at (w, μ, β). In the case of a parametric first step (i.e., $V_* = \varphi(X, Z, \alpha_*)$), the empirical likelihood statistic in (4) is modified by replacing $\tilde{g}_i(\beta)$ with

$$\begin{aligned}
\tilde{g}_i(\beta) &= g(W_i, \hat{\gamma}(X_i, \hat{V}_i), \beta) + \hat{\Delta}_1 \psi(X_i, Z_i, \hat{\alpha}) + \hat{\Delta}_{2i}\{Y_i - \hat{\gamma}(X_i, \hat{V}_i)\}, \\
\hat{\Delta}_1 &= \frac{1}{n}\sum_{i=1}^{n}\{(g_1(W_i, \hat{\gamma}(X_i, \hat{V}_i), \beta) - \hat{\kappa}(X_i, \hat{V}_i))\hat{\gamma}_v(X_i, \hat{V}_i)\varphi_\alpha(X_i, Z_i, \hat{\alpha}) \\
&\qquad + (Y_i - \hat{\gamma}(X_i, \hat{V}_i))\hat{\kappa}_v(X_i, \hat{V}_i)\varphi_\alpha(X_i, Z_i, \hat{\alpha})\},
\end{aligned}$$

$\hat{\kappa}(X_i, \hat{V}_i)$ and $\hat{\kappa}_v(X_i, \hat{V}_i)$ are the intercept and slope coefficient of $\hat{V}_i$ in a local polynomial regression of $g_1(W_i, \hat{\gamma}(X_i, \hat{V}_i), \beta)$ on $(X_i, \hat{V}_i)$, respectively, and $\hat{\Delta}_{2i} = \hat{\kappa}(X_i, \hat{V}_i)$.

In the case of a nonparametric first step (i.e., $V_* = \varphi(X, Z) = E[U|X, Z]$ for some observable U), the statistic in (4) is modified by replacing $\tilde{g}_i(\beta)$ with

$$\begin{aligned}\tilde{g}_i(\beta) &= g(W_i, \hat{\gamma}(X_i, \hat{V}_i), \beta) + \hat{\Delta}_{1i}(U_i - \hat{V}_i) + \hat{\Delta}_{2i}\{Y_i - \hat{\gamma}(X_i, \hat{V}_i)\}, \\ \bar{\Delta}_{1i} &= \{g_1(W_i, \hat{\gamma}(X_i, \hat{V}_i), \beta) - \hat{\kappa}(X_i, \hat{V}_i)\}\hat{\gamma}_v(X_i, \hat{V}_i) + \{Y_i - \hat{\gamma}(X_i, \hat{V}_i)\}\hat{\kappa}_v(X_i, \hat{V}_i),\end{aligned}$$

and $\hat{\Delta}_{1i}$ is the nonparametric regression fit of $\bar{\Delta}_{1i}$ on (X_i, Z_i).

For both cases, it can be shown that the empirical likelihood ratio $\ell(\beta_*)$ converges to the $\chi^2(1)$ distribution (without undersmoothing).

Remark 6. The above results can be extended to the general case where both g and β are vectors. In this case, as far as the dimension of g is greater than or equal to that of β, we can show that $\ell(\beta_*) \xrightarrow{d} \chi^2(\dim g)$ under analogous conditions. Also, the results can be generalized to the case of multidimensional μ.

3.2. **Partial mean case.** In this subsection, we consider an extension to

$$E[g(W, \mu_1(X, V_*), \ldots, \mu_K(X, V_*), \beta_*)] = 0,$$

where $\mu_k(X, V_*) = E[Y|X, V_*, D = d_{(k)}]$ for $k = 1, \ldots, K$ is a vector of conditional means associated with the discrete variable D supported on the values $d_{(1)}, \ldots, d_{(K)}$. This extension is useful to accommodate matching estimators of treatment effects, for example.

Let $g_{1k}(w, \mu_1, \ldots, \mu_K, \beta)$ be the partial derivative of $g(\cdot, \ldots, \cdot)$ with respect to its (d_w+k)-th argument evaluated at $(w, \mu_1, \ldots, \mu_K, \beta)$, $\kappa_k(X_i, V_{*i}) = E[g_{1k}(W_i, \mu_1(X, V_*), \ldots, \mu_K(X, V_*), \beta)|X_i, V_{*i}]$, and $\pi_k(X_i, V_{*i}) = \Pr\{D_i = d_{(k)}|X_i, V_{*i}\}$. In the case of a parametric first step, the empirical likelihood statistic in (4) is modified by replacing $\tilde{g}_i(\beta)$ with

$$\begin{aligned}\tilde{g}_i(\beta) &= g(W_i, \hat{\gamma}_1(X_i, \hat{V}_i), \ldots, \hat{\gamma}_K(X_i, \hat{V}_i), \beta) + \hat{\Delta}_1\psi(X_i, Z_i, \alpha) \\ &\quad + \sum_{k=1}^{K} \mathbb{I}\{D_i = d_{(k)}\}\{Y_i - \hat{\gamma}_k(X_i, \hat{V}_i)\}\frac{\hat{\kappa}_k(X_i, \hat{V}_i)}{\hat{\pi}_k(X_i, \hat{V}_i)},\end{aligned}$$

where $\mathbb{I}\{\cdot\}$ is the indicator function, $\hat{\gamma}_k(X_i, \hat{V}_i)$, $\hat{\gamma}_{k,v}(X_i, \hat{V}_i)$, $\hat{\pi}_k(X_i, \hat{V}_i)$, $\hat{\pi}_{k,v}(X_i, \hat{V}_i)$, $\hat{\kappa}_k(X_i, \hat{V}_i)$, and $\hat{\kappa}_{k,v}(X_i, \hat{V}_i)$ are the local polynomial estimators of $\mu_k(X_i, V_{*i})$, $\partial\mu_k(X_i, V_{*i})/\partial V_{*i}$, $\pi_k(X_i, V_{*i})$, $\partial\pi_k(X_i, V_{*i})/\partial V_{*i}$, $\kappa_k(X_i, V_{*i})$, and $\partial\kappa_k(X_i, V_{*i})/\partial V_{*i}$, respectively, and

$$\begin{aligned}\hat{\Delta}_1 &= \frac{1}{n}\sum_{i=1}^{n}\left[\sum_{k=1}^{K}\left(\begin{array}{c} g_{1k}(W_i, \hat{\gamma}_1(X_i, \hat{V}_i), \ldots, \hat{\gamma}_K(X_i, \hat{V}_i), \beta) \\ -\frac{\mathbb{I}\{D_i = d_{(k)}\}}{\hat{\pi}_k(X_i, \hat{V}_i)}\hat{\kappa}_k(X_i, \hat{V}_i)\end{array}\right)\hat{\gamma}_{k,v}(X_i, \hat{V}_i)\right]\varphi_\alpha(X_i, Z_i, \hat{\alpha}) \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{k=1}^{K}\frac{\mathbb{I}\{D_i = d_{(k)}\}}{\hat{\pi}_k(X_i, \hat{V}_i)}\{Y_i - \hat{\gamma}_k(X_i, \hat{V}_i)\}\hat{\kappa}'_k(X_i, \hat{V}_i)\right)\varphi_\alpha(X_i, Z_i, \hat{\alpha}) \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{k=1}^{K}\frac{\mathbb{I}\{D_i = d_{(k)}\}}{\hat{\pi}_k(X_i, \hat{V}_i)^2}\{Y_i - \hat{\gamma}_k(X_i, \hat{V}_i)\}\hat{\kappa}_k(X_i, \hat{V}_i)\hat{\pi}_{k,v}(X_i, \hat{V}_i)\right)\varphi_\alpha(X_i, Z_i, \hat{\alpha}).\end{aligned}$$

In the case of a nonparametric first step, the statistic in (4) is modified by replacing $\tilde{g}_i(\beta)$ with

$$\begin{aligned}\tilde{g}_i(\beta) &= g(W_i, \hat{\gamma}_1(X_i, \hat{V}_i), \ldots, \hat{\gamma}_K(X_i, \hat{V}_i), \beta) + \hat{\Delta}_{1i}(U_i - \hat{V}_i) \\ &\quad + \sum_{k=1}^{K} \mathbb{I}\{D_i = d_{(k)}\}\{Y_i - \hat{\gamma}_k(X_i, \hat{V}_i)\} \frac{\hat{\kappa}_k(X_i, \hat{V}_i)}{\hat{\pi}_k(X_i, \hat{V}_i)},\end{aligned}$$

where $\hat{\Delta}_{1i}$ is a nonparametric estimator of

$$\begin{aligned}\Delta_{1i} &= E\left[\sum_{k=1}^{K} \left(g_{1k}(W_i, \mu_1(X_i, V_{*i}), \ldots, \mu_K(X_i, V_{*i}), \beta) - \frac{\mathbb{I}\{D_i = d_{(k)}\}}{\pi_k(X_i, V_{*i})} \kappa_k(X_i, V_{*i}) \right) \frac{\partial \mu_k(X_i, V_{*i})}{\partial V_{*i}} \right| X_i, Z_i \Bigg] \\ &\quad + E\left[\sum_{k=1}^{K} \frac{\mathbb{I}\{D_i = d_{(k)}\}}{\pi_k(X_i, V_{*i})} \{Y_i - \mu_k(X_i, V_{*i})\} \frac{\partial \kappa_k(X_i, V_{*i})}{\partial V_{*i}} \right| X_i, Z_i \Bigg] \\ &\quad + E\left[\sum_{k=1}^{K} \frac{\mathbb{I}\{D_i = d_{(k)}\}}{\pi_k(X_i, V_{*i})^2} \{Y_i - \mu_k(X_i, V_{*i})\} \kappa_k(X_i, V_{*i}) \frac{\partial \pi_k(X_i, V_{*i})}{\partial V_{*i}} \right| X_i, Z_i \Bigg].\end{aligned}$$

For both cases, it can be shown that the empirical likelihood ratio $\ell(\beta_*)$ converges to the $\chi^2(1)$ distribution (without undersmoothing).

4. Examples

4.1. **Olley and Pakes (1996) type estimator.** In this subsection, we illustrate our empirical likelihood method using a partially linear model with a generated regressor. This model may be considered a simplified version of the production function model studied in Olley and Pakes (1996). In particular, we consider inference on the slope parameters β_* in the partial linear model with an unobservable regressor V^*:

$$Y = X\beta_* + m(V_*) + \epsilon,$$

where m is an unknown function and $E[\epsilon | X, V_*] = 0$. Estimation of β_* may be interpreted in a three step way. First, we compute the generated regressor $\hat{V}$ as a proxy for V_*. Second, the functions $\mu_1(v) = E[X | V_* = v]$ and $\mu_2(v) = E[Y | V_* = v]$ are estimated by $\hat{\gamma}_1(\hat{V}_i)$ and $\hat{\gamma}_2(\hat{V}_i)$, that is, a nonparametric regression of X on $\hat{V}$ and Y on $\hat{V}$, respectively. Third, the estimator $\hat{\beta}$ can be obtained by solving $\frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\gamma}_1(\hat{V}_i))\{(Y_i - \hat{\gamma}_2(\hat{V}_i)) - (X_i - \hat{\gamma}_1(\hat{V}_i))\hat{\beta}\} = 0$. Based on this condition for $\hat{\beta}$, we consider the moment function $g(X, \mu(V_i), \beta) = (X - \mu_1(V_*))\{(Y_i - \mu_2(V_*)) - (X - \mu_1(V_*))\beta\}$ to apply our empirical likelihood method.

In the case of a parametric first step (i.e., $V_* = \varphi(X, Z, \alpha_*)$), using the fact that $m(V_*) = E[Y - X\beta_* | V_*] = \mu_2(V_*) - \mu_1(V_*)\beta_*$ and a multidimensional version of Hahn and Ridder (2013, Theorem 4), the influence function of $\hat{\beta}$ is obtained as

$$\begin{aligned}&\{X_i - \mu_1(V_{*i})\}\{Y_i - X_i\beta_* - m(V_{*i})\} \\ &- \left[E[(Y_i - X_i\beta_* - m(V_{*i}))\frac{\partial \mu_1(v)}{\partial v}\varphi_\alpha(X, Z, \alpha_*)] + E[(X - \mu_1(V_*))\frac{\partial m(V_*)}{\partial V_*}\varphi_\alpha(X, Z, \alpha_*)]\right]\sqrt{n}(\hat{\alpha} - \alpha),\end{aligned}$$

The t-ratio is given by estimating the asymptotic variance of this function. We note that by Newey (1994, Proposition 2), there is no contribution from $\hat{\gamma}_1$ and $\hat{\gamma}_2$ in this example.

By applying the result in Section 3.1, our empirical likelihood ratio is defined by (4) with

$$\begin{aligned}\tilde{g}_i(\beta) &= \{X_i - \hat{\gamma}_1(\hat{V}_i)\}\{Y_i - X_i\beta - \hat{m}(\hat{V}_i)\} \\ &\quad -\frac{1}{n}\sum_{i=1}^{n}\left[\begin{array}{c}\{Y_i - X_i\beta - \hat{m}(\hat{V}_i)\}\hat{\gamma}_{1,v}(\hat{V}_i)\varphi_\alpha(X_i, Z_i, \hat{\alpha}) \\ +\{X_i - \hat{\gamma}_1(\hat{V}_i)\}\hat{m}_v(\hat{V}_i)\varphi_\alpha(X_i, Z_i, \hat{\alpha})\end{array}\right]\psi(X_i, Z_i, \hat{\alpha}),\end{aligned}$$

where $\hat{\gamma}_1(v)$, $\hat{\gamma}_{1,v}(v)$, $\hat{m}(v)$, and $\hat{m}_v(v)$ are the nonparametric estimators of $\mu_1(v)$, $\frac{\partial\mu_1(v)}{\partial v}$, $\{\mu_2(v) - \mu_1(v)\beta_*\}$, and $\left\{\frac{\partial\mu_2(v)}{\partial v} - \frac{\partial\mu_1(v)}{\partial v}\beta_*\right\}$, respectively.

In the case of a nonparametric first step (i.e., $V_* = \varphi(X, Z)$), our empirical likelihood ratio can be defined by (4) with

$$\tilde{g}_i(\beta) = \{X_i - \hat{\gamma}_1(\hat{V}_i)\}\{Y_i - X_i\beta - \hat{m}(\hat{V}_i)\} + \hat{\Delta}_{2i}(U_i - \hat{V}_i),$$

where $\hat{\Delta}_{2i}$ is the nonparametric regression fit of $-\left[\{Y_i - X_i\beta_* - \hat{m}(\hat{V}_i)\}\hat{\gamma}_{1,v}(\hat{V}_i) + \{X_i - \hat{\gamma}_1(\hat{V}_i)\}\hat{m}_v(\hat{V}_i)\right]$ on (X_i, Z_i).

4.2. **Average treatment effect and counterpart on treated population.** In this subsection, we consider the propensity score matching estimators for the average treatment effect and the one for the treated population. Let $Y_i(1)$ and $Y_i(0)$ denote potential outcomes of unit i with and without exposure to a treatment, respectively. Let $D_i \in \{0, 1\}$ be an indicator variable for the treatment such that $D_i = 1$ if unit i is exposed to the treatment and $D_i = 0$ otherwise. We observe $Z_i = (Y_i, X_i', D_i)'$, where $Y_i = D_iY_i(1) + (1 - D_i)Y_i(0)$ is the observable outcome, and X_i is a vector of covariates.

First, we consider inference on the average treatment effect $\beta_* = E[Y_i(1) - Y_i(0)]$. Let $\varphi(x) = \Pr\{D = 1|X = x\}$ be the propensity score and $\hat{\varphi}(x)$ be its nonparametric estimator (i.e., a nonparametric regression of D on X). Also let $\hat{\gamma}_1(\cdot)$ and $\hat{\gamma}_0(\cdot)$ be the nonparametric regression fits from Y on $\hat{\varphi}(X)$ for the treated and untreated samples, respectively. Then the propensity score matching estimator by Heckman, Ichimura and Todd (1998) is defined as $\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n}\{\hat{\gamma}_1(\hat{\varphi}(X_i)) - \hat{\gamma}_0(\hat{\varphi}(X_i))\}$. This can be interpreted as the method of moments estimator using the moment function $g(X, \mu(V_*), \beta_*) = \mu_1(V_*) - \mu_0(V_*) - \beta_*$, where $\mu_1(v) = E[Y|V = v, D = 1]$, $\mu_0(v) = E[Y|V = v, D = 0]$, and $V_* = \varphi(X)$.

From Hahn and Ridder (2013, Section 4), the influence function of the propensity score matching estimator $\hat{\beta}$ is given by

$$\begin{aligned}&(\mu_1(V_*) - \mu_0(V_*) - \beta_*) - \left(\frac{m_1(X_i) - \mu_1(V_{*i})}{\varphi(X_i)} + \frac{m_0(X_i) - \mu_0(V_{*i})}{1 - \varphi(X_i)}\right)(D_i - \varphi(X_i)) \\ &\quad + \left(\frac{D_i}{\varphi(X_i)}(Y_i - \mu_1(V_{*i})) - \frac{1 - D_i}{1 - \varphi(X_i)}(Y_i - \mu_0(V_{*i}))\right) \\ &= (m_1(X_i) - m_0(X_i) - \beta_*) + \frac{D_i}{\varphi(X_i)}(Y_i - m_1(X_i)) - \frac{1 - D_i}{1 - \varphi(X_i)}(Y_i - m_0(X_i)), \qquad (5)\end{aligned}$$

where $m_1(x) = E[Y|X = x, D = 1]$ and $m_0(x) = E[Y|X = x, D = 0]$. By applying the result in Section 3.2, our empirical likelihood ratio is defined by (4) with

$$\tilde{g}_i(\beta) = (\hat{m}_1(X_i) - \hat{m}_0(X_i) - \beta) + \frac{D_i}{\hat{\varphi}(X_i)}(Y_i - \hat{m}_1(X_i)) - \frac{1 - D_i}{1 - \hat{\varphi}(X_i)}(Y_i - \hat{m}_0(X_i)). \qquad (6)$$

We note that the influence function in (5) is identical for other asymptotically efficient estimators, such as the inverse probability weighted estimator (Hirano, Imbens and Ridder, 2003). Indeed, Bravo, Escanciano and van Keilegom (2015) modified the moment function for the inverse probability weighted estimator and obtained the same function in (6).

Next, we consider the average treatment effect on the treated population $\beta_* = E[Y_i(1) - Y_i(0)|D_i = 1]$. To simplify the presentation, we assume $p = \Pr\{D_i = 1\}$ is known as in Hahn and Ridder (2013). In this case, from Hahn and Ridder (2013, Section 4), the influence function of the propensity score matching estimator $\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n} \frac{D_i}{p}\{\hat{\gamma}_1(\hat{\varphi}(X_i)) - \hat{\gamma}_0(\hat{\varphi}(X_i))\}$ is given by

$$\begin{aligned}
&\frac{D_i}{p}(\mu_1(V_*) - \mu_0(V_*) - \beta_*) - \frac{m_0(X_i) - \mu_0(V_{*i})}{p(1-\varphi(X_i))}(D_i - \varphi(X_i)) \\
&\quad + \left(\frac{D_i}{p}(Y_i - \mu_1(V_{*i})) - \frac{(1-D_i)\varphi(X_i)}{p(1-\varphi(X_i))}(Y_i - \mu_0(V_{*i})) \right) \\
&= \frac{D_i}{p}(m_1(X_i) - m_0(X_i) - \beta_*) + \frac{D_i}{p}(Y_i - m_1(X_i)) - \frac{(1-D_i)\varphi(X_i)}{p(1-\varphi(X_i))}(Y_i - m_0(X_i)).
\end{aligned}$$

By applying the result in Section 3.2, our empirical likelihood ratio is defined by (4) with

$$\tilde{g}_i(\beta) = \frac{D_i}{p}(\hat{m}_1(X_i) - \hat{m}_0(X_i) - \beta) + \frac{D_i}{p}(Y_i - \hat{m}_1(X_i)) - \frac{(1-D_i)\hat{\varphi}(X_i)}{p(1-\hat{\varphi}(X_i))}(Y_i - \hat{m}_0(X_i)),$$

where $\hat{m}_1(X_i)$ and $\hat{m}_0(X_i)$ are nonparametric estimators of $m_1(X_i)$ and $m_0(X_i)$, respectively.

Appendix A. Appendix for Theorem 1

A.1. **Notation and Assumptions.** Let $\mathbb{X}$, $\mathbb{Z}$, and $\mathbb{V}$ be support of X, Z, and V_*, respectively, and

$$\Delta' = E[h_2(\mu(X,V_*))\mu_v(X,V_*)\{\mu(X,Z) - \mu(X,V_*)\}\varphi'_\alpha(X,Z,\alpha_*)].$$

We impose the following assumptions for Theorem 1.

Assumption P.

(i): *$\{Y_i, X'_i, Z'_i\}_{i=1}^n$ is an iid sample from $(Y, X', Z') \in \mathbb{R}\times\mathbb{X}\times\mathbb{Z}$. $\mathbb{X}$, $\mathbb{Z}$, and $\mathbb{V}$ are compact. The joint density $f(x,v)$ of (X,V_*) is continuously differentiable to order $s \geq 2$ and bounded away from zero on $\mathbb{X}\times\mathbb{V}$. $\mu(x,v)$ is continuously differentiable to order $s \geq 2$ on $\mathbb{X}\times\mathbb{V}$. For some $p \geq 4$, $E|Y|^p < \infty$ and $E[|Y|^p|X = x, V_* = v]$ is bounded over $\mathbb{X}\times\mathbb{V}$.*

(ii): *g is twice differentiable with the bounded second derivative. For some neighborhood $\mathcal{N}$ of α_*, $\varphi_{\alpha\alpha}(x,z,\alpha)$ is continuous over $\mathbb{X}\times\mathbb{V}\times\mathcal{N}$.*

(iii): *$K(\cdot)$ integrates to one, is compactly supported and twice differentiable with bounded derivatives, and satisfies $\int K(u)u_1^{j_1}\cdots u_{d_x+1}^{j_{d_x+1}}du = 0$ for all vector of non-negative integers $(j_1,\ldots,j_{d_x+1})$ such that $j_1+\cdots+j_{d_x+1} < s$.*

(iv): *As $n\to\infty$, it holds $n^{1/2}h^{d_x+1}/\log n \to \infty$ and $nh^{4s}\to 0$.*

(v): *$\hat\alpha$ satisfies (2) with $E|\psi(X,Z,\alpha_*)|^2 < \infty$ and $n^{-1}\sum_{i=1}^n \psi(X_i,Z_i,\hat\alpha) = o_p(n^{-1/2})$.*

(vi): $\hat\Delta \xrightarrow{p} \Delta$.

Hereafter, we use the following notation. By suppressing dependence on $(X_j - x)/h$, define

$$\begin{aligned}
\xi_{*j}(v) &= [1,(X_j-x)/h,(V_{*j}-v)/h]', \qquad \hat\xi_j(v) = [1,(X_j-x)/h,(\hat V_j - v)/h]',\\
\Phi(V_{*j},v) &= e'_1\left[\frac{1}{nh^{d_x+1}}\sum_{j=1}^n \xi_{*j}(v)\xi_{*j}(v)'K\left(\frac{X_j-x}{h},\frac{V_{*j}-v}{h}\right)\right]^{-1}\xi_{*j}(v)K\left(\frac{X_j-x}{h},\frac{V_{*j}-v}{h}\right),\\
\Phi(\hat V_j,v) &= e'_1\left[\frac{1}{nh^{d_x+1}}\sum_{j=1}^n \hat\xi_j(v)\hat\xi_j(v)'K\left(\frac{X_j-x}{h},\frac{\hat V_j-v}{h}\right)\right]^{-1}\hat\xi_j(v)K\left(\frac{X_j-x}{h},\frac{\hat V_j-v}{h}\right).
\end{aligned}$$

where $e'_1 = (1,0,\ldots,0)$. Then we denote

$$\begin{aligned}
\hat\mu(X_i,V_{*i}) &= \frac{1}{nh^{d_x+1}}\sum_{j=1}^n \Phi(V_{*j},V_{*i})Y_j,\\
\hat\gamma(X_i,V_{*i}) &= \frac{1}{nh^{d_x+1}}\sum_{j=1}^n \Phi(\hat V_j,V_{*i})Y_j,\\
\hat\gamma(X_i,\hat V_i) &= \frac{1}{nh^{d_x+1}}\sum_{j=1}^n \Phi(\hat V_j,\hat V_i)Y_j.
\end{aligned}$$

Also, let $\Phi_v(\cdot,\cdot)$ be the derivative with respect to its second argument. Finally, denote $\varphi_{\alpha,i} = \varphi_\alpha(X_i,Z_i,\alpha_*)$ and

$$\Omega = E[g(\mu(X,V_*),\beta_*) + \Delta'\psi(X,Z,\alpha_*) + h_1(\mu(X,V_*))\{Y-\mu(X,V_*)\}]^2.$$

A.2. **Lemmas.**

Lemma A.1. *Under Assumption P,*

$$\begin{aligned}\max_{1\le i\le n}|\hat{\mu}(X_i,V_{*i})-\mu(X_i,V_{*i})| &= o_p(n^{-1/4}),\\ \max_{1\le i\le n}|\hat{\gamma}(X_i,V_{*i})-\mu(X_i,V_{*i})| &= o_p(n^{-1/4}),\\ \max_{1\le i\le n}|\hat{\gamma}(X_i,\hat{V}_i)-\mu(X_i,V_{*i})| &= o_p(n^{-1/4}).\end{aligned}$$

Proof. By Assumption P (i), both X_i and V_{*i} are compactly supported, and their joint density is bounded away from zero. Thus, an application of Hansen (2008, Theorem 10) yields the first statement. The second and third statements follow by expansions around $\hat{\alpha}=\alpha_*$ combined with $\sqrt{n}(\hat{\alpha}-\alpha_*)=O_p(1)$ (by Assumption P (v)) and the first statement.

Lemma A.2. *Under Assumption P,*

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{g(\hat{\gamma}(X_i,\hat{V}_i),\beta_*)-g(\hat{\gamma}(X_i,V_{*i}),\beta_*)\}\\ =\;& E[g_1(\mu(X_i,V_{*i}),\beta_*)\mu_v(X_i,V_{*i})\varphi'_{\alpha,i}]\sqrt{n}(\hat{\alpha}-\alpha_*)+o_p(1).\end{aligned}$$

Proof. Observe that

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{g(\hat{\gamma}(X_i,\hat{V}_i),\beta_*)-g(\hat{\gamma}(X_i,V_{*i}),\beta_*)\}\\ =\;&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}g_1(\mu(X_i,V_{*i}),\beta_*)\frac{1}{nh^{d_x+1}}\sum_{j=1}^{n}\{\Phi(\hat{V}_j,\hat{V}_i)-\Phi(\hat{V}_j,V_{*i})\}Y_j+o_p(1)\\ =\;&\frac{1}{n}\sum_{i=1}^{n}g_1(\mu(X_i,V_{*i}),\beta_*)\left\{\frac{1}{nh^{d_x+1}}\sum_{j=1}^{n}\Phi_v(V_{*j},V_{*i})Y_j\right\}\varphi'_{\alpha,i}\sqrt{n}(\hat{\alpha}-\alpha_*)+o_p(1)\\ =\;&\frac{1}{n}\sum_{i=1}^{n}g_1(\mu(X_i,V_{*i}),\beta_*)\mu_v(X_i,V_{*i})\varphi'_{\alpha,i}\sqrt{n}(\hat{\alpha}-\alpha_*)+o_p(1)\\ =\;&E[g_1(\mu(X_i,V_{*i}),\beta_*)\mu_v(X_i,V_{*i})\varphi'_{\alpha,i}]\sqrt{n}(\hat{\alpha}-\alpha_*)+o_p(1),\end{aligned}$$

where the first equality follows from expansions around $\hat{\gamma}(X_i,\hat{V}_i)=\hat{\gamma}(X_i,V_{*i})$ and $\hat{\gamma}(X_i,V_{*i})=\mu(X_i,V_{*i})$, Lemma A.1, and boundedness of h_2 (Assumption P (ii)), the second equality follows from an expansion around $\hat{\alpha}=\alpha_*$and $\sqrt{n}(\hat{\alpha}-\alpha_*)=O_p(1)$ (by Assumption P (v)) combined with boundedness of $g_1(\mu(x,v),\beta)$ over $\mathbb{X}\times\mathbb{V}$ and $\varphi_\alpha(x,z,\alpha)$ and $\varphi_{\alpha\alpha}(x,z,\alpha)$ over $\mathbb{X}\times\mathbb{Z}\times\mathcal{N}$ (Assumption P (ii)), the third equality follows from the uniform convergence of the derivative of the local linear estimator, and the last equality follows from the law of large numbers.

Lemma A.3. *Under Assumption P,*

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{g(\hat{\gamma}(X_i,V_{*i}),\beta_*)-g(\hat{\mu}(X_i,V_{*i}),\beta_*)\}\\ =\;&-E[g_1(\mu(X_i,V_{*i}),\beta_*)\mu_v(X_i,V_{*i})\varphi'_{\alpha,i}]\sqrt{n}(\hat{\alpha}-\alpha_*)+\Delta'\sqrt{n}(\hat{\alpha}-\alpha_*)+o_p(1).\end{aligned}$$

Proof. Let $\mu_{xv,i}=\left(\mu(X_i,V_{*i}),\frac{\partial\mu(X_i,V_{*i})}{\partial x}h,\frac{\partial\mu(X_i,V_{*i})}{\partial v}h\right)'$. Decompose

$$Y_j=\mu'_{xv,i}\hat{\xi}_j(V_{*i})-\{\mu'_{xv,i}\hat{\xi}_j(V_{*i})-\mu'_{xv,i}\xi_{*j}(V_{*i})\}+\{\mu(X_j,V_{*j})-\mu'_{xv,i}\xi_{*j}(V_{*i})\}+\{\mu(X_j,Z_j)-\mu(X_j,V_{*j})\}+\epsilon_j,$$

where the error term $\epsilon_j = Y_j - \mu(X_j, Z_j)$ satisfies $E[\epsilon_j|X_j, Z_j] = 0$. By this expression, we can write as

$$\hat{\gamma}(X_i, V_{*i}) - \hat{\mu}(X_i, V_{*i}) = m_i^A + m_i^B + m_i^C + m_i^D + m_i^E,$$

where

$$\begin{aligned}
m_i^A &= \frac{1}{nh^{d_x+1}} \sum_{j=1}^{n} \Phi(\hat{V}_j, V_{*i}) \mu'_{xv,i} \hat{\xi}_j(V_{*i}) - \frac{1}{nh^{d_x+1}} \sum_{j=1}^{n} \Phi(V_{*j}, V_{*i}) \mu'_{xv,i} \xi_{*j}(V_{*i}), \\
m_i^B &= -\frac{1}{nh^{d_x+1}} \sum_{j=1}^{n} \Phi(\hat{V}_j, V_{*i}) \{\mu'_{xv,i} \hat{\xi}_j(V_{*i}) - \mu'_{xv,i} \xi_{*j}(V_{*i})\}, \\
m_i^C &= \frac{1}{nh^{d_x+1}} \sum_{j=1}^{n} \{\Phi(\hat{V}_j, V_{*i}) - \Phi(V_{*j}, V_{*i})\}\{\mu(X_j, V_{*j}) - \mu'_{xv,i} \xi_{*j}(V_{*i})\}, \\
m_i^D &= \frac{1}{nh^{d_x+1}} \sum_{j=1}^{n} \{\Phi(\hat{V}_j, V_{*i}) - \Phi(V_{*j}, V_{*i})\}\{\mu(X_j, Z_j) - \mu(X_j, V_{*j})\}, \\
m_i^E &= \frac{1}{nh^{d_x+1}} \sum_{j=1}^{n} \{\Phi(\hat{V}_j, V_{*i}) - \Phi(V_{*j}, V_{*i})\} \epsilon_j.
\end{aligned}$$

Note that $m_i^A = 0$ by construction. Thus, an expansion of $h(\hat{\gamma}(X_i, V_{*i}))$ around $\hat{\gamma}(X_i, V_{*i}) = \hat{\mu}(X_i, V_{*i})$ and Lemma A.1 yield

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{g(\hat{\gamma}(X_i, \hat{V}_i), \beta_*) - g(\hat{\gamma}(X_i, V_{*i}), \beta_*)\} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_1(\hat{\mu}(X_i, V_{*i}), \beta_*) \{\hat{\gamma}(X_i, V_{*i}) - \hat{\mu}(X_i, V_{*i})\} + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_1(\hat{\mu}(X_i, V_{*i}), \beta_*) \{m_i^B + m_i^C + m_i^D + m_i^E\} + o_p(1) \\
&\equiv M^B + M^C + M^D + M^E + o_p(1).
\end{aligned}$$

For M^B, we have

$$\begin{aligned}
M^B &= -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_1(\hat{\mu}(X_i, V_{*i}), \beta_*) \frac{1}{nh^{d_x+1}} \sum_{j=1}^{n} \Phi(\hat{V}_j, V_{*i}) (\mu'_{xv,i} \hat{\xi}_j(V_{*i}) - \mu'_{xv,i} \xi_{*j}(V_{*i})) \\
&= -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_1(\hat{\mu}(X_i, V_{*i}), \beta_*) \frac{1}{nh^{d_x+1}} \sum_{j=1}^{n} \Phi(\hat{V}_j, V_{*i}) \mu_v(X_i, V_{*i}) (\hat{V}_j - V_{*j}) \\
&= -\frac{1}{n} \sum_{i=1}^{n} g_1(\mu(X_i, V_{*i}), \beta_*) \frac{1}{nh^{d_x+1}} \sum_{j=1}^{n} \Phi(V_{*j}, V_{*i}) \mu_v(X_i, V_{*i}) \varphi'_{\alpha,j} \sqrt{n}(\hat{\alpha} - \alpha_*) + o_p(1) \\
&= -\frac{1}{n} \sum_{j=1}^{n} \left\{ \frac{1}{nh^{d_x+1}} \sum_{i=1}^{n} \Phi(V_{*i}, V_{*j}) g_1(\mu(X_i, V_{*i}), \beta_*) \mu_v(X_i, V_{*i}) \right\} \varphi'_{\alpha,j} \sqrt{n}(\hat{\alpha} - \alpha_*) + o_p(1) \\
&= -\frac{1}{n} \sum_{j=1}^{n} g_1(\mu(X_j, V_{*j}), \beta_*) \mu_v(X_j, V_{*j}) \varphi'_{\alpha,j} \sqrt{n}(\hat{\alpha} - \alpha_*) + o_p(1) \\
&= -E[g_1(\mu(X_i, V_{*i}), \beta_*) \mu_v(X_i, V_{*i}) \varphi'_{\alpha,i}] \sqrt{n}(\hat{\alpha} - \alpha_*) + o_p(1),
\end{aligned}$$

where the first equality is the definition of M^B, the second equality follows from the definitions of $\hat{\xi}_j(V_{*i})$ and $\xi_{*j}(V_{*i})$, the third equality follows from expansions around $\hat{\mu}(X_i, V_{*i}) = \mu(X_i, V_{*i})$ and $\hat{\alpha} = \alpha_*$ combined with Lemma A.1, $\sqrt{n}(\hat{\alpha} - \alpha_*) = O_p(1)$, and Assumption P (ii), the fourth equality follows by exchanging the order of summations and the fact that $\sum_{i=1}^n \Phi(V_{*j}, V_{*i})a_i = \sum_{i=1}^n \Phi(V_{*i}, V_{*j})a_i$ for any a_i (because it is the intercept of the weighted OLS), the fifth equality follows from the uniform convergence of the local linear estimator, and the last equality follow from the law of large numbers.

For M^C, we have

$$\begin{aligned} M^C &= \frac{1}{\sqrt{n}} \sum_{i=1}^n g_1(\hat{\mu}(X_i, V_{*i}), \beta_*) \frac{1}{nh^{d_x+1}} \sum_{j=1}^n \{\Phi(V_{*i}, \hat{V}_j) - \Phi(V_{*i}, V_{*j})\}\{\mu(X_j, V_{*j}) - \mu'_{xv,i}\xi_{*j}(V_{*i})\} \\ &= \frac{1}{n} \sum_{i=1}^n g_1(\mu(X_i, V_{*i}), \beta_*) \frac{1}{nh^{d_x+1}} \sum_{j=1}^n \Phi_v(V_{*i}, V_{*j})\{\mu(X_j, V_{*j}) - \mu'_{xv,i}\xi_{*j}(V_{*i})\}\varphi'_{\alpha,j}\sqrt{n}(\hat{\alpha} - \alpha_*) + o_p(1) \\ &= o_p(1), \end{aligned}$$

where the first equality is the definition of M^C and the fact that $\sum_{i=1}^n \Phi(V_{*j}, V_{*i})a_i = \sum_{i=1}^n \Phi(V_{*i}, V_{*j})a_i$ for any a_i, the second equality follows from expansions around $\hat{\mu}(X_i, V_{*i}) = \mu(X_i, V_{*i})$ and $\hat{\alpha} = \alpha_*$combined with Lemma A.1, $\sqrt{n}(\hat{\alpha} - \alpha_*) = O_p(1)$, and Assumption P (ii), and the third equality follows by exchanging the order of summations and the uniform convergence of the derivative of the local linear estimator.

For M^D, we have

$$\begin{aligned} M^D &= \frac{1}{\sqrt{n}} \sum_{i=1}^n g_1(\hat{\mu}(X_i, V_{*i}), \beta_*) \frac{1}{nh^{d_x+1}} \sum_{j=1}^n \{\Phi(V_{*i}, \hat{V}_j) - \Phi(V_{*i}, V_{*j})\}\{\mu(X_j, Z_j) - \mu(X_j, V_{*j})\} \\ &= \frac{1}{n} \sum_{i=1}^n g_1(\mu(X_i, V_{*i}), \beta_*) \frac{1}{nh^{d_x+1}} \sum_{j=1}^n \Phi_v(V_{*i}, V_{*j})\{\mu(X_j, Z_j) - \mu(X_j, V_{*j})\}\varphi'_{\alpha,j}\sqrt{n}(\hat{\alpha} - \alpha_*) + o_p(1) \\ &= \frac{1}{n} \sum_{j=1}^n g_2(\mu(X_j, V_{*j}), \beta_*)\mu_v(X_j, V_{*j})\{\mu(X_j, Z_j) - \mu(X_j, V_{*j})\}\varphi'_{\alpha,j}\sqrt{n}(\hat{\alpha} - \alpha_*) + o_p(1) \\ &= \Delta'\sqrt{n}(\hat{\alpha} - \alpha_*) + o_p(1), \end{aligned}$$

where the first equality is the definition of M^D and the fact that $\sum_{i=1}^n \Phi(V_{*j}, V_{*i})a_i = \sum_{i=1}^n \Phi(V_{*i}, V_{*j})a_i$ for any a_i, the second equality follows from expansions around $\hat{\mu}(X_i, V_{*i}) = \mu(X_i, V_{*i})$ and $\hat{\alpha} = \alpha_*$combined with Lemma A.1, $\sqrt{n}(\hat{\alpha} - \alpha_*) = O_p(1)$, and Assumption P (ii), the third equality follows by exchanging the order of summations and the uniform convergence of the derivative of the local linear estimator, and the last equality follows from the law of large numbers.

For M^E, a similar argument to M^C using $E[\epsilon|X, Z] = 0$ yields $M_E = o_p(1)$. Therefore, combining the results for all terms, the conclusion follows.

Lemma A.4. *Under Assumption P,* $\frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{g}_i(\beta_*) \xrightarrow{d} N(0, \Omega)$.

Proof. By Lemmas A.2 and A.3,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \{g(\hat{\gamma}(X_i, \hat{V}_i), \beta_*) - g(\hat{\mu}(X_i, V_{*i}), \beta_*)\} = \Delta'\sqrt{n}(\hat{\alpha} - \alpha_*) + o_p(1).$$

By this and an expansion of $g(\hat{\mu}(X_i, V_{*i}), \beta_*)$ around $\hat{\mu}(X_i, V_{*i}) = \mu(X_i, V_{*i})$, we can decompose

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{g}_i(\beta_*) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(\mu(X_i, V_{*i}), \beta_*) + M_1 + M_2 + o_p(1),$$

where

$$\begin{aligned} M_1 &= \Delta' \sqrt{n}(\hat{\alpha} - \alpha_*) + \hat{\Delta}' \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(X_i, Z_i, \hat{\alpha}), \\ M_2 &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[g_1(\mu(X_i, V_{*i}), \beta_*)\{\hat{\mu}(X_i, V_{*i}) - \mu(X_i, V_{*i})\} + g_1(\hat{\gamma}(X_i, \hat{V}_i), \beta_*)\{Y_i - \hat{\gamma}(X_i, \hat{V}_i)\} \right]. \end{aligned}$$

Thus, suppose we have

$$M_1 = \Delta' \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(X_i, Z_i, \alpha_*) + o_p(1), \tag{7}$$

$$M_2 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_1(\mu(X_i, V_{*i}), \beta_*)\{Y_i - \mu(X_i, V_{*i})\} + o_p(1), \tag{8}$$

Then the central limit theorem implies the conclusion.

Since the relation (7) follows from Assumption P (v)-(vi), it remains to show (8). Decompose

$$M_2 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_1(\mu(X_i, V_{*i}), \beta_*)\{Y_i - \mu(X_i, V_{*i})\} + M_{21} + M_{22} + M_{23},$$

where

$$\begin{aligned} M_{21} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_1(\mu(X_i, V_{*i}), \beta_*)\{\hat{\mu}(X_i, V_{*i}) - \mu(X_i, V_{*i})\}, \\ M_{22} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{g_1(\hat{\gamma}(X_i, \hat{V}_i), \beta_*) - g_1(\mu(X_i, V_{*i}), \beta_*)\}\{Y_i - \mu(X_i, V_{*i})\}, \\ M_{23} &= -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_1(\hat{\gamma}(X_i, \hat{V}_i), \beta_*)\{\hat{\gamma}(X_i, \hat{V}_i) - \mu(X_i, V_{*i})\}. \end{aligned}$$

For M_{22}, the same argument to the proof of Lemma A.2 and Lemma A.3 implies

$$M_{22} = \frac{1}{n} \sum_{i=1} \Delta' \{Y_i - \mu(X_i, V_{*i})\} \sqrt{n}(\hat{\alpha} - \alpha_*) + o_p(1) = o_p(1).$$

For M_{23}, we further decompose

$$\begin{aligned} M_{23} &= -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_1(\mu(X_i, V_{*i}), \beta_*)\{\hat{\gamma}(X_i, \hat{V}_i) - \mu(X_i, V_{*i})\} \\ &\quad -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{g_1(\hat{\gamma}(X_i, \hat{V}_i), \beta_*) - g_1(\mu(X_i, V_{*i}), \beta_*)\}\{\hat{\gamma}(X_i, \hat{V}_i) - \mu(X_i, V_{*i})\} \\ &= M_{231} + M_{232}. \end{aligned}$$

From the same argument to the proof of Lemma A.2 and Lemma A.3 (by setting $g(\cdot)$ as the identity map), we have

$$\begin{aligned}
M_{231} &= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_1(\mu(X_i,V_{*i}),\beta_*)\left[\{\hat{\gamma}(X_i,\hat{V}_i)-\hat{\mu}(X_i,V_{*i})\}+\{\hat{\mu}(X_i,V_{*i})-\mu(X_i,V_{*i})\}\right] \\
&= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_1(\mu(X_i,V_{*i}),\beta_*)\{\hat{\mu}(X_i,V_{*i})-\mu(X_i,V_{*i})\}+o_p(1).
\end{aligned}$$

For M_{232}, we have

$$\begin{aligned}
M_{232} &= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big[\{g_1(\hat{\gamma}(X_i,\hat{V}_i),\beta_*)-g_1(\hat{\gamma}(X_i,V_{*i}),\beta_*)\}+\{g_1(\hat{\gamma}(X_i,V_{*i},\beta_*))-g_1(\hat{\mu}(X_i,V_{*i}),\beta_*)\} \\
&\qquad +\{g_1(\hat{\mu}(X_i,V_{*i}),\beta_*)-g_1(\mu(X_i,V_{*i}),\beta_*)\}] \\
&\quad \times\left[\{\hat{\gamma}(X_i,\hat{V}_i)-\hat{\gamma}(X_i,V_{*i})\}+\{\hat{\gamma}(X_i,V_{*i})-\hat{\mu}(X_i,V_{*i})\}+\{\hat{\mu}(X_i,V_{*i})-\mu(X_i,V_{*i})\}\right] \\
&= o_p(1).
\end{aligned}$$

The last equality follows the same argument as above combined with the standard argument for degenerated U-statistics.

Finally, note that M_{21} and the main term of M_{231} are cancelled out. Therefore, the conclusion follows.

Lemma A.5. *Under Assumption P,* $\max_{1\leq i\leq n}|\tilde{g}_i(\beta_*)| = o_p(n^{1/2})$.

Proof. The result follows as in Owen (1990, Lemma 3) by the Borel-Cantelli lemma based on $E|Y|^2<\infty$ and Assumption P (i).

Lemma A.6. *Under Assumption P,* $n^{-1}\sum_{i=1}^{n}\tilde{g}_i(\beta_*)^2 \xrightarrow{p} \Omega$.

Proof. The proof follows by a similar argument to the proof of Lemma A.4.

A.3. **Proof of Theorem 1.** First, by Lemmas A.4, A.5 and A.6, the same arguments as in the proof of Owen (1990, eq. (2.14)) implies that $\hat{\lambda}=O_p(n^{-1/2})$.

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

$$\begin{aligned}
0 &= \frac{1}{n}\sum_{i=1}^{n}\frac{\tilde{h}_i(\beta_*)}{1+\hat{\lambda}\tilde{g}_i(\beta_*)} \\
&= \frac{1}{n}\sum_{i=1}^{n}\tilde{g}_i(\beta_*)-\frac{1}{n}\sum_{i=1}^{n}\tilde{g}_i(\beta_*)^2\hat{\lambda}+\frac{1}{n}\sum_{i=1}^{n}\frac{\tilde{g}_i(\beta_*)^3\hat{\lambda}^2}{1+\hat{\lambda}\tilde{g}_i(\beta_*)},
\end{aligned}$$

where the second equality follows from the identity $(1+x)^{-1}=1-x+x^2(1+x)^{-1}$. By applying Lemmas A.4, A.5 and A.6, and $\hat{\lambda}=O_p(n^{-1/2})$, we have

$$\hat{\lambda}=\frac{\sum_{i=1}^{n}\tilde{g}_i(\beta_*)}{\sum_{i=1}^{n}\tilde{g}_i(\beta_*)^2}+o_p(n^{-1/2}).$$

Finally, a Taylor expansion yields

$$
\begin{aligned}
2\sum_{i=1}^{n}\log(1+\hat{\lambda}\tilde{g}_i(\beta_*)) &= 2\sum_{i=1}^{n}\left[\hat{\lambda}\tilde{g}_i(\beta_*)-\frac{1}{2}\{\hat{\lambda}\tilde{g}_i(\beta_*)\}^2\right]+o_p(1) \\
&= \frac{\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{g}_i(\beta_*)\right]^2}{\frac{1}{n}\sum_{i=1}^{n}\tilde{g}_i(\beta_*)^2}+o_p(1).
\end{aligned}
$$

The conclusion follows by Lemmas A.4 and A.6.

Appendix B. Appendix for Theorem 2

B.1. **Notation and Assumptions.** Define

$$
\begin{aligned}
\Delta_i &= E[\{Y_i-\mu(X_i,V_{*i})\}g_2(\mu(X_i,V_{*i}),\beta_*)\mu_v(X_i,V_{*i})|X_i,Z_i], \\
\Omega &= E[g(\mu(X_i,V_{*i}),\beta_*)+\Delta_i'(U_i-V_{*i})+h_1(\mu(X_i,V_{*i}))\{Y_i-\mu(X_i,V_{*i})\}]^2.
\end{aligned}
$$

We impose the following assumptions for Theorem 2.

Assumption NP. *In addition to Assumption P (i)-(iv), suppose*

- **(i):** *As $n\to\infty$, it holds $n^{1/2}b^{d_x+d_z}/\log n\to\infty$ and $nb^{4s}\to 0$.*
- **(ii):** *The joint density $f(x,z)$ of (X,Z) is continuously differentiable to order $s\geq 2$ and bounded away from zero on $\mathbb{X}\times\mathbb{Z}$. The functions $\varphi(x,z)=E[U|X=x,Z=z]$ is continuously differentiable to order $s\geq 2$ on $\mathbb{X}\times\mathbb{Z}$. For some $p\geq 4$, $E|U^p|<\infty$ and $E[|U|^p|X=x,Z=z]f(x,z)$ is bounded on $\mathbb{X}\times\mathbb{Z}$.*
- **(iii):** $\max_{1\leq i\leq n}|\hat{\Delta}_{1i}-\Delta_i|\xrightarrow{p}0$.

B.2. **Lemmas.**

Lemma B.1. *Under Assumption NP,*

$$
\begin{aligned}
&\max_{1\leq i\leq n}|\hat{V}_i-V_{*i}|=o_p(n^{-1/4}), \\
&\max_{1\leq i\leq n}|\hat{\mu}(X_i,V_{*i})-\mu(X_i,V_{*i})|=o_p(n^{-1/4}), \\
&\max_{1\leq i\leq n}|\hat{\gamma}(X_i,V_{*i})-\mu(X_i,V_{*i})|=o_p(n^{-1/4}), \\
&\max_{1\leq i\leq n}|\hat{\gamma}(X_i,\hat{V}_i)-\mu(X_i,V_{*i})|=o_p(n^{-1/4}).
\end{aligned}
$$

Proof. The first statement follows from Assumption NP (ii) and the same argument as in Lemma A.1. The second statement is the same as in Lemma A.1. The third and fourth statements follow by expansions around $\hat{V}_i=V_{*i}$ combined with the first and second statements.

Lemma B.2. *Under Assumption NP,*

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{g(\hat{\gamma}(X_i,\hat{V}_i),\beta_*)-g(\hat{\gamma}(X_i,V_{*i}),\beta_*)\} \\
=\ &\frac{1}{\sqrt{n}}\sum_{i=1}^{n}g_1(\mu(X_i,V_{*i}),\beta_*)\mu_v(X_i,V_{*i})(\hat{V}_i-V_{*i})+o_p(1).
\end{aligned}
$$

Proof. This follows from Lemma B.1 and the same argument as in Lemma A.2.

Lemma B.3. *Under Assumption NP,*

$$\begin{aligned} & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{g(\hat{\gamma}(X_i,V_{*i}),\beta_*)-g(\hat{\mu}(X_i,V_{*i}),\beta_*)\} \\ = & -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}g_1(\mu(X_i,V_{*i}),\beta_*)\mu_v(X_i,V_{*i})(\hat{V}_i-V_{*i}) \\ & +\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Delta_i(\hat{V}_i-V_{*i})+o_p(1). \end{aligned}$$

Proof. By the same argument as in Lemma A.3, we have

$$\begin{aligned} & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{g(\hat{\gamma}(X_i,V_{*i}),\beta_*)-g(\hat{\mu}(X_i,V_{*i}),\beta_*)\} \\ = & -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}g_1(\mu(X_i,V_{*i}),\beta_*)\mu_v(X_i,V_{*i})(\hat{V}_i-V_{*i}) \\ & +\frac{1}{\sqrt{n}}\sum_{i=1}^{n}g_2(\mu(X_j,V_{*j}),\beta_*)\mu_v(X_j,V_{*j})\{\mu(X_j,Z_j)-\mu(X_j,V_{*j})\}(\hat{V}_i-V_{*i})+o_p(1). \end{aligned}$$

Applying the standard argument using degenerated U-statistics to the last term yields the conclusion.

Lemma B.4. *Under Assumption NP,* $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{g}_i(\beta_*)\xrightarrow{d}N(0,\Omega)$.

Proof. By Lemmas B.2 and B.3,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{h(\hat{\gamma}(X_i,\hat{V}_i))-h(\hat{\mu}(X_i,V_{*i}))\}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Delta_i(\hat{V}_i-V_{*i})+o_p(1).$$

By this and an expansion of $h(\hat{\mu}(X_i,V_{*i}))$ around $\hat{\mu}(X_i,V_{*i})=\mu(X_i,V_{*i})$, we can decompose

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{g}(\beta_*)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}g(\mu(X_i,V_{*i}),\beta)+M_1+M_2+o_p(1),$$

where

$$\begin{aligned} M_1 &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Delta_i(\hat{V}_i-V_{*i})+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\Delta}_{1i}(U_i-\hat{V}_i), \\ M_2 &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{g_1(\mu(X_i,V_{*i}),\beta_*)(\hat{\gamma}(X_i,V_{*i})-\mu(X_i,V_{*i}))+g_1(\hat{\gamma}(X_i,\hat{V}_i),\beta_*)\{Y_i-\hat{\gamma}(X_i,\hat{V}_i)\}\right\}. \end{aligned}$$

Suppose we have

$$M_1 = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Delta_i(U_i-V_{*i})+o_p(1), \tag{9}$$

$$M_2 = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}g_1(\mu(X_i,V_{*i}),\beta_*)\{Y_i-\mu(X_i,V_{*i})\}+o_p(1), \tag{10}$$

Then the central limit theorem implies the conclusion. For (9), by using the relation that $\hat{V}_i - V_{*i} = (U_i - V_{*i}) - (U_i - \hat{V}_i)$, we have

$$
\begin{aligned}
M_1 &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Delta_i(U_i - V_{*i}) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\hat{\Delta}_{1i} - \Delta_i)(U_i - \hat{V}_i) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Delta_i(U_i - V_{*i}) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\hat{\Delta}_{1i} - \Delta_i)(U_i - V_{*i}) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\hat{\Delta}_{1i} - \Delta_i)(\hat{V}_i - V_{*i}) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Delta_i(U_i - V_{*i}) + o_p(1).
\end{aligned}
$$

The last equality follows from the standard argument using degenerated U-statistics. Finally, (10) follows from the same argument as in Lemma A.4.

B.3. **Proof of Theorem 2.** We can show Theorem 2 by arguments that are similar to those which were used in the proof of Theorem 1, using Lemmas B.2-B.4. Therefore, we omit the details.

References

[1] Bravo, F., Escanciano, J. C. and I. van Keilegom (2015) Wilks' phenomenon in two-step semiparametric empirical likelihood inference, Working paper.

[2] Hahn, J. and G. Ridder (2013) Asymptotic variance of semiparametric estimators with generated regressors, *Econometrica*, 81, 315-340.

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

[4] Heckman, J. J., Ichimura, H. and P. Todd (1998) Matching as an econometric evaluation estimator, *Review of Economic Studies*, 65, 261-294.

[5] Hirano, K., Imbens, G. W. and G. Ridder (2003) Efficient estimation of average treatment effects using the estimated propensity score, *Econometrica*, 71, 1161-1189.

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

[7] Mykland, P. A. (1999) Bartlett identities and large deviations in likelihood theory, *Annals of Statistics*, 27, 1105-1117.

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

[9] Olley, G. S. and A. Pakes (1996) The dynamics of productivity in the telecommunications equipment industry, *Econometrica*, 64, 1263-1297.

[10] Owen, A. B. (1988) Empirical likelihood ratio confidence intervals for a single functional, *Biometrika*, 75, 237-249.

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

[12] Pagan, A. (1984) Econometric issues in the analysis of regressions with generated regressors, *International Economic Review*, 25, 221-247.

[13] 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.

Department of Industrial Engineering and Economics, Graduate School of Engineering, Tokyo Institute of Technology, 2-12-1, Ookayama, Meguro-ku, Tokyo 152-8550, Japan.

E-mail address: matsushita.y.ab@m.titech.ac.jp

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

E-mail address: t.otsu@lse.ac.uk