Econometric Theory, **38**, 2022, 621–688.
doi:10.1017/S0266466621000281

IDENTIFICATION AND ESTIMATION IN A CORRELATED RANDOM COEFFICIENTS TRANSFORMATION MODEL

ZHENGYU ZHANG
Shanghai University of Finance and Economics

ZEQUN JIN
Shanghai University of Finance and Economics

BEILI MU
Chinabond Pricing Center,
China Central Depository & Clearing Co., LTD.

This study examines identification and estimation in a correlated random coefficients (CRC) model with an unknown transformation of the dependent variable, namely $\lambda(Y^*) = B_0 + X'B$, where the latent outcome Y^* may be subject to a certain kind of censoring mechanism, $\lambda(\cdot)$ is an unknown, one-to-one monotone function, and the random coefficients (B_0, B) are allowed to be correlated with one or several components of X. Under a conditional median independence plus a conditional median zero restriction, the mean of B is shown to be identified up to scale. Moreover, we show the derivative of the median structural function (MSF) is point identified. This derivative of MSF resembles the marginal treatment effect introduced by Heckman and Vytlacil (2005, *Econometrica* 73, 669–738).

It generalizes the usual average treatment effect in a linear CRC model and coincides with $E(B)$ when λ is equal to the identity function; it is invariant to both location and scale normalization on the coefficients. We develop estimators for the identified parameters and derive asymptotic properties for the derivative of MSF. An empirical example using the U.K. Family Expenditure Survey is provided.

1. INTRODUCTION

This study examines a semiparametric transformation model with random coefficients that are allowed to be correlated with regressors. We call this a correlated random coefficients transformation model (CRCT model for short), which takes

We are grateful to the editor Peter C.B. Phillips, the co-editor Arthur Lewbel, and four anonymous referees for their insightful comments. This research is supported by the National Science Foundation of China (71873080), the Key Project of the National Science Foundation of China (71833004), the Fundamental Research Funds for the Central Universities (2021110070), and the 2018 Program for Innovative Research Team of the Shanghai University of Finance and Economics. Address correspondence to Zequn Jin, School of Economics, Shanghai University of Finance and Economics, No. 777, Guoding Road, Shanghai, P.R. China; e-mail: polaris@163.sufe.edu.cn.

© The Author(s), 2021. Published by Cambridge University Press.

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

the form:

$$\lambda\left(Y^*\right) = B_0(U) + X'B(U). \tag{1.1}$$

Here, Y^* is a one-dimensional (latent) dependent variable and $\lambda(\cdot)$ is an unknown smooth monotone function. X takes values in $\mathbb{R}^{k_X}$; (B_0, B) is a vector of random variables, which implicitly relies on multidimensional unobservables U. Moreover, some explanatory variables can be endogenous, in the sense that (B_0, B) may be correlated with one or several components of X. For this model, we will provide sufficient conditions under which the mean of the distribution of random coefficients and other economically meaningful parameters, such as the average derivative of the median structural function (MSF), can be identified. The average derivative of MSF (AD-MSF) measures the average marginal effect of the median potential outcome for people with observed characteristics and a given propensity to take up treatment. We also develop a $\sqrt{n}$-consistent and asymptotically normal estimator for the AD-MSF.

The first motivation for considering the CRCT model is that it substantially extends the traditional nonrandom coefficients transformation model, which has been of longstanding interest in the statistics and econometrics literature. With only a few exceptions (Lewbel and Pendakur, 2017), most prior studies focus on transformation models with nonrandom coefficients and an additively separable scalar error term.[1] Unlike the existing literature, our model allows for random coefficients and multidimensional unobservables entering nonadditively.

The second motivation is that equation (1.1) incorporates the linear correlated random coefficients (CRC) model as a special case by restricting $\lambda(\cdot)$ to the identity transformation, which means the CRCT model can be used to decrease the risk of model misspecification and thus help to interpret the model better. One well-known example is Box and Cox (1964), who propose a power transform of the dependent variable, which may lead to normality in a linear regression. Xu and Lee (2015) argue that taking transformation of the dependent variable through a nonlinear function is particularly useful in dealing with share data, which refers to samples with observed dependent variables whose values are between zero and one. In Section 6, we give an application to the estimation of a triangular simultaneous equations model for Engel curves. In this application, the dependent variable is the share of expenditure on a commodity, which takes values between

[1] The statistics and econometrics literature on transformation models is vast. In a fully exogenous setting, Horowitz (2001) and Jacho-Chávez, Lewbel, and Linton (2010) consider identification in a purely nonparametric transformation model. Others have analyzed semiparametric transformation models by either assuming a nonparametric form for $\lambda(\cdot)$ but with a linear regression function as in Horowitz (1996) and Moon (2013), or a parametric form for $\lambda(\cdot)$ as in Linton, Sperlich, and Van Keilegom (2008). For transformation models with endogenous regressors, Chiappori, Komunjer, and Kristensen (2015) consider a fully nonparametric setting. Florens and Sokullu (2017) and Feve and Florens (2010) consider a semiparametric form for the regression function. Vanhems and Van Keilegom (2018; hereafter VV) consider a parametric form for $\lambda(\cdot)$ but impose an additive structure on the regression function and estimate the model using a control function approach. In the econometrics literature, semiparametric transformation models have been studied by Asparouhova et al. (2002), Abrevaya (1999, 2003), Cavanagh and Sherman (1998), Chen (2002, 2010), Chen, Jin, and Ying (2002), Khan and Tamer (2007), Shin (2010), and Zhang (2016), to mention only a few. However, all these works focus on models with nonrandom coefficients and an additively separable scalar error term.

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

zero and one. Overall, by taking an unknown transformation of the dependent variable, the CRCT model provides a flexible and parsimonious way to extend the linear CRC model.[2]

Third, in addition to unobserved heterogeneity, censoring is another common phenomenon in microdata analysis. For example, income survey data are often censored because of top-coding and many economic variables, such as hours worked, wages, and expenditure shares are naturally bounded from below by zero. Indeed, in our application to the estimation of Engel curves, about 25% of households in the sample report zero expenditure on their leisure goods. However, the existing literature lacks formal studies concerning identification and estimation in a CRC model with censoring. In our analysis, we allow an observed outcome Y censored from the latent variable Y^* through a general two-sided fixed censoring mechanism, namely,

$$Y = \underline{c} \times 1\{Y^* < \underline{c}\} + Y^* \times 1\{\underline{c} \le Y^* \le \bar{c}\} + \bar{c} \times 1\{Y^* > \bar{c}\}, \tag{1.2}$$

with $-\infty < \underline{c} < \bar{c} < +\infty$.[3]

Our CRCT model allows one or several components of X to be correlated with U, that is, it allows observed covariates to interact with unobservable covariates. A typical example is that in the human capital earning model (Heckman and Vytlacil, 1998), the effect of years of schooling on wages after logarithmic transformation may vary from person to person because of the influence of unobserved ability or motivational factors. Moreover, these unmeasured factors may be positively correlated with the level of schooling. Another example is that in estimating Engel curves, total expenditure is endogenous for individual commodity demands; thus, we use gross earnings of the household head as an instrument. Let $\mathfrak{X}$ denote an endogenous component of X. We assume that there exists a vector of instruments Z for $\mathfrak{X}$ such that

$$\mathfrak{X} = g(Z, V), \tag{1.3}$$

where V is an unknown scalar random variable, $Z \perp (U, V)$, and $g(z, v)$ are unknown functions invertible in v for each possible value z. For clarity, we assume that Z always incorporates all exogenous components in X. These restrictions imposed on equation (1.3) are more general than those in Heckman and Vytlacil (1998) and Wooldridge (1997, 2003, 2008), which impose a homogeneous causal effect of Z on $\mathfrak{X}$. The same restrictions are maintained by Imbens and Newey (2009) for general nonseparable models, by Hoderlein and Sherman (2015; hereafter HS) for CRC binary response models, and by Masten and Torgovitsky (2016) for linear CRC models. Last but not least, because $\lambda(\cdot)$ is a one-to-one monotone function, equation (1.1) can be rewritten as

$$Y^* = \lambda^{-1}(B_0 + X'B).$$

[2]Linear CRC models have attracted considerable interest over the years, e.g., Heckman and Vytlacil (1998), Wooldridge (1997, 2003, 2008), and Masten and Torgovitsky (2016).

[3]Equation (1.2) incorporates the case of no censoring by artificially setting $\bar{c}$ sufficiently large and $\underline{c}$ sufficiently small.

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

In view of this, this study considers identification in a nonseparable model with a linear index structure restriction.

Our general idea for identification combines Chen's (2010) discretizing approach with the nonrandom coefficients transformation model and HS localize-then-average approach for the CRC binary response model. Under a conditional median restriction that generalizes Manski (1975, 1985), we show that the mean of (B_0, B) is identified up to scale. The same assumption was used by HS to identify the mean of coefficients (up to scale) in a CRC binary response model. As in HS, our identifying assumption is separated into a conditional median independence (CMI) condition and a conditional median zero (CMZ) condition. The CMI condition holds whenever X is independent of U given V, which is implied by equation (1.3) and $Z \perp (U, V)$. The CMZ condition holds nontrivially, but it is satisfied under a wide range of sufficient conditions. For example, the CMZ condition is implied by symmetry conditions on the joint distribution of (B_0, B, V). Moreover, there are two important special cases in which the CMZ condition is automatically satisfied, and no symmetry conditions are required. The first is the traditional nonrandom coefficients transformation model with endogenous regressors. The second is the nonrandom coefficients transformation model with exogenous regressors, which has been treated by many authors cited above. A more detailed statement of the sufficient conditions for the CMI and CMZ conditions to hold is provided in Section 2.

It is worth noting that attaining identification of $\mathbb{E}(B_0, B)$ up to scale means that we can only learn about the expected effect of X on the transformed dependent variable, relative to the effect of a given component in X whose coefficient is normalized to unity. In this study, in addition to $\mathbb{E}(B_0, B)$, a more readily interpretable parameter, namely the derivative of the MSF of the latent outcome, is shown to be point identified. This derivative of the MSF resembles the marginal treatment effect (MTE) introduced by Heckman and Vytlacil (2005), and has two major merits relative to $\mathbb{E}(B_0, B)$. First, it generalizes the usual average treatment effect focused on by the literature on linear CRC models (Wooldridge, 1997, 2003, 2008; Masten and Torgovitsky, 2016). When the transformation function $\lambda(\cdot)$ is confined to the identity function, the AD-MSF coincides with *EB*. Second, unlike the mean of coefficients that can only be identified up to scale, the derivative of the MSF can be point identified and is invariant to both location and scale normalization on the coefficients. We also propose a novel and computationally convenient estimator for the derivative of the MSF, which obviates the need for nonparametric estimation of the inverse and the derivative of $\lambda(\cdot)$.

The central idea of this study is closely related to two papers. The first is HS, who study a CRC binary response model with CMI and CMZ conditions. Nevertheless, the CRC binary response model is fundamentally different from the transformation model considered here. Moreover, HS aims to identify the mean of the coefficients only up to scale, whereas we focus on identification of the derivative of the MSF. The second is Zhang (2016), who studies a partially linear transformation model, motivated by Chen's (2010) discretizing approach. There are three main

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

differences between Zhang (2016) and this study. First, Zhang studies quantile regression estimation in a partially linear transformation model, whereas this study utilizes a generalized conditional median restriction to identify the derivative of the MSF in a random coefficient transformation model. Second, Zhang focuses on the exogeneity case, whereas we explicitly account for endogeneity and use the control function approach to handle it. The control function is nonparametrically specified in this study, thus complicating the asymptotic analysis a great deal. Third, Zhang's model is additively separable, whereas this study considers a random coefficients model with nonseparable unobservables. Consequently, the identification, estimation, and asymptotic analysis are all more complex than Zhang (2016).

Lewbel and Pendakur (2017) consider nonparametric identification in a generalized random coefficients model. Their study differs from ours in three aspects. First, their main identification result imposes an additive structure on the model, namely,

$$Y = B_0 + \sum_{l=1}^{k_X} \lambda_l (X_l B_l), \tag{1.4}$$

with each $\lambda_l(\cdot)$, $l = 1, \ldots, k_X$, an unknown strictly monotonically increasing function. Obviously, neither equation (1.1) nor (1.4) includes the other as a special case. Second, Lewbel and Pendakur focus on identifying the joint distribution of $(B_0, B_1, \ldots, B_{k_X})$ as well as $(\lambda_1(\cdot), \ldots, \lambda_{k_X}(\cdot))$, whereas we focus on identifying the derivative of a certain structural function of the outcome. Third, the main identifying assumptions in Lewbel and Pendakur include the additive structure and certain conditional independence between random coefficients and regressors, whereas we introduce the CMI and CMZ conditions. Moreover, our model can accommodate a certain kind of censoring in the latent outcome, whereas Lewbel and Pendakur's cannot.

The remainder of the paper is structured as follows. In Section 2, we introduce the key identification assumptions, discuss their implications, and introduce the parameters of interest that can be identified under these assumptions. The identification strategy is then translated into a multistage estimation procedure described in Section 3. Large sample properties of the proposed estimators are provided in Section 4. Section 5 reports some Monte Carlo results. An empirical example is presented in Section 6. Section 7 concludes. All technical details are collected in the Appendix.

2. IDENTIFICATION

This section is structured as follows. In Section 2.1, we introduce a CMI and a CMZ assumption, then discuss their implications. Moreover, we provide a detailed description of the sufficient conditions for these assumptions to hold in practice. Section 2.2 shows that the mean of the distribution of the random coefficients

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

is identified up to scale. Section 2.3 is the main part of Section 2. It introduces another economically meaningful parameter, namely the derivative of an MSF of the potential outcome. We discuss its economic implication, and show it can be point identified under CMI and CMZ conditions. Other meaningful parameters that might be possibly identified under the maintained assumptions are also discussed.

2.1. The CMI and CMZ Assumptions

Let $(\beta_0(v), \beta(v))$ be a measure of center of the distribution of (B_0, B) given $V = v$. In the spirit of HS, we call $(\beta_0(v), \beta(v))$ the "localizing function."[4]

The model can be written as

$$\lambda\left(Y^*\right) = \beta_0(v) + X'\beta(v) - \underbrace{\left(\beta_0(v) - B_0(U) + X'\beta(v) - X'B(U)\right)}_{\varepsilon(v,U,X)}, \tag{2.1.1}$$

where $\varepsilon = \varepsilon(v, U, X) = \beta_0(v) - B_0(U) + X'(\beta(v) - B(U))$ is the generalized error term.

Let S_V and S_X denote the support of X and V, respectively. In the spirit of HS,

we introduce the following two assumptions, namely the CMI and CMZ assumptions:

$$\text{CMI.} \quad \text{Med}\ (\varepsilon | X = x, V = v) = \text{Med}\ (\varepsilon | V = v), \text{ for any } x \in S_X \text{ and } v \in S_V. \tag{2.1.2}$$

$$\text{CMZ.} \quad \text{Med}\ (\varepsilon | V = v) = 0, \text{ for any } x \in S_X \text{ and } v \in S_V. \tag{2.1.3}$$

Both (2.1.1) and (2.1.2) imply that

$$\text{Med}\left(\varepsilon \middle| V = v, X = x\right) = 0, \text{ for any } x \in S_X \text{ and } v \in S_V. \tag{2.1.4}$$

As pointed out by HS, both CMI and CMZ conditions are nontrivial, but satisfied under a range of sufficient conditions. In Remarks 2.1.1–2.1.4 below, we summarize several important sufficient conditions particularly relevant to our model. For a more detailed discussion of these sufficient conditions, see HS (Sect. 2, Classes 1–4, pp. 138–139).

Remark 2.1.1. Clearly, the CMI condition holds whenever $U \perp X \middle| V$, where V is a control variable to break the dependence between X and U. Assume that V is a continuously distributed scalar with cdf.

$F_V(\cdot)$, which is strictly increasing on its support. As demonstrated by Imbens and Newey (2009), $F_V(V)$ can be identified from the data as $F_V(V) = F_{\mathfrak{X}|Z}\left(\mathfrak{X} \middle| Z\right)$, provided that (1.3) holds and $Z \perp (U, V)$. This approach to constructing control

[4] In this paper, we consider cases in which conditional mean $(\beta_0(v), \beta(v)) = (E(B_0 | v = v), E(B | V = v))$ and conditional median $(\beta_0(v), \beta(v)) = (Med(U|V), \beta)$ are used.

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

variables has also been taken by Florens et al. (2008), Masten and Torgovitsky (2016), and HS.

For Remarks 2.1.2 and 2.1.3, the localizing function $\beta(v)$ equals $E(B|V=v)$.

Remark 2.1.2. Assume that CMI holds, then the CMZ condition is implied by symmetry conditions on the distribution of (B_0, B, V), which guarantee that certain conditional means and medians coincide. These symmetry conditions are nontrivial, but are satisfied for a wide range of possible (B_0, B, V) distributions. For example, CMZ holds when the joint distribution of (B_0, B, V) is an elliptically symmetric random vector. Elliptically symmetric distributions form a large class encompassing, for example, multivariate normal distributions, normal variance mixture distributions, multivariate Laplace distributions, multivariate uniform distributions, multivariate t-distributions with degrees of freedom greater than two, multivariate logistic distributions, and all symmetric stable distributions with finite mean vectors and variance matrices.

Remark 2.1.3. Sometimes joint symmetry on the distribution of (B_0, B, V) might be too strong. Alternative sufficient conditions exist for both CMZ and CMI to hold with no joint symmetry. For example, both CMZ and CMI hold when random coefficients are decomposed into a systematic component correlated with V plus a symmetric idiosyncratic error term, namely,

$$B_l = \theta_l(V) + e_l, \quad l = 0, 1, \ldots, k_X,$$

where B_l, $l = 1, \ldots, k_X$, is the lth component of B, each $\theta_l(\cdot)$ is a real-valued function, and each e_l is symmetric and independent of (X, V). The independence and symmetry assumptions on e_l are strong, but the $\theta_l(V)$ component, through which X is correlated with (B_0, B), is unrestricted. In this case, X is correlated with (B_0, B) through $\theta_l(V)$, but X is also independent of (B_0, B) given V, so CMI holds. Moreover, CMZ follows directly from the symmetry of e_l.

Remark 2.1.4. In particular, for a nonrandom coefficients transformation model with an endogenous regressor, the CMZ condition is automatically satisfied, with no symmetry conditions required. To understand this, consider the model

$$\lambda\left(Y^*\right) = \beta_1 X_1 + X_2\beta_2 + U, \tag{2.1.5}$$

where the intercept term has been absorbed into $\lambda(\cdot)$. For simplicity, we assume that X_1 is the single endogenous regressor generated by $X_1 = g(Z, V)$ and V is the control variable such that

$$\text{Med}\left(U\middle|X, V\right) = \text{Med}\left(U\middle|V\right). \tag{2.1.6}$$

Equation (2.1.6) can be satisfied if $g(z, v)$ is invertible in v and $Z = (Z_1, X_2)$ is independent of (U, V). Equation (2.1.5) can be rewritten as

$$\lambda\left(Y^*\right) = \beta_1 X_1 + X_2\beta_2 + \underbrace{\text{Med}\left(U\middle|V=v\right)}_{\beta_0(v)} + \underbrace{\left(U - \text{Med}\left(U\middle|V=v\right)\right)}_{\varepsilon(v,U,X)}.$$

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

Then, (2.1.4) is automatically satisfied, because

$$\begin{aligned}&\text{Med}\left(\varepsilon(v,U,X)\middle|X=x,V=v\right)=\text{Med}\left(U-\text{Med}\left(U\middle|V=v\right)\middle|X=x,V=v\right)\\&=\text{Med}\left(U\middle|X=x,V=v\right)-\text{Med}\left(U\middle|V=v\right)=0.\end{aligned}$$

Moreover, in this case, $\beta(v)\equiv(\beta_1,\beta_2)$ and $\beta_0(v)=\text{Med}\left(U\middle|V=v\right)$. ■

2.2. Identification of $\mathbb{E}(B)$

From now on, we always take the localizing function $\beta(v)$ as $\mathbb{E}(B|V=v)$. First, consider the identification of $\mathbb{E}B$. Our general idea is based on Chen's (2010) discretizing approach that converts the transformation model into a sequence of binary choice models. Specifically, for a given cutoff point, $c\in\left(\underline{c},\bar{c}\right)$, define $D(c)=1\{Y>c\}$. Because $\lambda(\cdot)$ is a strictly increasing function, we have

$$\begin{aligned}D(c)&=1\left\{Y^*>c\right\}=1\left\{\lambda\left(Y^*\right)>\lambda(c)\right\}\\&=1\left\{B_0+X'B>\lambda(c)\right\}=1\left\{\beta_0(v)+X'\beta(v)-\varepsilon>\lambda(c)\right\},\end{aligned}$$

where the last equality follows from (2.1.1). Let $X=\left(X_1,\widetilde{X}'\right)'$, where X_1 denotes the first component of X, $\widetilde{X}$ denotes the remaining components of X except X_1, and $B=\left(B_1,B_2'\right)'$. As discussed in Horowitz (1996), for models subject to unknown transformation, the distribution of the coefficients can only be identified up to scale. Without loss of generality, we assume $B_1>0$ almost surely, which implies that $\beta_1(v)>0$, for any $v\in S_V$. We have

$$\begin{aligned}D(c)&=1\left\{\beta_0(v)+X'\beta(v)-\varepsilon>\lambda(c)\right\}\\&=1\left\{\beta_0(v)-\lambda(c)+X_1\beta_1(v)+\widetilde{X}'\beta_2(v)>\varepsilon\right\}\\&=1\left\{\frac{\beta_0(v)-\lambda(c)}{\beta_1(v)}+X_1+\widetilde{X}'\frac{\beta_2(v)}{\beta_1(v)}>\frac{\varepsilon}{\beta_1(v)}\right\}\\&=1\left\{X_1+\widetilde{X}'\widetilde{\beta}_2(v)+\widetilde{\lambda}(c,v)>\widetilde{\varepsilon}(v)\right\},\end{aligned}\tag{2.2.1}$$

where $\widetilde{\beta}_2(v)=\dfrac{\beta_2(v)}{\beta_1(v)}$, $\widetilde{\lambda}(c,v)=\dfrac{\beta_0(v)-\lambda(c)}{\beta_1(v)}$, and $\widetilde{\varepsilon}(v)=\dfrac{\varepsilon}{\beta_1(v)}$. Under (2.1.4),

$$\text{Med}\left(\widetilde{\varepsilon}(v)\middle|V=v,X=x\right)=\frac{1}{\beta_1(v)}\text{Med}\left(\varepsilon\middle|V=v,X=x\right)=0.\tag{2.2.2}$$

This zero conditional median restriction on the generalized error term $\widetilde{\varepsilon}(v)$ means that we can identify $\widetilde{\beta}_2(v)$ and $\widetilde{\lambda}(c,v)$ by combining Manski (1985) and Horowitz's (1992) maximum score method and some localizing technique. We formalize the above reasoning in the following assumptions and Lemma 2.1.

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

Assumption 2.1. (i) The observed variables consist of (Y, X, Z), which are generated by (1.1)–(1.3). (ii) $\lambda(\cdot)$ is a strictly increasing function.

Assumption 2.2. (i) $(\beta_0(v), \beta_1(v), \beta_2(v)) = \mathbb{E}\left((B_0, B_1, B_2)\middle|V = v\right)$. (ii) $\beta_1(v) > 0$, for each $v \in S_V$.

Assumption 2.3. Let $\varepsilon = \beta_0(v) - B_0 + X'(\beta(v) - B)$. $\text{Med}(\varepsilon|X = x, V = v) = \text{Med}(\varepsilon|V = v)$ and $\text{Med}(\varepsilon|V = v) = 0$, for every $v \in S_V$ and $x \in S_X$.

Assumption 2.4. (i) For each $c \in (\underline{c}, \overline{c})$, $0 < P\left(Y^* > c\middle|X = x, V = v\right) < 1$, for every $v \in S_V$, and $x \in S_X$. (ii) The distribution of X_1 conditional on $\widetilde{X} = \widetilde{x}$ and $V = v$ has an everywhere positive density with respect to Lebesgue measure. (iii) The support of X given $V = v$ is not contained in any proper linear subspace of $\mathbb{R}^{k_X}$.

Remark 2.2.1. Assumption 2.1 describes the data generating process (DGP). Assumption 2.2 clarifies the localizing function. $\beta_1(v) > 0$ facilitates scale normalization. Assumption 2.3 is the key identifying condition on the model. Assumption 2.4 is adapted from Manski (1985) and Horowitz (1992) for a binary response model. For example, see Horowitz (1992, Assumption 2, p. 509).

LEMMA 2.1. *If Assumptions 2.1–2.4 hold, then* $\widetilde{\beta}_2(v) = \beta_2(v)/\beta_1(v)$ *is identified for each* $v \in S_V$. *Moreover,*

$$\widetilde{\lambda}(c, v) = \frac{\beta_0(v) - \lambda(c)}{\beta_1(v)} \tag{2.2.3}$$

is also identified for any $c \in (\underline{c}, \overline{c})$ *and* $v \in S_V$.

The proof of Lemma 2.1 can be found in the Appendix. Specifically, the proof for the first part, i.e., the identification of $\widetilde{\beta}_2(v)$, can be found in Lemma A.4. Then, this result together with Lemma A.2 can achieve the identification of $\widetilde{\lambda}(c, v)$.

Remark 2.2.2. Let $\widetilde{B}_2 = B_2/B_1$. Because the distribution of $F_V(V)$ is identified (Remark 2.1.1), $\mathbb{E}\widetilde{B}_2 = \mathbb{E}_V\left(\widetilde{\beta}_2(V)\right)$ is also identified.

2.3. Average Derivative of Median Structural Function

There are several drawbacks, associated with the already identified $\mathbb{E}(B_2/B_1)$. First, it means that the average partial effect of X can only be measured up to scale. $\mathbb{E}(B_2/B_1)$ is not scale invariant, in the sense that it has to be interpreted differently if the coefficient on the regressor other than X_1 is normalized to one. Second, it merely tells us the average effect on Y after transformation, not the direct effect on Y. Thus, an interesting question is whether it is possible to identify a parameter which can summarize some central impact of X on the latent dependent variable Y^* while remedying the above drawbacks at the same time. To overcome these disadvantages, we will focus on the MSF of the latent outcome variable evaluated

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

at $X = x$ and conditional on $V = v$:

$$\mathrm{MSF}(x, v) = \mathrm{Med}\left(\lambda^{-1}\left(B_0 + x'B\right) \middle| V = v\right), \tag{2.3.1}$$

which is the median potential outcome for people with observed characteristics $X = x$ and a given propensity $V = v$ to take up treatment. If we are looking at the increase in the median potential outcomes as one increases x a little bit, but only for people with $X = x$ and $V = v$, we can work with the derivative of $\mathrm{MSF}(x, v)$ with respect to x, namely,

$$\mathrm{MSF}'(x, v) = \frac{\partial \mathrm{MSF}(x, v)}{\partial x}. \tag{2.3.2}$$

Furthermore, we can investigate the following parameter:

$$\vartheta = \mathbb{E}\left[\mathrm{MSF}'(X, V)\right],$$

called the AD-MSF. Identification of this parameter builds on several preliminary objects, such as $\mathrm{MSF}'(x, v)$ in Lemma 2.2.

Assume that X_1 is a continuous endogenous treatment variable, such as tax rate, years of schooling, or total expenditure. Then, $\mathrm{MSF}'_1(x, v)$, the first component $\mathrm{MSF}'(x, v)$, can be interpreted as the marginal median treatment effect, resembling the MTE introduced by Heckman and Vytlacil (2005, 2007). It is the median gain of the outcome for people who are just indifferent to the receipt of treatment level $X_1 = x_1$ at the given value of the unobservable $V = v$. $\mathrm{MSF}'_1(x, v) \times \Delta$ then is a willingness to pay measure that informs us how much an individual at the margin of receiving $X_1 = x_1$ would be willing to pay, in order to move from x_1 to $x_1 + \Delta$.

To give a concrete example, let Y be the wage, X_1 be the years of schooling, and Z_1 be some cost shifter of schooling, such as the distance to college. In the sense of Chesher (2003), the first-stage error term V can be interpreted as the ability and the outcome disturbance U can be multidimensional, for example, $U = (V, \epsilon)$, where V is the ability and ϵ captures fortune in the labor market. Our $\mathrm{MSF}'(x, v)$ provides a measure of the median gain of receiving more than 1 year of schooling for individuals having attended x_1 years of schooling and with the rank of ability $V = v \in (0, 1)$. As another example, let Y denote output of a firm and X_1 be the log labor demand of this firm. $\lambda(\cdot)$ takes a logarithm function if firms produce according to a Cobb–Douglas production function. Because labor inputs are chosen by the firm to maximize profits, X_1 is likely endogenous. Let Z_1 be some cost shifter of labor, such as log wage. The first-stage error V can be interpreted as unobserved economic shocks, which affect both demand for labor and output of a firm. Then, $\mathrm{MSF}'(x, v)$ provides a measure of the median change of the elasticity of output with respect to labor for firms with labor inputs $X_1 = x$ and with positive $(v > 0)$ or negative $(v < 0)$ economic shock.

LEMMA 2.2. *Under the assumptions of Lemma 2.1, if* $\lambda(\cdot)$ *is continuously differentiable,* $\mathrm{MSF}'(x, v)$ *is identified for any* $x \in S_X$ *and* $v \in S_V$.

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

Proof. By definition,

$$\begin{aligned}\mathrm{MSF}'(x,v) &= \frac{\partial \operatorname{Med}\left(\lambda^{-1}\left(B_0 + x'B\right)\middle| V=v\right)}{\partial x} = \frac{\partial \lambda^{-1}\left(\operatorname{Med}\left(B_0 + x'B\middle| V=v\right)\right)}{\partial x} \\ &= \frac{\partial \lambda^{-1}\left(\operatorname{Med}\left(\beta_0(v) + x'\beta(v) + \varepsilon\middle| V=v\right)\right)}{\partial x} = \frac{\partial \lambda^{-1}\left(\beta_0(v) + x'\beta(v)\right)}{\partial x} \\ &= \frac{\beta(v)}{\lambda'\left(\lambda^{-1}\left(\beta_0(v) + x'\beta(v)\right)\right)}. \end{aligned} \tag{2.3.3}$$

The first equality is by definition, the second by strict monotonicity of $\lambda(\cdot)$, the third by $\varepsilon = B_0 - \beta_0(v) + x'(B - \beta(v))$, the fourth by (2.1.3), and the last by the differentiability of $\lambda(\cdot)$.

Because $\beta_1(v) > 0$, we can denote $\Lambda(c,v) = \dfrac{\lambda(c)}{\beta_1(v)}$. Let $\Lambda^{-1}(c,v)$ ($\Lambda'(c,v)$, resp.) be the inverse (derivative, resp.) of $\Lambda(c,v)$ with respect to its first argument. Then, $\Lambda'(c,v) = \dfrac{\lambda'(c)}{\beta_1(v)}$ and $\Lambda^{-1}(c,v) = \lambda^{-1}(\beta_1(v)\cdot c)$, because

$$\Lambda\left(\Lambda^{-1}(c,v), v\right) = \frac{\lambda\left(\lambda^{-1}(\beta_1(v)\cdot c)\right)}{\beta_1(v)} = \frac{\beta_1(v)\cdot c}{\beta_1(v)} = c.$$

Dividing both the numerator and denominator in (2.3.4) by $\beta_1(v)$ gives

$$\begin{aligned}\mathrm{MSF}'(x,v) &= \frac{\beta(v)/\beta_1(v)}{\lambda'\left(\lambda^{-1}\left(\beta_0(v) + x'\beta(v)\right)\right)\Big/\beta_1(v)} \\ &=^{\dagger_1} \frac{\left(1, \widetilde{\beta}_2'(v)\right)'}{\Lambda'\left(\Lambda^{-1}\left(\widetilde{\beta}_0(v) + x_1 + \widetilde{x}'\widetilde{\beta}_2(v), v\right), v\right)}, \end{aligned} \tag{2.3.4}$$

where $x = \left(x_1, \widetilde{x}'\right)'$, $\beta(v) = (\beta_1(v), \beta_2(v))$, $\widetilde{\beta}_2(v) = \dfrac{\beta_2(v)}{\beta_1(v)}$, and $\widetilde{\beta}_0(v) = \dfrac{\beta_0(v)}{\beta_1(v)}$. The last equality $\dagger_1$ follows from the definition of $\Lambda^{-1}(c,v)$ and $\Lambda'(c,v)$. Because $\widetilde{\beta}_2(v)$ is already identified by Lemma 2.1, to complete the proof, it suffices to show that both $\Lambda(c,v)$ and $\widetilde{\beta}_0(v)$ are identifiable. Note that $\widetilde{\lambda}(c,v)$ is identified by Lemma 2.1, and it can be rewritten as

$$\widetilde{\lambda}(c,v) = \widetilde{\beta}_0(v) - \Lambda(c,v). \tag{2.3.5}$$

Below we show that $\Lambda(c,v)$ and $\widetilde{\beta}_0(v)$ are identifiable from (2.3.5) by the marginal integration method. Assume that $\omega(c)$ is any nonnegative real-valued function, satisfying

$$\int_{\underline{c}}^{\overline{c}} \omega(c)dc = 1.$$

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

For example, $\omega(c) = 1\{\underline{c} < c \leq \overline{c}\} \big/ \left(\overline{c} - \underline{c}\right)$. Because (1.1) is observationally equivalent to

$$\lambda\left(Y^*\right) - a = (B_0 - a) + X'B,$$

for any $a \in \mathbb{R}$, we can always assume without loss of generality that

$$\int_{\underline{c}}^{\overline{c}} \lambda(c)\omega(c)dc = 0. \tag{2.3.6}$$

Equation (2.3.6) actually places a location normalization restriction on $\lambda(\cdot)$. Then, $\widetilde{\beta}_0(v)$ is identified, because

$$\begin{aligned} &\int_{\underline{c}}^{\overline{c}} \widetilde{\lambda}(c,v)\,\omega(c)\,dc = \int_{\underline{c}}^{\overline{c}} \left(\widetilde{\beta}_0(v) - \Lambda(c,v)\right)\omega(c)\,dc \\ &\quad = \widetilde{\beta}_0(v)\int_{\underline{c}}^{\overline{c}} \omega(c)\,dc - \frac{\int_{\underline{c}}^{\overline{c}} \lambda(c)\omega(c)dc}{\beta_1(v)} = \widetilde{\beta}_0(v). \end{aligned} \tag{2.3.7}$$

Finally, $\Lambda(c,v)$ is identified, because

$$\Lambda(c,v) = \widetilde{\beta}_0(v) - \widetilde{\lambda}(c,v).$$

■

The reason why we focus on the derivative of MSF is that it has several advantages relative to $\mathbb{E}(B_2/B_1)$. First, it provides a reasonable measure of center of the distributional effect of X on Y^*. Second, when $\lambda(\cdot)$ is the identify function, namely, (1.1) is a linear CRC model, it follows from (2.3.3) that

$$\mathrm{MSF}'(x,v) = \beta(v), \quad \mathbb{E}\left[\,\mathrm{MSF}'(X,V)\right] = \mathbb{E}B.$$

In view of this, the AD-MSF generalizes the usual average treatment effect parameter in a linear CRC model. Third, unlike $\mathbb{E}(B_2/B_1)$, $\mathrm{MSF}'(x,v)$ is location and scale invariant in the sense that its interpretation does not rely on which component of B has been normalized to one.

Remark 2.3.1. Alternatively, another parameter of interest one may consider is $\mathrm{MSF}(x)$,

$$\mathrm{MSF}(x) = \mathrm{Med}\left(\lambda^{-1}\left(B_0 + x'B\right)\right),$$

which is the median potential outcome for people with observed characteristics $X = x$, and its derivative $\mathrm{MSF}'(x)$,

$$\mathrm{MSF}'(x) = \frac{\partial \mathrm{MSF}(x)}{\partial x} = \frac{\partial \mathrm{Med}\left(\lambda^{-1}\left(B_0 + x'B\right)\right)}{\partial x},$$

which measures the median potential outcome change with a small increase in x. However, $\mathrm{MSF}'(x)$ generally cannot be identified by the CMZ and CMI conditions as maintained in this paper. To understand this, by reasoning similar to (2.3.3),

$$\mathrm{MSF}'(x) = \frac{\partial \lambda^{-1}\left(\,\mathrm{Med}\left(B_0 + x'B\right)\right)}{\partial x} = \frac{\partial \lambda^{-1}\left(\,\mathrm{Med}\left(\mathbb{E}(B_0) + x'\mathbb{E}(B) + \varepsilon(x)\right)\right)}{\partial x},$$

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

where the first equality again follows from strict monotonicity of λ and the second by $\varepsilon(x) = B_0 - \mathbb{E}(B_0) + x'(B - \mathbb{E}(B))$. To proceed, assume that $\text{Med}(\varepsilon(x)) = 0$, for each $x \in S_X$, which can be satisfied, for example, if one imposes symmetry conditions on the distribution of (B_0, B), as discussed in Remark 2.1.2. Then,

$$\text{MSF}'(x) = \frac{\partial \lambda^{-1}\left(\mathbb{E}(B_0) + x'\mathbb{E}(B)\right)}{\partial x} = \frac{\mathbb{E}(B)}{\lambda'\left(\lambda^{-1}(\mathbb{E}(B_0) + x'\mathbb{E}(B))\right)}. \tag{2.3.8}$$

As discussed in Horowitz (1996), for models subject to unknown transformation, the distribution of the coefficients as well as the link function $\lambda(\cdot)$ can only be identified up to scale. Thus, (2.3.8) is likely identified if the mean of B can be, at least, identified up to scale. But according to Lemma 2.1, we can only identify quantities like $\mathbb{E}(B_2/B_1)$, which does not equal to $\mathbb{E}B_2/\mathbb{E}B_1$ in general.

Remark 2.3.2. There are several cases in which $\text{MSF}'(x)$ is equal to $\text{MSF}'(x, v)$, and thus identified. In the first situation, the slope coefficient B is nonrandom, which can be seen from Remark 2.1.4. Second, if B is independent of V, or X is exogenous, $\mathbb{E}(B|V = v) \equiv \mathbb{E}B$, for any $v \in S_V$. Third, when $\lambda(y) = ay + b$, including the linear CRC model as a special case, the two parameters coincide.

It follows from the proof of Lemma 2.2 that $\Lambda(c, v) = \lambda(c)/\beta_1(v)$ is point identified for any $c \in (\underline{c}, \overline{c})$ and $v \in S_V$. One corollary is that $\lambda(\cdot)$ can be identified up to scale if at least one component of (B_0, B), say, B_1 is a nonzero constant.

By localizing on $X = x$ and $V = v$, the estimator of $\text{MSF}'(x, v)$ suffers from a curse of dimensionality. To overcome the curse, it is convenient to focus on the average of $\text{MSF}'(x, v)$ over the joint distribution of (X, V).

COROLLARY 2.3. *Under the assumptions of Lemma 2.2, the AD-MSF, namely,*

$$\vartheta = \mathbb{E}\left[\text{MSF}'(X, V)\right], \tag{2.3.9}$$

is identified.

3. ESTIMATION

3.1. The Control Variable

Throughout this paper, we will limit our focus to the case where there is only one endogenous variable. The case for more than one endogenous regressor is similar. Let $\widetilde{V}_i = F_{\mathfrak{X}|Z}(\mathfrak{X}_i, Z_i)$, $i = 1, \ldots, n$, where $F_{\mathfrak{X}|Z}(\cdot, z)$ is the cdf of $\mathfrak{X}$ conditional on $Z = z$. We estimate $\widetilde{V}_i$ by

$$\widehat{\widetilde{V}}_i = \widehat{F}_{\mathfrak{X}|Z}(\mathfrak{X}_i, Z_i),$$

where $\widehat{F}_{\mathfrak{X}|Z}(\cdot, \cdot)$ is the Nadaraya–Watson estimator of $F_{\mathfrak{X}|Z}(\cdot, \cdot)$. Let $Z = \left(Z^c, Z^d\right)$ be the $k_Z \times 1$ instrument vector, where Z^c is the k_{Z^c}-dimensional continuously distributed instrument, and Z^d is the remaining k_{Z^d}-dimensional discretely distributed instrument. To simplify the notation, we assume that there is only one continuous

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

instrument in Z, that is, $k_{Z^c} = 1$.[5] Then, the Nadaraya–Watson estimator of $\widetilde{V}_i$ can be written as

$$\widehat{\widetilde{V}}_i = \frac{\sum_{j\neq i} 1\left\{\mathfrak{X}_j \leq \mathfrak{X}_i\right\} K_0\left(\frac{Z_j^c - Z_i^c}{h_0}\right) 1\left\{Z_j^d = Z_i^d\right\}}{\sum_{j\neq i} K_0\left(\frac{Z_j^c - Z_i^c}{h_0}\right) 1\left\{Z_j^d = Z_i^d\right\}},$$

where $K_0(\cdot)$ is a univariate kernel function and h_0 is the corresponding bandwidth converging to zero as the sample size increases to infinity.

3.2. Estimation of $\widetilde{\beta}_2(v)$

For any $v \in S_{\widetilde{V}} \subset (0,1)$, we estimate $\widetilde{\beta}_2(v) = \beta_2(v)/\beta_1(v)$ by the following two steps. First, for any given $c \in (\underline{c}, \overline{c})$, we employ a local version of the smoothed maximum score approach originally proposed by Manski (1985) and Horowitz (1992) to estimate $\widetilde{\beta}_2(v)$ by $\widehat{\widetilde{\beta}}_2(c,v)$ in a sequence of binary response models defined by (2.2.1). Specifically, $\left(\widehat{\widetilde{\beta}}_2(c,v), \widehat{\widetilde{\lambda}}(c,v)\right)$ maximizes

$$\widehat{S}_n(b,a|c,v) = \frac{1}{nh_2}\sum_{i=1}^{n}\left(2D_i(c)-1\right) G_1\left(\frac{X_{1i} + \widetilde{X}_i' b + a}{h_1}\right) K_2\left(\frac{\widehat{\widetilde{V}}_i - v}{h_2}\right) t_\kappa\left(Z_i^c\right) \tag{3.2.1}$$

with respect to (a,b) for any given (c,v). In $\widehat{S}_n(b,a|c,v)$,

$$G_1(s) = \int_{-\infty}^{s} K_1(v)dv$$

with $K_1(\cdot)$ a density kernel function and $G_1\left(\frac{X_{1i}+\widetilde{X}_i'b+a}{h_1}\right)$ aims to mimic the behavior of $1\left\{X_{1i} + \widetilde{X}_i' b + a > 0\right\}$ with h_1 decreasing to zero as $n \to \infty$. $t_\kappa(z) = 1\left\{|z| \leq \kappa\right\}$ is a trimming device. If Z_i has binary or discrete support, $t_\kappa(Z_i^c) \equiv 1$. As explained in HS, trimming the continuous components of Z_i prevents ratio bias when the denominator of $\widehat{\widetilde{V}}_i$ is too close to zero.

Remark 3.2.1. The rationale behind the above local maximum score estimator is that with a sufficiently small h_2, only the $\widehat{\widetilde{V}}_i$'s lying within a small neighborhood around v contribute to the objective function $\widehat{S}_n(b,a|c,v)$. As a result, the maximizer $\left(\widehat{\widetilde{\beta}}_2(c,v), \widehat{\widetilde{\lambda}}(c,v)\right)$ will be a consistent estimator of $\left(\widetilde{\beta}_2(v), \widetilde{\lambda}(c,v)\right)$.

However, due to localizing on (c,v), $\widehat{\widetilde{\beta}}_2(c,v)$ converges to $\widetilde{\beta}_2(v)$ at a slower rate. To improve its convergence rate, we can follow Chen's (2010) integration method,

[5] Here, we describe the estimator with only one continuous instrument, which is consistent with the application provided in Section 6. The only one continuous instrument there is the log gross earnings.

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

that is, $\widehat{\widetilde{\beta}}_2(v)$ is defined as the maximizer of $\widehat{S}_n(b|v) =$

$$\frac{1}{nh_2}\sum_{i=1}^{n}\int_{\underline{c}}^{\bar{c}}\left(2D_i(c)-1\right)G_1\left(\frac{X_{1i}+\widetilde{X}_i'b+\widehat{\widetilde{\lambda}}(c,v,b)}{h_1}\right)K_2\left(\frac{\widehat{\widetilde{V}}_i-v}{h_2}\right)t_\kappa\left(Z_i^c\right)\varpi(c)\,dc$$

with respect to b, for any given v, where $\varpi(c)$ is a proper integrable weight function and $\widehat{\widetilde{\lambda}}(c,v,b)$ maximizes (3.2.1) with respect to a, for any given (c,v,b).

3.3. Estimation of $\widetilde{\beta}_0(v)$ and $\Lambda(c,v)$

Based on the estimation of $\widetilde{\beta}_2(v)$, we can estimate $\widetilde{\beta}_0(v) = \beta_0(v)/\beta_1(v)$ and $\Lambda(c,v) = \lambda(c)/\beta_1(v)$, using the marginal integration described in the proof of Lemma 2.2. First, given $\widehat{\widetilde{\beta}}_2(v)$, we can estimate $\widetilde{\lambda}(c,v) = \widetilde{\beta}_0(v) - \Lambda(c,v)$ by $\widehat{\widetilde{\lambda}}(c,v)$, which maximizes

$$\begin{aligned}&\widehat{S}_n(a|c,v) = \widehat{S}_n\left(\widehat{\widetilde{\beta}}_2(v), a\middle|c,v\right)\\ &= \frac{1}{nh_2}\sum_{i=1}^{n}\left(2D_i(c)-1\right)G_1\left(\frac{X_{1i}+\widetilde{X}_i'\widehat{\widetilde{\beta}}_2(v)+a}{h_1}\right)K_2\left(\frac{\widehat{\widetilde{V}}_i-v}{h_2}\right)t_\kappa\left(Z_i^c\right)\end{aligned}$$

with respect to a. Based on (2.3.8), we can estimate $\widetilde{\beta}_0(v)$ by

$$\widehat{\widetilde{\beta}}_0(v) = \int_{\underline{c}}^{\bar{c}}\widehat{\widetilde{\lambda}}(c,v)\,\omega(c)\,dc,$$

where $\omega(c)$ is a nonnegative weighting function, satisfying

$$\int_{\underline{c}}^{\bar{c}}\omega(c)dc = 1.$$

$\Lambda(c,v)$ is estimated by

$$\widehat{\Lambda}(c,v) = \widehat{\widetilde{\beta}}_0(v) - \widehat{\widetilde{\lambda}}(c,v).$$

3.4. Average Derivative of MSF

In this subsection, we discuss the estimation of $\vartheta = \mathbb{E}\left[\,\mathrm{MSF}'(X,V)\right]$ defined by equation (2.3.10). Due to some technical reasons as noted in HS, instead of estimating ϑ directly, it is more convenient to focus on a trimmed version of this parameter, namely,

$$\vartheta_\kappa = \mathbb{E}\Big(\,\mathrm{MSF}'(X,V)t_\kappa(\mathfrak{X})t_\kappa(Z^c)\Big), \tag{3.4.1}$$

where $t_\kappa(z) = 1\{|z| \le \kappa\}$ and $t_\kappa(x) = 1\{|x| \le \kappa\}$. Compared with ϑ, ϑ_κ trims on the continuous endogenous regressor $\mathfrak{X}$ as well as continuous components of Z. If each component of Z has a binary or discrete support, $t_\kappa(Z^c) \equiv 1$.

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

Remark 3.4.1. There need be no practical difference between the trimmed parameter ϑ_κ and the untrimmed ϑ. ϑ_κ can be made arbitrarily close to ϑ by choosing a large enough trimming constant κ. As noted by HS, the trimming in ϑ_κ prevents boundary bias and ratio bias in the resulting estimator. If $\vartheta = \infty$, then ϑ_κ can be viewed as a robust measure of ϑ.

Remark 3.4.2. As detailed in the Appendix, trimming both $\mathfrak{X}_i$ and Z_i^c prevents boundary bias when $\widehat{\widetilde{V}}_i$ is too close to the boundary of $[0,1]$, while trimming Z_i^c prevents ratio bias when the denominator of $\widehat{\widetilde{V}}_i$ is too close to zero. We emphasize that such fixed trimming is purely technical and could be replaced with asymptotically negligible trimming, so that the limiting estimand would be identically ϑ. However, the cost involves lengthening an already long proof, while the practical benefit is negligible.

Apparently, it follows from equation (2.3.5) that a natural estimator for ϑ_κ can be obtained by replacing each ingredient in (2.3.5) with its sample counterpart, and then averaging over the distribution of (X, V). Such estimator, however, will entail constructing $\widehat{\Lambda}^{-1}(c,v)$ and $\widehat{\Lambda}'(c,v)$ from the estimate of $\widehat{\Lambda}(c,v)$. For example, one may estimate $\Lambda'(c,v)$ by

$$\widehat{\Lambda}'(c,v) = \frac{\widehat{\Lambda}(c+h,v) - \widehat{\Lambda}(c-h,v)}{2h},$$

for any given v and a small h. However, such direct plug-in procedures may cause a substantial computational burden and, even worse, make the subsequent asymptotic analysis a formidable task. To remedy this drawback, we propose a computationally simpler estimator for ϑ_κ, which is capable of obviating the need for nonparametric estimation of $\widehat{\Lambda}^{-1}(c,v)$ and $\widehat{\Lambda}'(c,v)$. The estimator we propose for ϑ_κ is

$$\widehat{\vartheta}_\kappa = \frac{1}{nh_3}\sum_{i=1}^{n}\int_{\underline{c}}^{\bar{c}} K_3\left(\frac{\widehat{\widetilde{\beta}}_0\left(\widehat{\widetilde{V}}_i\right) + X_{1i} + \widetilde{X}_i'\widehat{\widetilde{\beta}}_2\left(\widehat{\widetilde{V}}_i\right) - \widehat{\Lambda}\left(c, \widehat{\widetilde{V}}_i\right)}{h_3}\right) dc$$

$$\times \begin{pmatrix} 1 \\ \widehat{\widetilde{\beta}}_2\left(\widehat{\widetilde{V}}_i\right) \end{pmatrix} t_\kappa(\mathfrak{X}_i) t_\kappa(Z_i^c). \tag{3.4.2}$$

The key observation underlying (3.4.2) is that, for any v,

$$\begin{aligned}
&\int \frac{1}{h_3} K_3\left(\frac{\widetilde{\beta}_0(v) + X_1 + \widetilde{X}'\widetilde{\beta}_2(v) - \Lambda(c,v)}{h_3}\right) dc \\
&=^{\dagger_2} \int \frac{1}{h_3} K_3(t)\, d\Lambda^{-1}\left(\widetilde{\beta}_0(v) + X_1 + \widetilde{X}'\widetilde{\beta}_2(v) - th_3, v\right) \\
&=^{\dagger_3} \frac{1}{\Lambda'\left(\Lambda^{-1}\left(\widetilde{\beta}_0(v) + X_1 + \widetilde{X}'\widetilde{\beta}_2(v), v\right), v\right)} + O_p\left(h_3^{\kappa_3}\right),
\end{aligned} \tag{3.4.3}$$

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

where $\dagger_2$ follows by a change of variable $\widetilde{\beta}_0(v) + X_1 + \widetilde{X}'\widetilde{\beta}_2(v) - \Lambda(c, v) = h_3 t$ and $\dagger_3$ follows if we assume $K_3(\cdot)$ is a κ_3th-order bias reduction kernel. In the next section, we will show $\widehat{\vartheta}_\kappa$ is $\sqrt{n}$-consistent and derive its limiting distribution.

4. LARGE SAMPLE PROPERTIES

In this section, we study the asymptotic properties of $\widehat{\vartheta}_\kappa$ described in the preceding section. The asymptotic analysis of the estimator, though conceptually fairly straightforward, is very complicated, because the estimator is an average of optimization estimators evaluated at nonparametrically generated variables. For ease of presentation, we first sketch the overall strategy used to derive the limiting distribution before stating the main theorem formally. Then, we introduce the technical assumptions used to obtain the general asymptotic result. A detailed proof of the main theorem is given in the Appendix.

Let $\mathfrak{X}_\kappa = \{x \in S_\mathfrak{X} : |x| \le \kappa\}$, $\mathcal{Z}_\kappa = \{z \in S_Z : |z| \le \kappa\}$, $\widetilde{V}(x, z) = F_{\mathfrak{X}|Z}(x, z)$, and $\widetilde{V}_\kappa = \left\{\widetilde{V}(x, z) \in S_{\widetilde{V}} : x \in \mathfrak{X}_\kappa, z \in \mathcal{Z}_\kappa\right\}$ be a compact subset of $S_{\widetilde{V}}$. To derive the limiting distribution of $\widehat{\vartheta}_\kappa$, we first establish the rate of uniform convergence for $\widehat{\widetilde{V}}(x, z)$ and $\widehat{\widetilde{\lambda}}(c, v, b)$. For ease of notation, we abbreviate k_{Z^c} to e. Specifically, we establish in Lemmas A.1–A.3 that

$$\sup_{x \in \mathfrak{X}_\kappa, z \in \mathcal{Z}_\kappa} \left|\widehat{\widetilde{V}}(x, z) - \widetilde{V}(x, z)\right| = O_p\left(\frac{1}{\sqrt{n}h_0^e}\right), \tag{4.1}$$

$$\sup_{c \in [\underline{c}, \bar{c}], v \in \widetilde{V}_\kappa, b \in \mathbf{B}_{2v}} \left|\widehat{\widetilde{\lambda}}(c, v, b) - \widetilde{\lambda}(c, v, b)\right| = O_p\left(\frac{1}{\sqrt{n}h_0^e h_1 h_2^2}\right), \tag{4.2}$$

$$\sup_{c \in [\underline{c}, \bar{c}], v \in \widetilde{V}_\kappa, b \in \mathbf{B}_{2v}} \left|\frac{\partial\widehat{\widetilde{\lambda}}(c, v, b)}{\partial b} - \frac{\partial\widetilde{\lambda}(c, v, b)}{\partial b}\right| = O_p\left(\frac{1}{\sqrt{n}h_0^e h_1^2 h_2^2}\right), \tag{4.3}$$

where $\mathbf{B}_{2v}$ is a compact set containing $\widetilde{\beta}_2(v)$ as the interior. These results will be used to show the uniform consistency of $\widehat{\widetilde{\beta}}_2(v)$. Specifically, we will establish in Lemma A.4 that

$$\sup_{v \in \widetilde{V}_\kappa} \left|\widehat{\widetilde{\beta}}_2(v) - \widetilde{\beta}_2(v)\right| = O_p\left(\frac{1}{\sqrt{n}h_0^e h_1 h_2^2}\right). \tag{4.4}$$

Remark 4.1. The uniform convergence rates in (4.1)–(4.4) are larger than the usual in the kernel estimation literature. These larger rates are mainly suggested by Sherman (1994a). As pointed out by Sherman, proofs using such larger rates avoid much of the technical detail required by more traditional methods of analysis, but have almost no effect on inference in finite samples. Detailed discussions can be found in Sherman (1994a, 1994b).

Based on the uniform convergence result in equation (4.4), we derive the asymptotic representation for $\widehat{\widetilde{\beta}}_2(v) - \widetilde{\beta}_2(v)$ in Lemma A.5. Lemma A.6 quantifies

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

the penalty paid for having to replace $\widehat{\widetilde{\beta}}_2(\widetilde{V}_i)$ with $\widehat{\widetilde{\beta}}_2\left(\widehat{\widetilde{V}}_i\right)$ in the estimator. Then, we derive the asymptotic representation for $\widehat{\widetilde{\lambda}}(c,v) - \widetilde{\lambda}(c,v)$ in Lemma A.7, and similarly, Lemma A.8 quantifies the penalty paid for having to replace $\widehat{\widetilde{\lambda}}(c,\widetilde{V}_i)$ with $\widehat{\widetilde{\lambda}}\left(c,\widehat{\widetilde{V}}_i\right)$ in the estimator. Finally, the asymptotic distribution of $\widehat{\vartheta}_\kappa$ will be obtained by applying the Hoeffding decomposition and central limit theorem to each term represented as a sum of degenerate U-statistics. We now formally state the assumptions used to prove the main theorem.

Assumption 4.1. (i) The first-stage equation in (1.3) holds, $Z \perp (U,V)$, and $g(z,v)$ is an unknown function invertible in v for each possible value z. (ii) Let $F_V(\cdot)$ denote the cumulative distribution function of V. $F_V(\cdot)$ is a differentiable, invertible function from the support of V onto $[0,1]$.

Assumption 4.1(i) defines key instrument and control function assumptions. The invertibility condition implies that, for each $z \in S_Z$, there exists an invertible map from S_V onto $S_{\mathfrak{X}}$. Assumption 4.1(ii) implies the existence of a smooth invertible mapping from S_V onto $S_{\widetilde{V}}$. The existence of such mapping implies validity of localization on $\widetilde{V}$ in the estimation.

Assumption 4.2. (i) $g(z,v)$ is twice differentiable in v. (ii) Let $f(x_1|\widetilde{x},v)$ denote the probability density function of X_1 given $\widetilde{X} = \widetilde{x}$ and $\widetilde{V} = v$ evaluated at x_1. $f(x_1|\widetilde{x},v)$ is continuous in $\widetilde{x}$ and v. (iii) $f(\widetilde{x}|v)$ denotes the probability density function of $\widetilde{X}$ given $\widetilde{V} = v$ evaluated at $\widetilde{x}$. $f(\widetilde{x}|v)$ is continuous in v.

Assumption 4.2 imposes smoothness assumptions on the first-stage equation and the conditional density functions. This assumption is used to show the negligibility of a bias term in the proof of uniform consistency of $\widehat{\widetilde{\lambda}}(c,v,b)$ and $\left(\widehat{\widetilde{\beta}}_0(v), \widehat{\widetilde{\beta}}_2(v)\right)$.

Assumption 4.3. (i) $\widetilde{\lambda}(c,v)$, $\widehat{\widetilde{\lambda}}(c,v)$, $\left(\widetilde{\beta}_0(v), \widetilde{\beta}_2(v)\right)$, and $\left(\widehat{\widetilde{\beta}}_0(v), \widehat{\widetilde{\beta}}_2(v)\right)$ are twice continuously differentiable functions of v on $S_{\widetilde{V}}$. (ii) For each $v \in \widetilde{V}_\kappa$ and $c \in (\underline{c}, \overline{c})$, there is a compact set Ξ_{cv} containing $\widetilde{\lambda}(c,v)$ as the interior. For each $v \in \widetilde{V}_\kappa$, there is a compact set $\mathbf{B}_v$ that contains $\left(\widetilde{\beta}_0(v), \widetilde{\beta}_2(v)\right)$ as the interior. (iii) The correspondence from $v \in \widetilde{V}_\kappa$ to $\mathbf{B}_v$ is upper hemicontinuous.

Assumption 4.3(i) imposes smoothness conditions on $\widehat{\widetilde{\lambda}}(c,v)$ and $(\widehat{\widetilde{\beta}}_0(v), \widehat{\widetilde{\beta}}_2(v))$, which are sufficient to carry out certain Taylor expansions in the proof. Similar to Assumption A8 in HS, these smoothness conditions, as a matter of fact, are made solely for the sake of clarity and ease of exposition. They are unnecessary. For example, the smoothness of $\widehat{\widetilde{\lambda}}(c,v)$ and $\left(\widehat{\widetilde{\beta}}_0(v), \widehat{\widetilde{\beta}}_2(v)\right)$ (with probability tending to one as n approaches infinity) can be implied by the implicit function theorem and the smoothness of $\widehat{S}_n(b,a|c,v)$ and $\widehat{S}_n(b|v)$ in both (a,b) and (c,v). Assumption

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

4.3(ii) is a standard compactness assumption, and Assumption 4.3(iii) is a mild regularity condition used in establishing uniform consistency of $\widehat{\widetilde{\beta}}_2(v)$.

Assumption 4.4. (i) For $s = 0, 1, 2, 3$, $K_s(\cdot)$ is a continuously differentiable, κ_sth-order bias reduction kernel function with κ_s an even integer, or

$$\int K_s(t)\,dt = 1, \int K_s(t)\,t^r dt = 0,$$

for $1 \leq r < \kappa_s, s = 0, 1, 2, 3$. (ii) The corresponding bandwidths can be expressed as $h_s = c_s n^{-\delta_s}$, with $c_s > 0$, $\delta_s > 0$, for $s = 0, 1, 2, 3$. Moreover, the bandwidths should satisfy $\frac{1}{2\kappa_1} < \delta_1$, $\frac{1}{2\kappa_2} < \delta_2$, $\frac{1}{2\kappa_3} < \delta_3$, $3\delta_1 + 4\delta_2 < \frac{1}{2}$, $(\delta_2 - \delta_1)\kappa_2 > \frac{1}{2}$, when $e = 0$, and satisfy $\frac{1}{2\kappa_0} < \delta_0 < \frac{1}{4e}$, $\frac{1}{2\kappa_1} < \delta_1$, $\frac{1}{2\kappa_2} < \delta_2$, $\frac{1}{2\kappa_3} < \delta_3$, $2e\delta_0 + 3\delta_1 + 4\delta_2 < \frac{1}{2}$, $(\delta_2 - \delta_1)\kappa_2 > \frac{1}{2}$, when $e > 0$.

Assumption 4.5. (i) Let $f\left(z^c \middle| z^d\right)$ denote the density of Z^c given $Z^d = z^d$ evaluated at $Z^c = z^c$. $f\left(z^c \middle| z^d\right)$ is bounded away from zero and has κ_0th bounded derivatives with respect to z^c. (ii) Let $f\left(x'b \middle| \widetilde{x}\right)$ denote the density of $X'b$ given $\widetilde{X} = \widetilde{x}$ evaluated at $x'b$, for $b \in \mathbf{B}_v$. $f\left(x'b \middle| \widetilde{x}\right)$ is bounded away from zero and has κ_1th-order bounded derivatives with respect to $x'b$. (iii) $\lambda(\cdot)$ has κ_2th-order bounded derivatives. (iv) Let $f\left(v \middle| z, x'b\right)$ denote the density of $\widetilde{V}$ given $Z = z$ and $X'b = x'b$ evaluated at $\widetilde{V} = v$. $f\left(v|z, x'b\right)$ is bounded away from zero and has κ_2th-order bounded derivatives with respect to v. (v) Let

$$\mu(v, z, c) = 2P\left(\widetilde{\varepsilon}(v) < x_1 + \widetilde{x}'\widetilde{\beta}_2(v) + \widetilde{\lambda}(c, v) \middle| \widetilde{V} = v, Z = z\right) - 1.$$

$\mu(v, z, c)$ has κ_2th-order bounded derivatives with respect to v. (vi) Let $\Pi(t, v) = 1/\Lambda'\left(\Lambda^{-1}(t, v), v\right)$. $\Pi(t, v)$ has κ_3th-order bounded derivatives with respect to t.

Assumption 4.5 consists of a set of high-order smoothness conditions. Assumption 4.5(i) imposes smoothness on the joint density of the continuous components of Z given the remaining discrete components. These conditions are used along with Assumption 4.4 to show that the bias term $\mathbb{E}\widehat{\widetilde{V}}(x, z) - \widetilde{V}(x, z)$ is negligible. Assumption 4.5(ii)–(v) gives smoothness conditions on the transformation function, some other densities, and conditional expectations. They are used along with Assumption 4.4 to show that a bias term consisting of a certain expected gradient is negligible. Assumption 4.5(vi) is made solely for the sake of clarity and ease of exposition. Recall that

$$\Pi(t, v) = \frac{\partial \Lambda^{-1}(t, v)}{\partial t} = \frac{\partial \lambda^{-1}(\beta_1(v) \cdot t)}{\partial t}.$$

Thus, Assumption 4.5(vi) is satisfied if Assumption 4.5(iii) is strengthened to "$\lambda(\cdot)$ has $\max\{\kappa_2, \kappa_3 + 1\}$th-order bounded derivatives."

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

Assumption 4.6. For each $v \in \widetilde{V}_\kappa$, the matrix

$$\sigma(v) = \int_{\underline{c}}^{\overline{c}} \left\{ \alpha_{XX}(c,v) - \alpha_{OO}(c,v) \left[\frac{\partial \widetilde{\lambda}(c,v,b)}{\partial b} \frac{\partial \widetilde{\lambda}(c,v,b)}{\partial b'} \right]_{b=\widetilde{\beta}_2(v)} \right\} \omega(c)\, dc$$

is nonsingular, where

$$\alpha_{XX}(c,v) = \int \varphi\left(v, \Lambda(c,v) - \widetilde{\beta}_0(v), c, \widetilde{X}\right) \widetilde{X}\widetilde{X}' dF_{\widetilde{X}},$$

$$\alpha_{XO}(c,v) = \int \varphi\left(v, \Lambda(c,v) - \widetilde{\beta}_0(v), c, \widetilde{X}\right) \widetilde{X} dF_{\widetilde{X}},$$

$$\alpha_{OO}(c,v) = \int \varphi\left(v, \Lambda(c,v) - \widetilde{\beta}_0(v), c, \widetilde{X}\right) dF_{\widetilde{X}},$$

and $\varphi(v,s,c,\widetilde{x}) = -\dfrac{\partial \{\mathcal{I}_n(v,t+s,c) f(t+s|\widetilde{x})\}}{\partial t}|_{t=0}$, in which $f(\cdot|\widetilde{x})$ denotes the density of $X'b$ given $\widetilde{X} = \widetilde{x}$, and

$$\mathcal{I}_n(v,t,c) = \mathbb{E}\left[t_\kappa(Z^c)\, \mu(v,Z,c) f\left(v \middle| Z, X'b\right) \middle| X'b = t \right].$$

Assumption 4.6 gives regularity conditions on the Hessian of the population analog of the sample criterion function $\widehat{S}_n(b|v)$.

Remark 4.2. For ease of exposition, the above assumptions are stated in terms of the rescaled coefficients and variables instead of the original ones. Specifically, according to equations (2.1.1) and (2.2.1), the original model (1.1) is equivalent to

$$\frac{\lambda(Y^*)}{\beta_1(v)} = \widetilde{\beta}_0(v) + X_1 + \widetilde{X}'\widetilde{\beta}_2(v) - \widetilde{\varepsilon}(v), \tag{4.5}$$

for each $v \in S_V$, where $\widetilde{\varepsilon}(v) = \varepsilon / \beta_1(v)$, and ε is defined in equation (2.1.1). Then, Assumptions 4.1–4.6 are stated, as a matter of fact, explicitly for (4.5).

THEOREM 4.1. *Under the assumptions of Lemma 2.2 and Assumptions 4.1–4.6,*

$$\sqrt{n}\left(\widehat{\vartheta}_\kappa - \vartheta_\kappa\right) \to^d \mathcal{N}(0, \Sigma_\vartheta),$$

where $\Sigma_\vartheta = \mathbb{E}\left[f_\vartheta(\xi_i) f_\vartheta(\xi_i)'\right]$ *with* $\xi_i = \left(Y_i, X_i, Z_i, \widetilde{V}_i\right)$, $f_\vartheta(\xi_i) = f_{1n}(\xi_i) + \sum_{j=5}^{11} f_{jn}(\xi_i)$, *and each* $f_{ln}(\xi_i)$, $l = 1, 5, \ldots, 11$, *has zero mean and finite variance.*

The explicit expression of each $f_{ln}(\cdot)$ as well as the proof of this result is given in the Appendix. Because each $f_{ln}(\xi_i)$ is shown to have zero mean and finite variance, the result then follows by the application of any central limit theorem for i.i.d. random variables with finite variance, and Slutsky's Theorem.

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

Remark 4.3. A consistent estimator of the asymptotic variance can be obtained by replacing, in the expressions above, ξ_i with $\left(Y_i, X_i, Z_i, \widehat{\widetilde{V}}_i\right)$ and population expectations with the corresponding sample averages. Meanwhile, constructing a direct estimator of the asymptotic variance would be tedious and difficult, likely requiring an additional estimator of $\dfrac{\partial \widetilde{\lambda}(c,v,b)}{\partial b}$ as a function of (c,v,b). Instead, we propose a bootstrap procedure to approximate the limiting distribution of $\widehat{\vartheta}_\kappa$. We first outline the bootstrap procedure used in the simulation. Let us draw S sets of n observations with replacement from $\{(Y_i, X_{1i}, Z_i)\}_{i=1}^n$, where S and n denote the time of repetitions and sample size, respectively. In the simulation study, we set $S = 200$ and choose n to be 250 and 500. These observations are used to compute $\widehat{\vartheta}^*_{\kappa,s}$, for $s = 1, 2, \ldots, S$. Then,

$$\widehat{\Sigma}_\vartheta = \frac{1}{S-1}\sum_{s=1}^{S}\left(\widehat{\vartheta}^*_{\kappa,s} - \bar{\vartheta}^*_\kappa\right)\left(\widehat{\vartheta}^*_{\kappa,s} - \bar{\vartheta}^*_\kappa\right)'$$

with $\bar{\vartheta}^*_\kappa = \frac{1}{S}\sum_{s=1}^{S}\widehat{\vartheta}^*_{\kappa,s}$ forms a bootstrapped estimate of the variance of $\widehat{\vartheta}_\kappa$. The formal assumptions used to establish the validity of the bootstrap are given in Appendix B.

5. MONTE CARLO SIMULATION

In this section, we conduct a small-scale Monte Carlo simulation to explore the finite sample performance of our estimator. We consider the following three DGPs:

Design 1: $Y^* = B_0 + B_1 X_1$;

Design 2: $\lambda\left(Y^*, m\right) = B_0 + B_1 X_1$, with $m = 0.7$;

Design 3: $\lambda\left(Y^*, m\right) = B_0 + B_1 X_1$, with $m = 1.3$,

where $\lambda(y,m) = \dfrac{|y|^m \operatorname{sgn}(y) - 1}{m}$ is a modified Box–Cox transformation, and $\operatorname{sgn}(y) = 1$ if $y > 0$, $\operatorname{sgn}(y) = 0$ if $y = 0$, and $\operatorname{sgn}(y) = -1$ if $y < 0$. Let Y be generated by

$$Y = \underline{c} \times 1\left\{Y^* < \underline{c}\right\} + Y^* \times 1\left\{\underline{c} \le Y^* \le \bar{c}\right\} + \bar{c} \times 1\left\{Y^* > \bar{c}\right\},$$

with the left cutoff point $\underline{c}$ being the 15% empirical quantile of Y^* and the right cutoff point $\bar{c}$ the 85% empirical quantile of Y^*, resulting in about 30% degree of censoring. We specify the random coefficients (B_0, B_1) to be functions of V to generate their correlation with X_1:

$$X_1 = Z + V, \quad B_0 = b_0 + \rho_0 V + \epsilon_0, \quad B_1 = b_1 + \rho_1 V + \epsilon_1, \tag{5.1}$$

where $b_0 = 0.4$, $b_1 = 0.8$, $\rho_0 = 0.3$, and $\rho_1 = 0.7$. The generation processes of other random components in (5.1) are summarized below: $V \sim \mathit{Uniform}[-0.5, 0.5]$; each component of $(Z, \epsilon_0, \epsilon_1)$ follows a standard normal distribution; V, Z, ϵ_0, and ϵ_1 are

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

mutually independent. Throughout the simulation, we normalize B_1, namely the coefficient on X_1 to one.

The trimming functions $t_\kappa(Z_i) = 1\{|Z_i| \le \kappa\}$ and $t_\kappa(X_{1i}) = 1\{|X_{1i}| \le \kappa\}$, with $\kappa = 3$. This corresponds to trimming 4.8% of the data. We use the uniform weighting scheme

$$\omega(c) = \varpi(c) = \frac{1}{\bar{c} - \underline{c}} 1\{\underline{c} < c < \bar{c}\}.$$

We apply a numerical integration using a grid on $(\underline{c}, \bar{c})$ on the integration over c.[6] Assumption 4.4 requires sufficient high-order kernel. Thus, we choose a 16th-order Epanechnikov kernel for computing $\widehat{\widetilde{V}}_i$, that is,

$$K_0(t) = c_0(6864 - 240240t^2 + 2450448t^4 - 11085360t^6 + 25865840t^8 - 32449872t^{10} + 20801200t^{12} - 5348880t^{14})1\{-1 \le t \le 1\},$$

where $c_0 = 0.0006712228$. We use this same kernel to localize, that is, $K_1(t) = K_2(t) = K_0(t)$. We choose $K_3(\cdot)$ to be

$$K_3(t) = \left(\frac{15}{8} - \frac{5}{4}t^2 + \frac{1}{8}t^4\right)\exp\left(-\frac{t^2}{2}\right),$$

a Gaussian kernel with order six.

We specify the bandwidths $h_s = c_s n^{-a_s}$, for $s = 0, 1, 2, 3$, where h_0 denotes the bandwidth for estimating $\widehat{\widetilde{V}}_i$, h_1 for smoothing the integrated kernel $G_1(\cdot)$, h_2 for localizing, and h_3 for estimating $\widehat{\vartheta}_\kappa$. The bandwidths are determined by two steps. First, we choose the convergence rate, a_s, for $s = 0, 1, 2, 3$, by solving a system of inequalities, in accordance with the asymptotic order required by Assumption 4.4. Note that $\kappa_0 = \kappa_1 = \kappa_2 = 16$ and $\kappa_3 = 6$. We choose an arbitrary set of (a_0, a_1, a_2, a_3) satisfying

$$\begin{cases} 1/32 < a_0 < 1/4, \\ 1/32 < a_1, a_2, \\ 2a_0 + 3a_1 + 4a_2 < 1/2, \\ 1/32 < a_2 - a_1, \\ 1/12 < a_3. \end{cases}$$

In the second step, we choose the constants of bandwidths, i.e., c_s, for $s = 0, 1, 2, 3$. Because $\widehat{\widetilde{V}}_i$ is a Nadaraya–Watson estimator, c_0 can be determined by the standard cross-validation method (see, e.g., Li and Racine, 2007). Note that h_1 is used for smoothing the integrated kernel $G_1(\cdot)$. As noted by Chen (2010), we find the integrated local maximum score estimator like (3.2.1) without smoothing performs similarly as the estimator with smoothing. Thus, for simplicity, we avoid the selection of c_1, by using the indicator function directly. Moreover,

[6] Similar to Remark 3.1 in Chernozhukov, Fernández-Val, and Melly (2013), the number of grid points m and the number of observations n should satisfy $\sqrt{n}/m \to 0$.

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

TABLE 1. Simulation results

			Design 1 (linear)	Design 2 ($m = 0.7$)	Design 3 ($m = 1.3$)
$n = 250$	AD-MSF	Bias	0.0707	0.1377	0.0902
		SD	0.0889	0.1493	0.0856
		Coverage rate	89%	86.5%	78.5%
$n = 500$	AD-MSF	Bias	0.0125	0.0070	0.0204
		SD	0.0668	0.0936	0.0574
		Coverage rate	95%	95.5%	93.5%

because h_2 is used for localizing in the maximum score estimation, c_2 can be determined by a simple one-dimensional grid search. For example, we construct a sequence of possible values of c_2, indexed by $\underline{c}_2 < c_{21} < \cdots < c_{2L} < \bar{c}_2$. Then, we choose the optimal c_2^* from the sequence, which maximizes the sample objective function (3.2.1) evaluated at the estimated $\left(\widehat{\widehat{\beta}}_2(v), \widehat{\widehat{\lambda}}(c, v)\right)$. Finally, because $\widehat{\vartheta}_\kappa$ is $\sqrt{n}$-consistent, the estimation result is insensitive to the selection of c_3.

For each design, we compute $\widehat{\vartheta}_\kappa$ (AD-MSF) for $n \in \{250, 500\}$. Table 1 reports the empirical bias and empirical standard deviation for each estimator based on 200 repetitions. These results indicate that our estimator performs well for each DGP and sample size. The magnitude of bias decreases as the sample becomes large. As expected, the empirical standard deviation decreases steadily as the sample size increases and the magnitude of such decline is consistent with $\sqrt{n}$-asymptotic, as predicted by the theory. Moreover, we use the bootstrap to estimate the standard deviation, and construct the confidence intervals using the normal approximation based on 100 bootstrap repetitions. Then, we report the coverage rate of these confidence intervals with 95% confidence level.

From the simulation result, we can see the benefit of focusing on the AD-MSF. Take Design 1 as an example. Assume that we are interested in the average partial effect of X_1 on Y, which can be estimated by two approaches. The first approach estimates a linear CRC model, which is equivalent to imposing the restriction $\lambda(y) = y$. For example, the method proposed by Masten and Torgovitsky (2016) can be used to estimate EB_1 in an instrumental variable CRC model with continuously distributed endogenous regressors. The second approach addresses the problem from a CRCT model perspective, as we do in this section. By fitting the data with a CRCT model instead of a linear CRC model, we add more flexibility to the model and effectively avoid misspecification of the link function. One undesirable aspect of a CRCT model is that the distribution of the coefficient can only be identified up to scale, not point identified. But such scale normalization does not affect the estimation of the AD-MSF with respect to X_1, which remains equal to EB_1 in this case.

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

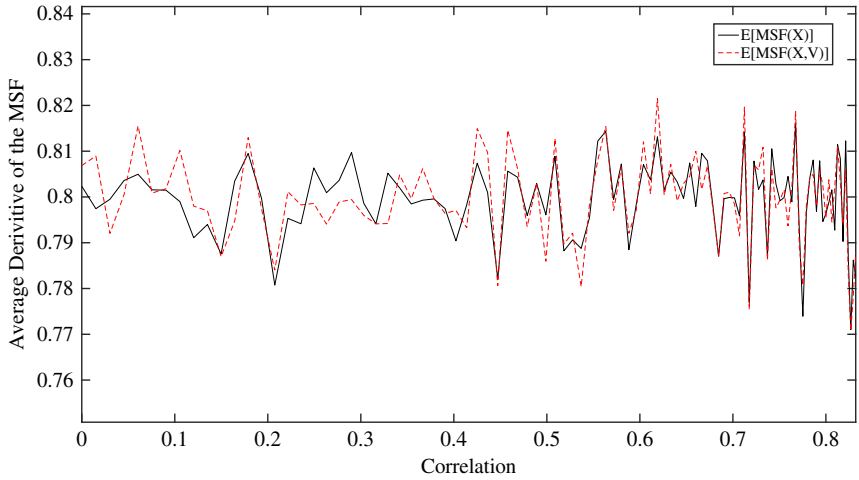


FIGURE 1. Comparison of $\mathbb{E}[\text{MSF}'(X)]$ and $\mathbb{E}[\text{MSF}'(X,V)]$ in a linear case.

Finally, we provide some numerical evidence on the relationship between $\text{MSF}'(x)$ and $\text{MSF}'(x,v)$, as discussed in Remark 2.3.2. We use the same DGPs 1–3, except that we choose $m = 0.9$ in Design 2 and $m = 1.1$ in Design 3. We vary ρ_1, namely the coefficient of V on B_1, to control the correlation (magnitude of endogeneity) between B_1 and V. The horizontal coordinate shows the correlation coefficient between B_1 and X_1, and the vertical coordinate shows the values of $\mathbb{E}[\text{MSF}'(X)]$ (solid line) and $\mathbb{E}[\text{MSF}'(X,V)]$ (dotted line). In Figure 1, we see the two parameters coincide when the transformation function is linear, which is consistent with the discussion in Remark 2.3.2. Figures 2 and 3 show the two parameters with nonlinear $\lambda(\cdot)$. The two parameters are approximately equivalent in the exogeneity situation (ρ_1 is small), and gradually diverge when the magnitude of endogeneity increases. Apparently, the difference of the two parameters depends on the correlation between V and B_1 in the nonlinear cases. Recall the definitions of two parameters in equations (2.3.3) and (2.3.8). Thus, $\mathbb{E}(B_1|V = v)$ does not equal $\mathbb{E}B_1$ if V is correlated with B_1, which leads to the difference of the two parameters. Moreover, these differences imply that the average marginal effect measured by $\mathbb{E}[\text{MSF}'(X,V)]$ varies with the magnitude of endogeneity, while $\mathbb{E}[\text{MSF}'(X)]$ exogenously measures the average marginal effect of the median potential outcome in nonlinear cases.

6. EMPIRICAL STUDY

In this section, we apply our model to the estimation of the Engel curve based on the U.K. Family Expenditure Survey (FES) as in Blundell et al. (2007; hereafter BCK). The Engel curve relationship describes how a household's demand for

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

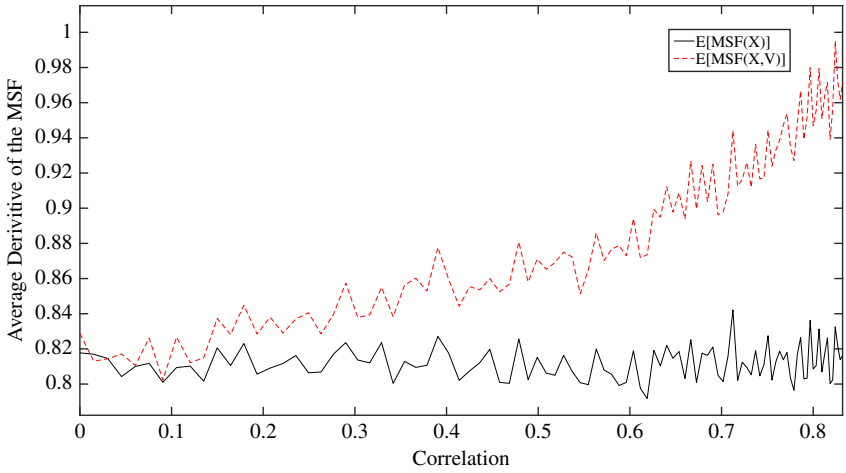


FIGURE 2. Comparison of $\mathbb{E}[\ \mathrm{MSF}'(X)]$ and $\mathbb{E}[\ \mathrm{MSF}'(X,V)]$ under DGP 2 with $m = 0.9$.

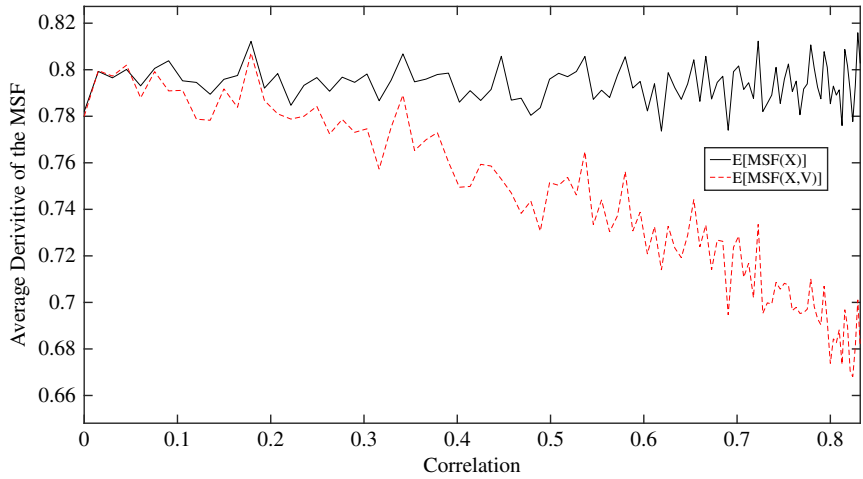


FIGURE 3. Comparison of $\mathbb{E}[\ \mathrm{MSF}'(X)]$ and $\mathbb{E}[\mathrm{MSF}'(X,V)]$ under DGP 3 with $m = 1.1$.

a commodity changes as the household's expenditure increases. The analysis is performed based on the FES data of a single year, 1995. There are a number of studies that estimate Engel curves using the FES data, such as BCK, Imbens and Newey (2009), and VV, among others. Especially, VV study the functional form of expenditure on leisure goods, travel, and household goods. Because the dependent variable (budget share) takes values between zero and one, it is common practice to apply a logarithm or Box–Cox transformation to the dependent variable, as these transformations convert a positive variable into a variable supported on

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

Table 2. Descriptive statistics of the data

Variable	Description (1,184 obs)	Mean	SD	Min	Max
shl	Share of total expenditure on leisure	0.0825	0.1475	0	1
lnexp	Log of total expenditure	5.7412	0.8151	−3.9120	8.5735
lnge	Log gross earnings	4.2944	1.1223	−1.6094	7.6811

the real line. Instead of taking a logarithmic transform as in BCK or a Box–Cox transformation as in VV, we use a more general monotone function $\lambda(\cdot)$ to mitigate the risk of model misspecification.

There are two additional motivations for our approach. First, the estimated regression function in VV actually merely reflects how the explanatory variable (log total expenditure) affects the budget share after the Box–Cox transformation. A more detailed interpretation of the estimation results should combine both the regression function and the estimated parameter in the Box–Cox transformation (θ in VV). By contrast, our method directly identifies some central effect of X on the distribution of the latent outcome, whose interpretation is invariant to both location and scale normalization on the coefficients. Second, as shown below, about 25% of families report zero expenditure on leisure goods. Therefore, it is more appropriate to use our method, as it is robust against censoring in the dependent variable.

We consider the effect of total expenditure on the share of leisure goods. To preserve some degree of demographic homogeneity, we select from the FES a subset of married or cohabiting couples with children, where the head of household is aged between 20 and 55. Because total expenditure is a choice variable in the consumer's allocation of income across consumption and savings, it is very likely to be jointly determined with individual demands and, thus, is an endogenous regressor. Moreover, because the allocation model of income to individual consumption and savings does suggest that exogenous sources of income are not correlated with household preferences over consumption, we use the gross earnings of the household head as the instrument (see BCK for a similar treatment). This leaves 1,184 observations, and 25% of the data are left censored by zero, because these families report zero expenditure on leisure goods. Table 2 presents the summary statistics for the variables we use in our model.

We estimate the following CRCT model:

$$\begin{cases} Y = \max\{0, Y^*\}, \\ \lambda(Y^*) = B_0 + BX, \\ X = \phi(Z, V), \end{cases}$$

where Y is the share of total expenditure on leisure goods, X the log of total expenditure, and Z the log of gross earnings. For this model, the distribution of (B_0, B) can only be identified up to scale. Based on the estimation results of Figure 4 in VV, and Figure 9 in BCK, we normalize the coefficient of the log of total

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

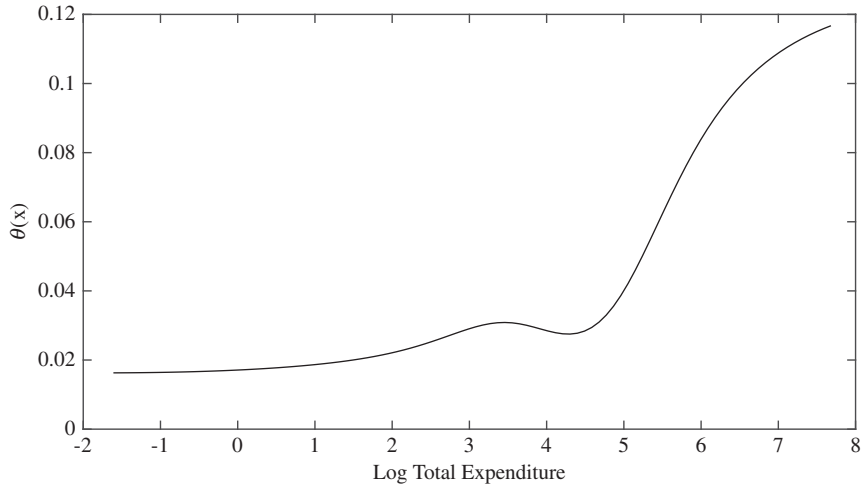


FIGURE 4. Marginal propensity to consume leisure goods conditional on log total expenditure.

expenditure to one, namely $B = 1$. For example, in their Figure 9, BCK report four IV estimates of Engel curves for the leisure goods corresponding to different values of smoothing parameters and all these estimates point to a positive relationship between the total expenditure and the expenditure on leisure goods. Although the coefficient of interest has been normalized to one, we can still identify an economically meaningful parameter, namely the AD-MSF, which captures the causal effect of total expenditure on the median of the distribution of leisure goods share, by averaging over both observable and unobservable characteristics. As discussed in Section 2.3, the AD-MSF is invariant to both location and scale normalization of the coefficients and resembles the usual average treatment effect in the policy evaluation literature.

As a benchmark, we first run a parametric Box–Cox median regression using the Buchinsky (1995) procedure, without correcting for either endogeneity or censoring. The result is displayed in Table 3. We report the estimated coefficients as well as the AD-MSF. Their standard deviations are computed by bootstrap based on 200 repetitions. The slope coefficient β_1 is slightly negative but not significant. Moreover, the average effect of total expenditure on the share of leisure goods is negative and significant, which is unlikely to be accurate. These results suggest the necessity for correcting for both endogeneity and censoring in the estimation.

Then, we compute $\widehat{\vartheta}_\kappa$ based on equation (3.4.2). Bandwidth selection follows the same procedure as in Section 5. Table 3 reports the estimation result together with the standard deviation and confidence interval obtained by bootstrapping based on 200 repetitions. The AD-MSF has an expected sign, a plausible magnitude, and is highly significant. Our result shows that raising one additional percentage of total expenditure, more than half of the households will spend more

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

Table 3. Estimators of AD-MSF based on parametric Box–Cox model and CRCT model

	Parameter	Estimate	Bootstrap std. err.	95% CI lower bound	95% CI upper bound
Box–Cox	λ	0.4326	0.0536	0.3275	0.5377
	β_0	−1.6539	0.7791	−3.1809	−0.1267
	β_1	−0.0387	0.1684	−0.3688	0.2914
	AD-MSF	−0.0022	0.00024	−0.0028	−0.0017
CRCT	AD-MSF	0.0563	0.0121	0.0326	0.0800

Table 4. Simple robust test for AD-MSF

Censoring ratio	30%	35%	40%
Estimate	0.0552	0.0538	0.0523

on leisure goods, and on average, the share of total expenditure on leisure goods will increase by 5.63%. An interesting feature of our estimator is its robustness against censoring. Thus, the result is expected to remain stable if we moderately change the cutoff points. For this purpose, we artificially raise the censoring ratio from 25% to 40% and reestimate the AD-MSF. The estimation results are reported in Table 4.[7] We see that the estimated AD-MSF decreases only by 8%, even when the degree of censoring is raised to 40%.

To see how the marginal propensity to consume leisure goods varies among families with different levels of total expenditure, we average the derivative of the MSF over the distribution of V alone but keeping $X = x$ fixed. In other words, we estimate $\theta(x) = \mathbb{E}\left[\, \mathrm{MSF}'(x, V)\right]$, for $x \in S_X$. Figure 4 plots $\theta(x)$, for $x \in [-1, 7]$. It is interesting to compare our results with BCK's Figure 9. It should be emphasized that the vertical coordinate in BCK denotes the structural mean function of the share of leisure goods, while our vertical coordinate denotes the *derivative* of the median structure function. The curve in Figure 4 is above zero everywhere, indicating that raising total expenditure always has a positive effect on consuming leisure goods, and this is consistent with the results in BCK. Nevertheless, while Figure 4 shows that the derivative of the MSF generally increases with the total expenditure, it seems that the slope of the structure mean function in BCK (Figure 9) is approximately constant. There are two explanations for this difference: first, censoring is a nonnegligible problem in the dataset on leisure goods. Our method is able to overcome this censoring problem, whereas BCK does not handle this. Second, our estimator is expected to remain consistent if the dependent variable is

[7]The standard error could be obtained by bootstrapping, but the computation is time-consuming; thus, we do not report it here.

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

subject to a certain complicated monotone transformation, thus avoiding possible misspecification due to parameterization of this transformation.

7. CONCLUSION

This study considers identification and estimation in a CRC model with an unknown transformation of the dependent variable, and coefficients allowed to be correlated with one or several components of regressors. Moreover, we allow the latent outcome subject to a certain kind of censoring. Under a CMI plus a CMZ restriction, we show the mean of coefficients is identified up to scale. However, this mean of coefficients only tells us the average effect of regressors on the outcome after transformation, not the direct effect on the outcome. Moreover, it has to be interpreted depending on which coefficient is normalized to one. To remedy these drawbacks, we show the derivative of the MSF is point identified. This derivative of MSF resembles the MTE introduced by Heckman and Vytlacil (2005). It generalizes the usual average treatment effect in a linear CRC model and is invariant to both location and scale normalization on the coefficients. We propose a novel and computationally convenient estimator for the AD-MSF, which obviates the need for nonparametric estimation of the inverse and the derivative of the link function.

The proposed estimator is shown to be $\sqrt{n}$-consistent and asymptotically normal. The asymptotic analysis of the estimator, although conceptually fairly straightforward, is very complicated due to the fact that the estimator is multi-stage and involves numerous nonparametrically preestimated random variables. Simulations illustrate that the estimator performs well in finite samples. Finally, we apply our method to estimating the Engel curve using the U.K. FES data.

APPENDICES

A. PROOFS

In this section, we provide a detailed proof of the main theorem accompanied by a series of lemmas. We give an overview of the proof before formally stating the lemmas.

Lemma A.1 shows the rate of uniform convergence for the estimated control variable $\widehat{\widetilde{V}}_i$, and the asymptotic representation for $\widehat{\widetilde{V}}_i - \widetilde{V}_i$. Lemma A.2 shows the uniform consistency for $\widehat{\widetilde{\lambda}}(c,v,b)$. The uniform convergence rate for $\widehat{\widetilde{\lambda}}(c,v,b)$, $\dfrac{\partial \widehat{\widetilde{\lambda}}(c,v,b)}{\partial b}$, and the asymptotic representation for $\widehat{\widetilde{\lambda}}(c,v,b) - \widetilde{\lambda}(c,v,b)$ are presented in Lemma A.3. Next, Lemma A.4 demonstrates the uniform convergence for $\widehat{\widetilde{\beta}}_2(v)$ based on Lemmas A.1 and A.2. Lemmas A.5 and A.6 contain the asymptotic representation for $\widehat{\widetilde{\beta}}_2\left(\widehat{\widetilde{V}}_i\right) - \widetilde{\beta}_2\left(\widetilde{V}_i\right)$. Specifically, Lemma A.5 shows the asymptotic representation for $\widehat{\widetilde{\beta}}_2\left(\widetilde{V}_i\right) - \widetilde{\beta}_2\left(\widetilde{V}_i\right)$, and Lemma A.6 for

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

$\widehat{\widetilde{\beta}}_2\left(\widehat{\widetilde{V}}_i\right) - \widehat{\widetilde{\beta}}_2\left(\widetilde{V}_i\right)$. Similarly, Lemmas A.7 and A.8 contain the asymptotic representation for $\widehat{\widetilde{\lambda}}\left(c, \widehat{\widetilde{V}}_i\right) - \widetilde{\lambda}\left(c, \widetilde{V}_i\right)$.

Finally, using Lemmas A.1–A.8, we can show

$$\sqrt{n}\left(\widehat{\vartheta}_\kappa - \vartheta_\kappa\right) \to^d \mathcal{N}\left(0, \Sigma_\vartheta\right),$$

where $\Sigma_\vartheta = \mathbb{E}\left[f_\vartheta\left(\xi_i\right) f_\vartheta\left(\xi_i\right)'\right]$ with $\xi_i = \left(Y_i, X_i, Z_i, \widetilde{V}_i\right)$, $f_\vartheta\left(\xi_i\right) = f_{1n}(\xi_i) + \sum_{j=5}^{11} f_{jn}(\xi_i)$, and each $f_{ln}(\xi_i)$, $l = 1, 5, \ldots, 11$, has zero mean and finite variance. $f_{5n}(\xi_i)$ consists of $f_{2n}(\xi_i)$, $f_{3n}(\xi_i)$, and $f_{4n}(\xi_i)$. Specifically, $f_{1n}(\xi_i)$ quantifies the first-order contribution to the asymptotic distribution of $\widehat{\vartheta}_\kappa$ if $\widetilde{V}_i$, $\widetilde{\beta}_2(v)$, and $\widetilde{\lambda}(c, v)$ were known. $f_{2n}(\xi_i)$ and $f_{6n}(\xi_i)$ quantify the penalty paid for having to estimate $\widetilde{\beta}_2(v)$ when $\widetilde{V}_i$'s are observed. $f_{3n}(\xi_i)$ and $f_{7n}(\xi_i)$ quantify the penalty paid when each $\widetilde{V}_i$ has to be estimated. $f_{4n}(\xi_i)$ and $f_{8n}(\xi_i)$ quantify the penalty paid for having to estimate $\widetilde{V}$ for $\widetilde{\beta}_2\left(\widetilde{V}\right)$. Similarly, $f_{9n}(\xi_i)$ quantifies the penalty paid for having to estimate $\widetilde{\lambda}(c, v)$ when $\widetilde{V}_i$'s are observed, whereas $f_{10n}(\xi_i)$ quantifies the penalty paid when each $\widetilde{V}_i$ should be estimated. Finally, $f_{11n}(\xi_i)$ quantifies the penalty paid for having to estimate $\widetilde{V}$ for $\widetilde{\lambda}\left(c, \widetilde{V}\right)$.

LEMMA A.1. *Define* $\widetilde{V}(x, z) = P\{\mathfrak{X} \le x | Z = z\}$, *and*

$$\widehat{\widetilde{V}}(x, z) = \frac{\frac{1}{nh_0}\sum_{i=1}^n \mathbf{1}\{\mathfrak{X}_i \le x\} K_0\left(\frac{Z_i^c - z^c}{h_0}\right)\mathbf{1}\left\{Z_i^d = z^d\right\}}{\frac{1}{nh_0}\sum_{i=1}^n K_0\left(\frac{Z_i^c - z^c}{h_0}\right)\mathbf{1}\left\{Z_i^d = z^d\right\}}.$$

If Assumptions 2.1–2.4 and 4.1–4.6 hold, as $n \to \infty$,

$$\sup_{x \in \mathfrak{X}_\kappa, z \in \mathcal{Z}_\kappa} \left|\widehat{\widetilde{V}}(x, z) - \widetilde{V}(x, z)\right| = O_p\left(\frac{1}{\sqrt{nh_0}}\right).$$

Moreover,

$$\widehat{\widetilde{V}}(x, z) - \widetilde{V}(x, z) = \frac{1}{nh_0}\sum_{i=1}^n \frac{\mathbf{1}\{\mathfrak{X}_i \le x\} - \widetilde{V}(x, z)}{f\left(z^c, z^d\right)} K_0\left(\frac{Z_i^c - z^c}{h_0}\right)\mathbf{1}\left\{Z_i^d = z^d\right\} + o_p\left(n^{-\frac{1}{2}}\right).$$

Proof. Recall that the number of the endogenous and instrumental variables are both equal to 1, as $Z = \left(Z^c, Z^d\right)$ is a $p_Z \times 1$ vector of instruments for the single endogenous regressor $\mathfrak{X}$, Z^c the single continuous instrument for $\mathfrak{X}$, and Z^d is the $(p_Z - 1)$-dimensional discretely distributed instrumental variable. Using the definition, fixing $x \in \mathfrak{X}_\kappa$ and $z = \left(z^c, z^d\right) \in \mathcal{Z}_\kappa$, we have that

$$\widehat{\widetilde{V}}(x, z) = \frac{\frac{1}{nh_0}\sum_{i=1}^n \mathbf{1}\{\mathfrak{X}_i \le x\} K_0\left(\frac{Z_i^c - z^c}{h_0}\right)\mathbf{1}\left\{Z_i^d = z^d\right\}}{\widehat{f}\left(z^c, z^d\right)}, \tag{1}$$

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

where

$$\widehat{f}\left(z^c,z^d\right)=\frac{1}{nh_0}\sum_{i=1}^{n}K_0\left(\frac{Z_i^c-z^c}{h_0}\right)\mathbf{1}\left\{Z_i^d=z^d\right\}.$$

Note that $\widehat{f}\left(z^c,z^d\right)$ estimates $f\left(z^c,z^d\right)=f\left(z^c|z^d\right)P\left(Z^d=z^d\right)$, where $f\left(z^c|z^d\right)$ is the conditional density of Z^c given z^d evaluated at z^c. Abbreviate $\widehat{f}\left(z^c,z^d\right)$ to $\widehat{f}$ and $f\left(z^c,z^d\right)$ to f. Note that

$$\begin{aligned}\widehat{\widehat{V}}(x,z)&=\frac{\frac{1}{nh_0}\sum_{i=1}^{n}\mathbf{1}\{\mathfrak{X}_i\le x\}K_0\left(\frac{Z_i^c-z^c}{h_0}\right)\mathbf{1}\left\{Z_i^d=z^d\right\}}{f\left(z^c,z^d\right)}\left[\frac{f}{\widehat{\widehat{f}}}\right]\\&=\frac{\frac{1}{nh_0}\sum_{i=1}^{n}\mathbf{1}\{\mathfrak{X}_i\le x\}K_0\left(\frac{Z_i^c-z^c}{h_0}\right)\mathbf{1}\left\{Z_i^d=z^d\right\}}{f\left(z^c,z^d\right)}\left[1-\left(1-\frac{\widehat{f}}{f}\right)\right]^{-1}\\&=\frac{\frac{1}{nh_0}\sum_{i=1}^{n}\mathbf{1}\{\mathfrak{X}_i\le x\}K_0\left(\frac{Z_i^c-z^c}{h_0}\right)\mathbf{1}\left\{Z_i^d=z^d\right\}}{f\left(z^c,z^d\right)}\\&\quad\times\left[1+\left(1-\frac{\widehat{f}}{f}\right)+\left(1-\frac{\widehat{f}}{f}\right)^2+\cdots\right].\end{aligned}$$

We analyze the leading term in the last expansion. Nonleading terms can be handled similarly and have a smaller stochastic order. By a slight abuse of notation, we take

$$\widehat{\widehat{V}}(x,z)=\frac{\frac{1}{nh_0}\sum_{i=1}^{n}\mathbf{1}\{\mathfrak{X}_i\le x\}K_0\left(\frac{Z_i^c-z^c}{h_0}\right)\mathbf{1}\left\{Z_i^d=z^d\right\}}{f\left(z^c,z^d\right)}.$$

As

$$\begin{aligned}&\left|\widehat{\widehat{V}}(x,z)-\widetilde{V}(x,z)\right|\\=&\left|\widehat{\widehat{V}}(x,z)-\mathbb{E}_z\widehat{\widehat{V}}(x,z)+\mathbb{E}_z\widehat{\widehat{V}}(x,z)-\widetilde{V}(x,z)\right|\\\le&\underbrace{\left|\widehat{\widehat{V}}(x,z)-\mathbb{E}_z\widehat{\widehat{V}}(x,z)\right|}_{\dagger_{11}}+\underbrace{\left|\mathbb{E}_z\widehat{\widehat{V}}(x,z)-\widetilde{V}(x,z)\right|}_{\dagger_{12}},\end{aligned}$$

where the expectation $\mathbb{E}_z$ is conditional on $Z=z$. For $\dagger_{12}$, using the standard Taylor expansion of κ_0 implies that the bias term $\mathbb{E}_z\widehat{\widehat{V}}(x,z)-\widetilde{V}(x,z)$ has order $O_p\left(h_0^{\kappa_0}\right)=o_p\left(n^{-\frac{1}{2}}\right)$ as $n\to\infty$, by Assumption 4.4. For $\dagger_{11}$, note that

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

$$\begin{aligned}
&\widehat{\widetilde{V}}(x,z)-\mathbb{E}_z\widehat{\widetilde{V}}(x,z)\\
=&\frac{\frac{1}{nh_0}\sum_{i=1}^n\mathbf{1}\{\mathfrak{X}_i\le x\}K_0\left(\frac{Z_i^c-z^c}{h_0}\right)\mathbf{1}\left\{Z_i^d=z^d\right\}}{f\left(z^c,z^d\right)}\\
-&\frac{\frac{1}{h_0}\mathbb{E}_z\mathbf{1}\{\mathfrak{X}_i\le x\}K_0\left(\frac{Z_i^c-z^c}{h_0}\right)\mathbf{1}\left\{Z_i^d=z^d\right\}}{f\left(z^c,z^d\right)},
\end{aligned}\tag{2}$$

which is a zero-mean empirical process. Using similar arguments, Lemma 3A in Sherman (1994a), we have $\sup_{x\in\mathfrak{X}_\kappa,z\in\mathcal{Z}_\kappa}\left|\widehat{\widetilde{V}}(x,z)-\mathbb{E}_z\widehat{\widetilde{V}}(x,z)\right|=O_p\left(\frac{1}{\sqrt{nh_0}}\right)$ almost surely, so the first part is proved.

Next, we derive the asymptotic representation for $\widehat{\widetilde{V}}-\widetilde{V}$. Let

$$\widehat{t}=\frac{1}{nh_0}\sum_{i=1}^n\mathbf{1}\{\mathfrak{X}_i\le x\}K_0\left(\frac{Z_i^c-z^c}{h_0}\right)\mathbf{1}\left\{Z_i^d=z^d\right\}.$$

We define $\widehat{\widetilde{V}}(x,z)=\frac{\widehat{t}}{\widehat{f}}$. Recall that f denotes the density function for $f\left(z^c,z^d\right)$. Then, applying similar arguments about $\widehat{\widetilde{V}}-\widetilde{V}$ above, we have $\sup_{x\in\mathfrak{X}_\kappa,z\in\mathcal{Z}_\kappa}\left|\widehat{f}-f\right|=O_p\left(\frac{1}{\sqrt{nh_0}}\right)$. Then,

$$\begin{aligned}
&\widehat{\widetilde{V}}(x,z)-\widetilde{V}(x,z)\\
=&\frac{\widehat{t}}{\widehat{f}}-\widetilde{V}\\
=&\frac{1}{f}\left(\widehat{t}-\widetilde{V}\widehat{f}\right)+\left(\frac{\widehat{t}}{\widehat{f}}-\frac{\widehat{t}}{f}-\widetilde{V}+\widetilde{V}\frac{\widehat{f}}{f}\right)\\
=&\underbrace{\frac{1}{f}\left(\widehat{t}-\widetilde{V}\widehat{f}\right)}_{\dagger_{13}}+\underbrace{\frac{1}{f}\left(\frac{\widehat{t}}{\widehat{f}}-\widetilde{V}\right)(f-\widehat{f})}_{\dagger_{14}}.
\end{aligned}$$

For $\dagger_{14}$, using the conclusion above, we have $\dagger_{14}=O_p\left(\frac{1}{nh_0^2}\right)=o_p\left(n^{-\frac{1}{2}}\right)$ provided $h_0^2\gg n^{-\frac{1}{2}}$, and this completes the proof. ■

LEMMA A.2. *Let* $\widehat{\widetilde{\lambda}}(c,v,b)=\underset{a}{\operatorname{argmax}}\widehat{S}_n(a|c,v,b)$, *where*

$$\widehat{S}_n(a|c,v,b)=\frac{1}{nh_2}\sum_{i=1}^n(2D_i(c)-1)G_1\left(\frac{X_i'b+a}{h_1}\right)K_2\left(\frac{\widehat{\widetilde{V}}_i-v}{h_2}\right)t_\kappa\left(Z_i^c\right).$$

If Assumptions 2.1–2.4 and 4.1–4.6 hold, then as $n\to\infty$,

$$\sup_{c\in(\underline{c},\,\overline{c}),\,v\in\widetilde{V}_\kappa,\,b\in\boldsymbol{B}_{2v}}\left|\widehat{\widetilde{\lambda}}(c,v,b)-\widetilde{\lambda}(c,v,b)\right|=o_p(1).$$

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

Proof. First, define the following criterion function:

$$S_n(a|c,v,b) = \frac{1}{nh_2}\sum_{i=1}^{n}(2D_i(c)-1)\,G_1\left(\frac{X_i'b+a}{h_1}\right)K_2\left(\frac{\widetilde{V}_i-v}{h_2}\right)t_\kappa\left(Z_i^c\right),$$

$$\widetilde{S}(a|c,v,b) = \frac{1}{h_2}\mathbb{E}\left\{(2D(c)-1)\,G_1\left(\frac{X'b+a}{h_1}\right)K_2\left(\frac{\widetilde{V}-v}{h_2}\right)t_\kappa\left(Z^c\right)\right\}.$$

By iterated expectations, with the inner expectation over $D(c)$ given $\widetilde{V}$ and Z, we get that

$$\widetilde{S}(a|c,v,b) = \frac{1}{h_2}\mathbb{E}\left\{\mu\left(\widetilde{V},Z,c\right)G_1\left(\frac{X_i'b+a}{h_1}\right)K_2\left(\frac{\widetilde{V}-v}{h_2}\right)t_\kappa\left(Z^c\right)\right\},$$

where the expectation $\mathbb{E}$ is over $\widetilde{V}$ and Z, and

$$\mu\left(\widetilde{V},Z,c\right) = 2P\left(\widetilde{\varepsilon}\left(\widetilde{V}\right) < X'\widetilde{\beta}\left(\widetilde{V}\right)+\widetilde{\lambda}\left(c,\widetilde{V}\right)\middle|\widetilde{V},Z\right)-1.$$

By Assumptions 2.1 and 4.1, $\mathfrak{X} = g\left(Z,F_V^{-1}\left(\widetilde{V}\right)\right)$, and therefore, given $\widetilde{V}$ and Z, $\mathfrak{X}$ is determined. As assumed in equation (1.3), Z contains all the exogenous components of X. It follows that given $\widetilde{V}$ and Z, X is determined. Thus, conditioning on $\widetilde{V}$ and Z in the above term implies conditioning on $\widetilde{V}$ and X. Deduce from Assumption 2.3 that $\mu\left(\widetilde{V},Z,c\right)=0$ whenever $X'\widetilde{\beta}\left(\widetilde{V}\right)+\widetilde{\lambda}\left(c,\widetilde{V}\right)=0$. Next, define

$$\overline{S}(a|c,v,b) = \mathbb{E}\left\{\mu\left(\widetilde{V},Z,c\right)G_1\left(\frac{X'b+a}{h_1}\right)f(v)t_\kappa\left(Z^c\right)\middle|\widetilde{V}=v\right\},$$

$$S(a|c,v,b) = \mathbb{E}\left\{\mu\left(\widetilde{V},Z,c\right)\mathbf{1}\left\{X'b+a>0\right\}f(v)t_\kappa\left(Z^c\right)\middle|\widetilde{V}=v\right\},$$

where $f(\cdot)$ is the density function of v.

To show the uniform consistency, we first show that $\sup|\widehat{S}_n(a|c,v,b)-S(a|c,v,b)| = o_p(1)$, uniformly over $c\in\left(\underline{c},\,\overline{c}\right), v\in\widetilde{V}_\kappa, b\in\mathbf{B}_{2v}$. Note that

$$\begin{aligned}
&|\widehat{S}_n(a|c,v,b)-S(a|c,v,b)|\\
\le\;&\underbrace{|\widehat{S}_n(a|c,v,b)-S_n(a|c,v,b)|}_{\dagger_{21}}+\underbrace{|S_n(a|c,v,b)-\widetilde{S}(a|c,v,b)|}_{\dagger_{22}}\\
+\;&\underbrace{|\widetilde{S}(a|c,v,b)-\overline{S}(a|c,v,b)|}_{\dagger_{23}}+\underbrace{|\overline{S}(a|c,v,b)-S(a|c,v,b)|}_{\dagger_{24}}.
\end{aligned}$$

First, based on the Taylor expansion of $\widehat{\widetilde{V}}_i$ about $\widetilde{V}_i$ and using Lemma A.1, we can show $\dagger_{21}=o_p(1)$. Standard empirical process results (see, for example, Sherman (1994a, Lem. 3A)) show that as $n\to\infty$, $\dagger_{22}$ is $o_p(1)$ uniformly over $c\in\left(\underline{c},\,\overline{c}\right), v\in\widetilde{V}_\kappa, b\in\mathbf{B}_{2v}$. Applying a change of variable argument with $th_2=\widetilde{V}-v$, and by the Taylor expansion of v, we get that as $n\to\infty$, $\dagger_{23}$ is $o_p(1)$ uniformly over $c\in\left(\underline{c},\,\overline{c}\right), v\in\widetilde{V}_\kappa, b\in\mathbf{B}_{2v}$. Finally, applying similar arguments as Lemma 1 in HS, we can show that $\dagger_{24}=o_p(1)$ uniformly over $c\in\left(\underline{c},\,\overline{c}\right), v\in\widetilde{V}_\kappa, b\in\mathbf{B}_{2v}$. Above all, we get that $\widehat{S}_n(a|c,v,b)$ converges in probability to $S(a|c,v,b)$ uniformly over the compact set $c\in\left(\underline{c},\,\overline{c}\right), v\in\widetilde{V}_\kappa, b\in\mathbf{B}_{2v}$. Based on Assumptions 2.3 and 2.4 and the arguments in Horowitz (1998, pp. 59–60), it is straightforward to show that $\widetilde{\lambda}(c,v,b)$ is continuous, and is the unique maximizer of $S(a|c,v,b)$. Then, it follows from

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

Carroll et al.'s (1997) Lemma A.1 that

$$\sup_{c\in(\underline{c},\,\overline{c}),\,v\in\widetilde{V}_\kappa,\,b\in\mathbf{B}_{2v}}\left|\widehat{\widetilde{\lambda}}(c,v,b)-\widetilde{\lambda}(c,v,b)\right|=o_p(1).$$

■

LEMMA A.3. *$\widehat{\widetilde{\lambda}}(c,v,b)$ and $\widetilde{\lambda}(c,v,b)$ are as defined in Lemma A.2. If Assumptions 2.1–2.4 and 4.1–4.6 hold, then as $n\to\infty$,*

$$\sup_{c\in(\underline{c},\,\overline{c}),\,v\in\widetilde{V}_\kappa,\,b\in\boldsymbol{B}_{2v}}\left|\widehat{\widetilde{\lambda}}(c,v,b)-\widetilde{\lambda}(c,v,b)\right|=O_p\left(\frac{1}{\sqrt{nh_0h_1h_2^2}}\right),$$

and

$$\widehat{\widetilde{\lambda}}(c,v,b)-\widetilde{\lambda}(c,v,b)=\alpha_{OO}^{-1}(c,v,b)\widehat{T}_n\left(\widetilde{\lambda}(c,v,b)\,|c,v,b\right)+o_p\left(n^{-\frac{1}{2}}\right),$$

with

$$\widehat{T}_n\left(\widetilde{\lambda}(c,v,b)\,|c,v,b\right)=\frac{1}{nh_1h_2}\sum_{i=1}^{n}(2D_i(c)-1)K_1\left(\frac{X_i'b+\widetilde{\lambda}(c,v,b)}{h_1}\right)K_2\left(\frac{\widehat{\widetilde{V}}_i-v}{h_2}\right)t_\kappa\left(Z_i^c\right),$$

$$\alpha_{OO}(c,v,b)=\int\varphi\left(v,-\widetilde{\lambda}(c,v,b),c,\widetilde{X}\right)dF_{\widetilde{X}},$$

$$\varphi(v,s,c,\widetilde{x})=-\left.\frac{\partial\left\{\mathcal{I}_n(v,t+s,c)f(t+s|\widetilde{x})\right\}}{\partial t}\right|_{t=0},$$

in which $f(\cdot|\widetilde{x})$ denotes the density of $X'b$ given $\widetilde{X}=\widetilde{x}$, and

$$\mathcal{I}_n(v,t,c)=\mathbb{E}\left[t_\kappa\left(Z^c\right)\mu(v,Z,c)f\left(v\middle|Z,X'b\right)\middle|X'b=t\right].$$

Moreover,

$$\sup_{c\in(\underline{c},\,\overline{c}),\,v\in\widetilde{V}_\kappa,\,b\in\boldsymbol{B}_{2v}}\left|\frac{\partial\widehat{\widetilde{\lambda}}(c,v,b)}{\partial b}-\frac{\partial\widetilde{\lambda}(c,v,b)}{\partial b}\right|=O_p\left(\frac{1}{\sqrt{nh_0h_1^2h_2^2}}\right).$$

Proof. Define the gradient and hessian of the criterion function $\widehat{S}_n(a|c,v,b)$ with respect to a, that is,

$$\widehat{T}_n(a|c,v,b)=\frac{1}{nh_1h_2}\sum_{i=1}^{n}(2D_i(c)-1)K_1\left(\frac{X_i'b+a}{h_1}\right)K_2\left(\frac{\widehat{\widetilde{V}}_i-v}{h_2}\right)t_\kappa\left(Z_i^c\right),$$

$$\widehat{H}_n(a|c,v,b)=\frac{1}{nh_1^2h_2}\sum_{i=1}^{n}(2D_i(c)-1)K_1'\left(\frac{X_i'b+a}{h_1}\right)K_2\left(\frac{\widehat{\widetilde{V}}_i-v}{h_2}\right)t_\kappa\left(Z_i^c\right).$$

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

For $\widehat{\widetilde{\lambda}}(c,v,b) = \underset{a}{\operatorname{argmax}} \widehat{S}_n(a|c,v,b)$, the first-order condition means that $\widehat{T}_n(\widehat{\widetilde{\lambda}}(c,v,b)|c,v,b) = 0$. The Taylor expansion of $\widehat{T}_n(\widehat{\widetilde{\lambda}}(c,v,b)|c,v,b)$ about $\widetilde{\lambda}(c,v,b)$ implies that

$$\widehat{\widetilde{\lambda}}(c,v,b) - \widetilde{\lambda}(c,v,b) = \left[-\widehat{H}_n\left(\overline{\lambda}(c,v,b)|c,v,b\right)\right]^{-1} \widehat{T}_n\left(\widetilde{\lambda}(c,v,b)|c,v,b\right), \tag{3}$$

where $\overline{\lambda}(c,v,b)$ lies between $\widehat{\widetilde{\lambda}}(c,v,b)$ and $\widetilde{\lambda}(c,v,b)$. Next, define the gradient and hessian of the criterion function $S_n(a|c,v,b)$ with respect to a, that is,

$$T_n(a|c,v,b) = \frac{1}{nh_1h_2}\sum_{i=1}^{n}(2D_i(c)-1)K_1\left(\frac{X_i'b+a}{h_1}\right)K_2\left(\frac{\widetilde{V}_i-v}{h_2}\right)t_\kappa\left(Z_i^c\right),$$

$$H_n(a|c,v,b) = \frac{1}{nh_1^2h_2}\sum_{i=1}^{n}(2D_i(c)-1)K_1'\left(\frac{X_i'b+a}{h_1}\right)K_2\left(\frac{\widetilde{V}_i-v}{h_2}\right)t_\kappa\left(Z_i^c\right).$$

Note that

$$\widehat{T}_n\left(\widetilde{\lambda}(c,v,b)|c,v,b\right) = \frac{1}{nh_1h_2}\sum_{i=1}^{n}(2D_i(c)-1)K_1\left(\frac{X_i'b+\widetilde{\lambda}(c,v,b)}{h_1}\right)K_2\left(\frac{\widehat{\widetilde{V}}_i-v}{h_2}\right)t_\kappa\left(Z_i^c\right).$$

By applying the Taylor expansion about $\widetilde{V}_i$, we get

$$\widehat{T}_n\left(\widetilde{\lambda}(c,v,b)|c,v,b\right) = T_n\left(\widetilde{\lambda}(c,v,b)|c,v,b\right) + \frac{1}{nh_1h_2}\sum_{i=1}^{n}\zeta_n\left(\widetilde{V}_i,v\right)\left(\widehat{\widetilde{V}}_i-\widetilde{V}_i\right) + \frac{1}{2nh_1h_2}\sum_{i=1}^{n}\zeta_n'\left(\overline{V}_i,v\right)\left(\widehat{\widetilde{V}}_i-\widetilde{V}_i\right)^2,$$

where $\overline{V}_i$ lies between $\widehat{\widetilde{V}}_i$ and $\widetilde{V}_i$, and

$$\begin{aligned}\zeta_n\left(\widetilde{V},v\right) &= \frac{\partial}{\partial\widetilde{V}}\left[(2D(c)-1)K_1\left(\frac{X'b+\widetilde{\lambda}(c,v,b)}{h_1}\right)K_2\left(\frac{\widetilde{V}-v}{h_2}\right)t_\kappa\left(Z^c\right)\right]\\ &= \frac{1}{h_2}(2D(c)-1)K_1\left(\frac{X'b+\widetilde{\lambda}(c,v,b)}{h_1}\right)K_2'\left(\frac{\widetilde{V}-v}{h_2}\right)t_\kappa\left(Z^c\right),\end{aligned}$$

$$\begin{aligned}\zeta_n'\left(\widetilde{V},v\right) &= \frac{\partial^2}{\partial^2\widetilde{V}}\left[(2D(c)-1)K_1\left(\frac{X'b+\widetilde{\lambda}(c,v,b)}{h_1}\right)K_2\left(\frac{\widetilde{V}-v}{h_2}\right)t_\kappa\left(Z^c\right)\right]\\ &= \frac{1}{h_2^2}(2D(c)-1)K_1\left(\frac{X'b+\widetilde{\lambda}(c,v,b)}{h_1}\right)K_2''\left(\frac{\widetilde{V}-v}{h_2}\right)t_\kappa\left(Z^c\right).\end{aligned}$$

Apply Lemma A.1 to get that, uniformly over $c\in(\underline{c},\,\overline{c}), v\in\widetilde{V}_\kappa, b\in\mathbf{B}_{2v}$,

$$\frac{1}{nh_1h_2}\sum_{i=1}^{n}\zeta_n'\left(\overline{V}_i,v\right)\left(\widehat{\widetilde{V}}_i-\widetilde{V}_i\right)^2 = O_p\left(\frac{1}{nh_0^2h_1h_2^3}\right)$$

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

as long as $h_0^2 h_1 h_2^3 \gg n^{-\frac{1}{2}}$, then $\frac{1}{nh_1h_2}\sum_{i=1}^n \zeta_n'\left(\overline{V}_i, v\right)\left(\widehat{\widetilde{V}}_i - \widetilde{V}_i\right)^2 = o_p\left(n^{-\frac{1}{2}}\right)$; therefore, we can write $\widehat{T}_n\left(\widetilde{\lambda}\left(c,v,b\right)|c,v,b\right)$ as

$$\widehat{T}_n\left(\widetilde{\lambda}\left(c,v,b\right)|c,v,b\right) = \underbrace{T_n\left(\widetilde{\lambda}\left(c,v,b\right)|c,v,b\right)}_{\dagger_{31}} + \underbrace{\frac{1}{nh_1h_2}\sum_{i=1}^n \zeta_n\left(\widetilde{V}_i, v\right)\left(\widehat{\widetilde{V}}_i - \widetilde{V}_i\right)}_{\dagger_{32}} + o_p\left(n^{-\frac{1}{2}}\right). \tag{4}$$

First, we analyze $\dagger_{31}$, as

$$T_n\left(\widetilde{\lambda}\left(c,v,b\right)|c,v,b\right) = \left[T_n\left(\widetilde{\lambda}\left(c,v,b\right)|c,v,b\right) - \mathbb{E}T_n\left(\widetilde{\lambda}\left(c,v,b\right)|c,v,b\right)\right] + \mathbb{E}T_n\left(\widetilde{\lambda}\left(c,v,b\right)|c,v,b\right).$$

Noting that the term in brackets is a zero-mean empirical process, standard empirical process arguments show that, uniformly over $c \in \left(\underline{c}, \overline{c}\right), v \in \widetilde{V}_\kappa, b \in \mathbf{B}_{2v}$, as $n \to \infty$,

$$T_n\left(\widetilde{\lambda}\left(c,v,b\right)|c,v,b\right) - \mathbb{E}T_n\left(\widetilde{\lambda}\left(c,v,b\right)|c,v,b\right) = O_p\left(\frac{1}{\sqrt{n}h_1h_2}\right).$$

For the bias term $\mathbb{E}T_n\left(\widetilde{\lambda}\left(c,v,b\right)|c,v,b\right)$, by the definition of $\widetilde{\lambda}(c,v,b)$, and applying the Taylor expansion, we have $\mathbb{E}T_n\left(\widetilde{\lambda}\left(c,v,b\right)|c,v,b\right) = O_p\left(h_1^{\kappa_1} + h_2^{\kappa_2}\right) = o_p\left(n^{-\frac{1}{2}}\right)$, that is, $\dagger_{31} = O_p\left(\frac{1}{\sqrt{n}h_1h_2}\right)$ uniformly over $c \in \left(\underline{c}, \overline{c}\right), v \in \widetilde{V}_\kappa, b \in \mathbf{B}_{2v}$, as $n \to \infty$. Next, for $\dagger_{32}$, applying Lemma A.1, we have $\dagger_{32} = O_p\left(\frac{1}{\sqrt{n}h_0h_1h_2^2}\right)$ uniformly over $c \in \left(\underline{c}, \overline{c}\right), v \in \widetilde{V}_\kappa, b \in \mathbf{B}_{2v}$, as $n \to \infty$. Combined with the results above, we have that uniformly over $c \in \left(\underline{c}, \overline{c}\right), v \in \widetilde{V}_\kappa, b \in \mathbf{B}_{2v}$, as $n \to \infty$,

$$\widehat{T}_n\left(\widetilde{\lambda}\left(c,v,b\right)|c,v,b\right) = O_p\left(\frac{1}{\sqrt{n}h_0h_1h_2^2}\right). \tag{5}$$

Next, consider

$$H_n\left(\widetilde{\lambda}\left(c,v,b\right)|c,v,b\right) = \frac{1}{nh_1^2h_2}\sum_{i=1}^n \left(2D_i\left(c\right)-1\right)K_1'\left(\frac{X_i'b + \widetilde{\lambda}\left(c,v,b\right)}{h_1}\right)K_2\left(\frac{\widetilde{V}_i - v}{h_2}\right)t_\kappa\left(Z_i^c\right). \tag{6}$$

First, we need to get the expectation by integration. Define $H\left(\widetilde{\lambda}\left(c,v,b\right)|c,v,b\right) = \mathbb{E}H_n\left(\widetilde{\lambda}\left(c,v,b\right)|c,v,b\right)$. To get the expectation of $H_n\left(\widetilde{\lambda}\left(c,v,b\right)|c,v,b\right)$, we will evaluate it in five steps: (i) average over $D_i\left(c\right)$ given $\widetilde{V}_i$ and Z_i, (ii) average over $\widetilde{V}_i$ given Z_i and $X_i'b$, (iii) average over Z_i given $X_i'b$, (iv) average over $X_i'b$ given $\widetilde{X}_i$, and (v) average over $\widetilde{X}_i$. First, apply step (i), then (6) equals

$$\frac{1}{h_1^2h_2}\mu\left(\widetilde{V}, Z, c\right)K_1'\left(\frac{X'b + \widetilde{\lambda}\left(c,v,b\right)}{h_1}\right)K_2\left(\frac{\widetilde{V}-v}{h_2}\right)t_\kappa\left(Z^c\right).$$

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

Next, apply step (ii). Making a change of variables $th_2 = \widetilde{V} - v$, the integral above equals

$$\begin{aligned}
&\int \frac{1}{h_1^2 h_2} \mu\left(\widetilde{V}, Z, c\right) K_1'\left(\frac{X'b + \widetilde{\lambda}(c,v,b)}{h_1}\right) K_2\left(\frac{\widetilde{V}-v}{h_2}\right) t_\kappa\left(Z^c\right) f\left(\widetilde{V}|Z, X'b\right) d\widetilde{V} \\
&= \int \frac{1}{h_1^2} \mu\left(v + th_2, Z, c\right) K_1'\left(\frac{X'b + \widetilde{\lambda}(c,v,b)}{h_1}\right) K_2(t)\, t_\kappa\left(Z^c\right) f\left(v + th_2|Z, X'b\right) dt \\
&= \left[\int \mu\left(v + th_2, Z, c\right) f\left(v + th_2|Z, X'b\right) K_2(t)\, dt\right] t_\kappa\left(Z^c\right) \frac{1}{h_1^2} K_1'\left(\frac{X'b + \widetilde{\lambda}(c,v,b)}{h_1}\right) \\
&= \frac{1}{h_1^2} t_\kappa\left(Z^c\right) \mu(v, Z, c) f\left(v|Z, X'b\right) K_1'\left(\frac{X'b + \widetilde{\lambda}(c,v,b)}{h_1}\right) + O_p\left(\frac{h_2^{\kappa_2}}{h_1^2}\right),
\end{aligned}$$

where the last equation comes from the Taylor expansion around v, and the $K_2(\cdot)$ is a κ_2th-order bias kernel. In applying step (iii), write $\mathcal{I}_n\left(v, X'b, c\right)$ for the expectation of $t_\kappa\left(Z^c\right)\mu(v, Z, c) f\left(v|Z, X'b\right)$ over Z, given $X'b$. After applying step (iii), the integrand in (6) becomes

$$\frac{1}{h_1^2} \mathcal{I}_n\left(v, X'b, c\right) K_1'\left(\frac{X'b + \widetilde{\lambda}(c,v,b)}{h_1}\right),$$

where the term of order $h_2^{\kappa_2}/h_1^2$ is higher than $n^{-\frac{1}{2}}$ and has been neglected. Recall that $X = \left(X_1, \widetilde{X}\right)$. Let $R = X'b = X_1 + \widetilde{X}'b_2$, and consider the transformation $\left(X_1, \widetilde{X}\right) \mapsto \left(R, \widetilde{X}\right)$. This transformation is one-to-one, and the Jacobian of the transformation is unity. Let $f\left(x'b|\widetilde{x}\right)$ denote the density of $X'b$ given $\widetilde{X} = \widetilde{x}$ evaluated at $x'b$. Considering this, we apply steps (iv) and (v) together, and then

$$\begin{aligned}
&\int\int \frac{1}{h_1^2} \mathcal{I}_n\left(v, X'b, c\right) K_1'\left(\frac{X'b + \widetilde{\lambda}(c,v,b)}{h_1}\right) f\left(X'b|\widetilde{X}\right) dR dF_{\widetilde{X}}\left(\widetilde{X}\right) \\
&= \int\int \frac{1}{h_1^2} \mathcal{I}_n(v, R, c) K_1'\left(\frac{R + \widetilde{\lambda}(c,v,b)}{h_1}\right) f\left(R|\widetilde{X}\right) dR dF_{\widetilde{X}}\left(\widetilde{X}\right) \\
&= \int\int \frac{1}{h_1} \mathcal{I}_n\left(v, th_1 - \widetilde{\lambda}(c,v,b), c\right) f\left(th_1 - \widetilde{\lambda}(c,v,b)|\widetilde{X}\right) K_1'(t)\, dt dF_{\widetilde{X}}\left(\widetilde{X}\right) \\
&= \int \varphi\left(v, -\widetilde{\lambda}(c,v,b), c, \widetilde{X}\right) dF_{\widetilde{X}}\left(\widetilde{X}\right),
\end{aligned}$$

where $F_{\widetilde{X}}(\cdot)$ denotes the cumulative distribution function of $\widetilde{X}$, and

$$\varphi\left(v, s, c, \widetilde{x}\right) = -\left.\frac{\partial\left\{\mathcal{I}_n(v, t+s, c) f\left(t + s|\widetilde{x}\right)\right\}}{\partial t}\right|_{t=0}.$$

Note that the negative sign in $\varphi\left(v, s, c, \widetilde{x}\right)$ comes from the fact that $\int t K_1'(t)\, dt = -1$. Now, consider $\widehat{H}_n\left(\overline{\lambda}(c,v,b)|c,v,b\right)$ in (3). Note that

$$\begin{aligned}
&\left|\widehat{H}_n\left(\overline{\lambda}(c,v,b)|c,v,b\right) - H\left(\widetilde{\lambda}(c,v,b)|c,v,b\right)\right| \\
&\quad \leq \underbrace{\left|\widehat{H}_n\left(\overline{\lambda}(c,v,b)|c,v,b\right) - H_n\left(\overline{\lambda}(c,v,b)|c,v,b\right)\right|}_{\dagger_{33}}
\end{aligned}$$

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

$$+\underbrace{\left|H_n\left(\overline{\lambda}\left(c,v,b\right)|c,v,b\right)-H\left(\overline{\lambda}\left(c,v,b\right)|c,v,b\right)\right|}_{\dagger_{34}}$$
$$+\underbrace{\left|H\left(\overline{\lambda}\left(c,v,b\right)|c,v,b\right)-H\left(\widetilde{\lambda}\left(c,v,b\right)|c,v,b\right)\right|}_{\dagger_{35}}.$$

For $\dagger_{33}$, after applying the Taylor expansion to $\widetilde{V}_i$, we have

$$\dagger_{33}=\frac{1}{nh_1^2h_2^2}\sum_{i=1}^{n}\left(2D_i\left(c\right)-1\right)K_1'\left(\frac{X_i'b+\overline{\lambda}\left(c,v,b\right)}{h_1}\right)K_2'\left(\frac{\overline{V}_i-v}{h_2}\right)t_\kappa\left(Z_i^c\right)\left(\widehat{\widetilde{V}}_i-\widetilde{V}_i\right),$$

where $\bar{V}_i$ is between $\widehat{\widetilde{V}}_i$ and $\widetilde{V}_i$, for each i. Applying Lemma A.1, we know that $\dagger_{33}$ has order $O_p\left(\frac{1}{\sqrt{n}h_0h_1^2h_2^2}\right)$ uniformly over $c\in\left(\underline{c},\,\overline{c}\right),v\in\widetilde{V}_\kappa,b\in\mathbf{B}_{2v}$, as $n\to\infty$. For $\dagger_{34}$, the difference $H_n\left(\overline{\lambda}_n\left(c,v,b\right)|c,v,b\right)-H\left(\overline{\lambda}\left(c,v,b\right)|c,v,b\right)$ has mean zero for each (c,v,b). Standard empirical process arguments show that $\dagger_{34}$ has order $O_p\left(\frac{1}{\sqrt{n}h_1^2h_2}\right)$ uniformly over $c\in\left(\underline{c},\,\overline{c}\right),v\in\widetilde{V}_\kappa,b\in\mathbf{B}_{2v}$. For $\dagger_{35}$, making the Taylor expansion of each components of $H\left(\overline{\lambda}\left(c,v,b\right)|c,v,b\right)$ about $\widetilde{\lambda}\left(c,v,b\right)$, we can get

$$\begin{aligned}&H\left(\overline{\lambda}\left(c,v,b\right)|c,v,b\right)-H\left(\widetilde{\lambda}\left(c,v,b\right)|c,v,b\right)\\=&\frac{\partial H\left(t|c,v,b\right)}{\partial t}|_{t=\overline{\overline{\lambda}}(c,v,b)}\left(\overline{\lambda}\left(c,v,b\right)-\widetilde{\lambda}\left(c,v,b\right)\right),\end{aligned}$$

where $\overline{\overline{\lambda}}\left(c,v,b\right)$ lies between $\overline{\lambda}\left(c,v,b\right)$ and $\widetilde{\lambda}\left(c,v,b\right)$. By assumption, $\frac{\partial H\left(t|c,v,b\right)}{\partial t}$ is a continuous function on the compact set. Thus, $\dagger_{35}$ has order $o_p\left(1\right)$ uniformly over $c\in\left(\underline{c},\,\overline{c}\right),v\in\widetilde{V}_\kappa,b\in\mathbf{B}_{2v}$ by Lemma A.2. Then, we have that uniformly over $c\in\left(\underline{c},\,\overline{c}\right),v\in\widetilde{V}_\kappa,b\in\mathbf{B}_{2v}$, as $n\to\infty$,

$$\begin{aligned}&\widehat{H}_n\left(\overline{\lambda}\left(c,v,b\right)|c,v,b\right)-H\left(\widetilde{\lambda}\left(c,v,b\right)|c,v,b\right)\\&=O_p\left(\frac{1}{\sqrt{n}h_0h_1^2h_2^2}\right)+O_p\left(\frac{1}{\sqrt{n}h_1^2h_2}\right)+O_p\left(\sup_{c\in\left(\underline{c},\,\overline{c}\right),v\in\widetilde{V}_\kappa,b\in\mathbf{B}_{2v}}\left|\overline{\lambda}\left(c,v,b\right)-\widetilde{\lambda}\left(c,v,b\right)\right|\right)\\&=o_p\left(1\right).\end{aligned}$$

Now, applying the Taylor expansion of $\left[\widehat{H}_n\left(\overline{\lambda}\left(c,v,b\right)|c,v,b\right)\right]^{-1}$ to $H\left(\widetilde{\lambda}\left(c,v,b\right)|c,v,b\right)$, we can get, uniformly over $c\in\left(\underline{c},\,\overline{c}\right),v\in\widetilde{V}_\kappa,b\in\mathbf{B}_{2v}$, as $n\to\infty$,

$$\begin{aligned}&\left[\widehat{H}_n\left(\overline{\lambda}\left(c,v,b\right)|c,v,b\right)\right]^{-1}-\left[H\left(\widetilde{\lambda}\left(c,v,b\right)|c,v,b\right)\right]^{-1}\\&=O_p\left(\frac{1}{\sqrt{n}h_0h_1^2h_2^2}\right)+O_p\left(\frac{1}{\sqrt{n}h_1^2h_2}\right)+O_p\left(\sup_{c\in\left(\underline{c},\,\overline{c}\right),v\in\widetilde{V}_\kappa,b\in\mathbf{B}_{2v}}\left|\overline{\lambda}\left(c,v,b\right)-\widetilde{\lambda}\left(c,v,b\right)\right|\right)\\&=o_p\left(1\right).\end{aligned}\tag{7}$$

Furthermore, the continuity of the inverse function implies that

$$\left[H\left(\widetilde{\lambda}\left(c,v,b\right)|c,v,b\right)\right]^{-1}=O\left(1\right).\tag{8}$$

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

Deduce from (3), (5), and (8), uniformly over $c \in (\underline{c}, \bar{c}), v \in \widetilde{V}_\kappa, b \in \mathbf{B}_{2v}$, as $n \to \infty$,

$$\begin{aligned}\widehat{\widetilde{\lambda}}(c,v,b) - \widetilde{\lambda}(c,v,b) &= \left[-H\left(\widetilde{\lambda}(c,v,b)|c,v,b\right) + o_p(1)\right]^{-1} \widehat{T}_n\left(\widetilde{\lambda}(c,v,b)|c,v,b\right) \\ &= O_p(1)\, \widehat{T}_n\left(\widetilde{\lambda}(c,v,b)|c,v,b\right),\end{aligned}$$

which means

$$\sup_{c\in(\underline{c},\,\bar{c}),\, v\in\widetilde{V}_\kappa, b\in\mathbf{B}_{2v}} \left|\widehat{\widetilde{\lambda}}(c,v,b) - \widetilde{\lambda}(c,v,b)\right| = O_p\left(\frac{1}{\sqrt{n}h_0h_1h_2^2}\right). \tag{9}$$

Applying (9)–(7), we have that uniformly over $c \in (\underline{c}, \bar{c}), v \in \widetilde{V}_\kappa, b \in \mathbf{B}_{2v}$, as $n \to \infty$,

$$\left[\widehat{H}_n\left(\bar{\lambda}(c,v,b)|c,v,b\right)\right]^{-1} = \left[H\left(\widetilde{\lambda}(c,v,b)|c,v,b\right)\right]^{-1} + O_p\left(\frac{1}{\sqrt{n}h_0h_1^2h_2^2}\right). \tag{10}$$

Combining (3), (5), and (10) together, we have

$$\widehat{\widetilde{\lambda}}(c,v,b) - \widetilde{\lambda}(c,v,b) = \left[-H\left(\widetilde{\lambda}(c,v,b)|c,v,b\right)\right]^{-1} \widehat{T}_n\left(\widetilde{\lambda}(c,v,b)|c,v,b\right) + O_p\left(\frac{1}{nh_0^2h_1^3h_2^4}\right). \tag{11}$$

Provided $h_0^2h_1^3h_2^4 \gg n^{-\frac{1}{2}}$, applying (4)–(11), and letting $\alpha_{OO}(c,v,b) = -H(\widetilde{\lambda}(c,v,b)|c,v,b)$, we have that uniformly over $c \in (\underline{c}, \bar{c}), v \in \widetilde{V}_\kappa, b \in \mathbf{B}_{2v}$, as $n \to \infty$,

$$\widehat{\widetilde{\lambda}}(c,v,b) - \widetilde{\lambda}(c,v,b) = \alpha_{OO}^{-1}(c,v,b)\, \widehat{T}_n\left(\widetilde{\lambda}(c,v,b)|c,v,b\right) + o_p\left(n^{-\frac{1}{2}}\right).$$

Next, we analyze $\dfrac{\partial\widehat{\widetilde{\lambda}}(c,v,b)}{\partial b} - \dfrac{\partial\widetilde{\lambda}(c,v,b)}{\partial b}$. Note that

$$\begin{aligned}\frac{\partial\widehat{\widetilde{\lambda}}(c,v,b)}{\partial b} - \frac{\partial\widetilde{\lambda}(c,v,b)}{\partial b} =& \underbrace{\frac{\partial\left[\alpha_{OO}^{-1}(c,v,b)\right]}{\partial b}\widehat{T}_n\left(\widetilde{\lambda}(c,v,b)|c,v,b\right)}_{\dagger_{36}} \\ &+ \underbrace{\alpha_{OO}^{-1}(c,v,b)\frac{\partial\widehat{T}_n\left(\widetilde{\lambda}(c,v,b)|c,v,b\right)}{\partial b}}_{\dagger_{37}} + o_p\left(n^{-\frac{1}{2}}\right).\end{aligned}$$

Combined with (5), $\dagger_{36}$ has order $O_p\left(\dfrac{1}{\sqrt{n}h_0h_1h_2^2}\right)$ uniformly over $c \in (\underline{c}, \bar{c}), v \in \widetilde{V}_\kappa, b \in \mathbf{B}_{2v}$, as $n \to \infty$. Then, for $\dagger_{37}$, note that

$$\begin{aligned}\frac{\partial\widehat{T}_n\left(\widetilde{\lambda}(c,v,b)|c,v,b\right)}{\partial b} =& \frac{1}{nh_1^2h_2}\sum_{i=1}^n (2D_i(c)-1)K_1'\left(\frac{X_i'b + \widetilde{\lambda}(c,v,b)}{h_1}\right) \\ &\times K_2\left(\frac{\widehat{\widetilde{V}}_i - v}{h_2}\right)\left(\widetilde{X}_i + \frac{\partial\widetilde{\lambda}(c,v,b)}{\partial b}\right),\end{aligned}$$

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

as similar discussion of order of $\widehat{T}_n\left(\widetilde{\lambda}\left(c,v,b\right)|c,v,b\right)$, and we can get $\dfrac{\partial \widehat{T}_n\left(\widetilde{\lambda}\left(c,v,b\right)|c,v,b\right)}{\partial b}$ $= O_p\left(\dfrac{1}{\sqrt{n}h_0h_1^2h_2^2}\right)$ uniformly over $c \in \left(\underline{c},\,\overline{c}\right), v \in \widetilde{V}_\kappa, b \in \mathbf{B}_{2v}$, as $n \to \infty$. So

$$\frac{\partial \widehat{\widetilde{\lambda}}\left(c,v,b\right)}{\partial b} - \frac{\partial \widetilde{\lambda}\left(c,v,b\right)}{\partial b} = O_p\left(\frac{1}{\sqrt{n}h_0h_1^2h_2^2}\right),$$

uniformly in $c \in \left(\underline{c},\,\overline{c}\right), v \in \widetilde{V}_\kappa, b \in \mathbf{B}_{2v}$. ■

LEMMA A.4. *If Assumptions 2.1–2.4 and 4.1–4.6 hold, then as* $n \to \infty$,

$$\sup_{v\in\widetilde{V}_\kappa}\left|\widehat{\widetilde{\beta}}_2\left(v\right) - \widetilde{\beta}_2\left(v\right)\right| = o_p\left(1\right).$$

Proof. Recall that $\widetilde{\beta}(v) = \left(1, \widetilde{\beta}_2'(v)\right)'$. So it is sufficient to show that

$$\sup_{v\in\widetilde{V}_\kappa}\left|\widehat{\widetilde{\beta}}\left(v\right) - \widetilde{\beta}\left(v\right)\right| = o_p\left(1\right).$$

Note that $\widehat{\widetilde{\beta}}\left(v\right) = \underset{b}{\operatorname{argmax}}\widehat{S}_n\left(b|v\right)$, where

$$\widehat{S}_n\left(b|v\right) = \int_{\underline{c}}^{\overline{c}} \frac{1}{nh_2}\sum_{i=1}^{n}\left(2D_i\left(c\right)-1\right)G_1\left(\frac{X_i'b + \widehat{\widetilde{\lambda}}\left(c,v,b\right)}{h_1}\right)K_2\left(\frac{\widehat{\widetilde{V}}_i - v}{h_2}\right)t_\kappa\left(Z_i^c\right)\varpi\left(c\right)dc,$$

among which $\widehat{\widetilde{\lambda}}\left(c,v,b\right) = \underset{a}{\operatorname{argmax}}\widehat{S}_n\left(a|c,v,b\right)$, where

$$\widehat{S}_n\left(a|c,v,b\right) = \frac{1}{nh_2}\sum_{i=1}^{n}\left(2D_i\left(c\right)-1\right)G_1\left(\frac{X_i'b + a}{h_1}\right)K_2\left(\frac{\widehat{\widetilde{V}}_i - v}{h_2}\right)t_\kappa\left(Z_i^c\right).$$

Next, define

$$S\left(b|v\right) = \int_{\underline{c}}^{\overline{c}} \mathbb{E}\left\{\mu\left(v,Z,c\right)\mathbf{1}\left\{X'b + \widetilde{\lambda}\left(c,v,b\right) > 0\right\}t_\kappa\left(Z^c\right)f(v)|\widetilde{V} = v\right\}\varpi\left(c\right)dc,$$

among which $\widetilde{\lambda}\left(c,v,b\right) = \underset{a}{\operatorname{argmax}}S\left(a|c,v,b\right)$, where

$$S\left(a|c,v,b\right) = \mathbb{E}\left\{\mu\left(v,Z,c\right)\mathbf{1}\left\{X'b + a > 0\right\}t_\kappa\left(Z^c\right)f(v)|\widetilde{V} = v\right\}.$$

To show uniform consistency, we will take two steps to show that (i) a standard identification, called a strong maximization condition, holds uniformly over $\widetilde{V}_\kappa$ and (ii) $\widehat{S}_n\left(b|v\right)$ converges in probability to $S\left(b|v\right)$ uniformly over the compact set $\widetilde{V}_\kappa \otimes \mathbf{B}_{2v}$. First, start with the uniform strong maximization condition (i), for any $\delta > 0$,

$$\inf_{v\in\widetilde{V}_\kappa}\left[S\left(\widetilde{\beta}\left(v\right)|v\right) - \sup_{\left|b-\widetilde{\beta}(v)\right|\geq\delta} S\left(b|v\right)\right] > 0.$$

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

To establish the uniform condition, we first prove the pointwise result. Fix $v \in \widetilde{V}_\kappa$. We show that $S(b|v)$ is continuous in b on the compact $\mathbf{B}_{2v}$ and is uniquely maximized at $\widetilde{\beta}(v)$. Note that

$$S\left(\widetilde{\lambda}(c,v,b)\,|c,v,b\right) = \mathbb{E}\left\{\mu(v,Z,c)\,\mathbf{1}\left\{X'b+\widetilde{\lambda}(c,v,b)>0\right\}t_\kappa\left(Z^c\right)f(v)|\widetilde{V}=v\right\},$$

first, as $\widetilde{\lambda}(c,v,b)$ is continuous in b, we have $S\left(\widetilde{\lambda}(c,v,b)\,|c,v,b\right)$ is continuous in b by a dominated convergence argument using the almost sure continuity and uniform boundedness of $S\left(\widetilde{\lambda}(c,v,b)\,|c,v,b\right)$, then $S(b|v)=\int_{\underline{c}}^{\overline{c}} S\left(\widetilde{\lambda}(c,v,b)\,|c,v,b\right)\varpi(c)\,dc$ is continuous in b. Note that Z^c is the instrumental variable in the model. So we have that $\widetilde{\varepsilon}(v)$ is independent with Z^c given X and $\widetilde{V}$. Thus, we can derive that

$$med\left(\widetilde{\varepsilon}(v)\,|X=x,\widetilde{V}=v\right) = med\left(\widetilde{\varepsilon}(v)\,|X=x,\widetilde{V}=v,Z^c=z^c\right).$$

Deduce from this fact and Assumption 2.3 that, for each $z^c \in \mathcal{Z}^c_\kappa$, $S(b|v)$ is maximized at $b=\widetilde{\beta}(v)$. Following Manski's (1985) identification argument, we have $\widetilde{\lambda}\left(c,v,\widetilde{\beta}(v)\right)=\widetilde{\lambda}(c,v)$, and

$$S\left(\widetilde{\lambda}(c,v)\,|c,v,\widetilde{\beta}(v)\right) > S\left(\widetilde{\lambda}(c,v,b)\,|c,v,b\right),$$

for any $b \neq \widetilde{\beta}(v)$. Then,

$$S\left(\widetilde{\beta}(v)\,|v\right)-S(b|v)=\int_{\underline{c}}^{\overline{c}}\left[S\left(\widetilde{\lambda}(c,v)\,|c,v,\widetilde{\beta}(v)\right)-S\left(\widetilde{\lambda}(c,v,b)\,|c,v,b\right)\right]\varpi(c)\,dc>0,$$

for any $b \neq \widetilde{\beta}(v)$. The pointwise result is established. Next, we establish the uniform result. Like the discussion about $S(b|v)$ is continuous in b, we can also have that $S(b|v)$ is continuous in v. Since $S(b|v)$ is both continuous in b and v, we have $\widetilde{\beta}(v)$ is an upper hemicontinuous correspondence by the maximum theorem. Moreover, as $\widetilde{\beta}(v)$ uniquely maximizes $S(b|v)$, then $\widetilde{\beta}(v)$ is a continuous function. Next, since $\mathbf{B}_{2v}$ is compact, then the correspondence from v to $\widetilde{\beta}(v)$ is compact-valued. The last two facts and Assumption 4.3 (the correspondence from $v \in \widetilde{V}_\kappa$ to $\mathbf{B}_{2v}$ is upper hemicontinuous) imply that the constraint set $\left\{b \in \mathbf{B}_{2v} : |b-\widetilde{\beta}(v)| \geq \delta\right\}$ is a compact-valued upper hemicontinuous correspondence. This and continuity of $S(b|v)$ in both b and v imply that the constrained value function $\sup_{|b-\widetilde{\beta}(v)|\geq\delta} S(b|v)$ is an upper hemicontinuous function of v. Since the unconstrained value function $S\left(\widetilde{\beta}(v)\,|v\right)$ is continuous in v, then $S\left(\widetilde{\beta}(v)\,|v\right) - \sup_{|b-\widetilde{\beta}(v)|\geq\delta} S(b|v)$ is a lower hemicontinuous function of v. By the Weierstrass theorem, this function must attain its minimum value on the compact set $\widetilde{V}_\kappa$. By the pointwise result, this function is positive for each $v \in \widetilde{V}_\kappa$, then its minimized value must be positive, which establishes condition (i).

To show condition (ii), we first define

$$\widehat{S}_{nh}(b|v)=\int_{\underline{c}}^{\overline{c}}\frac{1}{nh_2}\sum_{i=1}^{n}(2D_i(c)-1)\,G_1\left(\frac{X_i'b+\widetilde{\lambda}(c,v,b)}{h_1}\right)K_2\left(\frac{\widehat{\widetilde{V}}_i-v}{h_2}\right)t_\kappa\left(Z_i^c\right)\varpi(c)\,dc,$$

$$S_n(b|v)=\int_{\underline{c}}^{\overline{c}}\frac{1}{nh_2}\sum_{i=1}^{n}(2D_i(c)-1)\,G_1\left(\frac{X_i'b+\widetilde{\lambda}(c,v,b)}{h_1}\right)K_2\left(\frac{\widetilde{V}_i-v}{h_2}\right)t_\kappa\left(Z_i^c\right)\varpi(c)\,dc,$$

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

$$\overline{S}(b|v)=\int_{\underline{c}}^{\overline{c}}\frac{1}{h_2}\mathbb{E}\left\{(2D(c)-1)G_1\left(\frac{X'b+\widetilde{\lambda}(c,v,b)}{h_1}\right)K_2\left(\frac{\widetilde{V}-v}{h_2}\right)t_\kappa\left(Z_i^c\right)\right\}\varpi(c)\,dc,$$

$$\widetilde{S}(b|v)=\int_{\underline{c}}^{\overline{c}}\mathbb{E}\left\{\mu(v,Z,c)G_1\left(\frac{X'b+\widetilde{\lambda}(c,v,b)}{h_1}\right)f(v)t_\kappa\left(Z^c\right)|\widetilde{V}=v\right\}\varpi(c)\,dc.$$

Then,

$$\begin{aligned}|\widehat{S}_n(b|v)-S(b|v)| &\leq \underbrace{|\widehat{S}_n(b|v)-\widehat{S}_{nh}(b|v)|}_{\dagger_{41}}+\underbrace{|\widehat{S}_{nh}(b|v)-S_n(b|v)|}_{\dagger_{42}}\\ &+ \underbrace{|S_n(b|v)-\overline{S}(b|v)|}_{\dagger_{43}}+\underbrace{|\overline{S}(b|v)-\widetilde{S}(b|v)|}_{\dagger_{44}}\\ &+ \underbrace{|\widetilde{S}(b|v)-S(b|v)|}_{\dagger_{45}}.\end{aligned}$$

First, based on the Taylor expansion about $\widetilde{\lambda}(c,v,b)$ and using Lemma A.3, we have $\dagger_{41}=o_p(1)$ uniformly over $\widetilde{V}_\kappa\otimes\mathbf{B}_{2v}$, as $n\to\infty$. Then, based on the Taylor expansion about $\widetilde{V}_i$ and Lemma A.1, we can show $\dagger_{42}=o_p(1)$ uniformly over $\widetilde{V}_\kappa\otimes\mathbf{B}_{2v}$, as $n\to\infty$. Standard empirical process results show that $\dagger_{43}=o_p(1)$ uniformly over $\widetilde{V}_\kappa\otimes\mathbf{B}_{2v}$, as $n\to\infty$. Applying conditional expectation of $D(c)$ given $\widetilde{V}$ and Z, making a change of variables $th_2=\widetilde{V}-v$, and the Taylor expansion around v, we can get that $\dagger_{44}=o_p(1)$ uniformly over $\widetilde{V}_\kappa\otimes\mathbf{B}_{2v}$, as $n\to\infty$. Finally, following similar arguments in Lemma 1 in HS, we have $\dagger_{45}=o_p(1)$ uniformly over $\widetilde{V}_\kappa\otimes\mathbf{B}_{2v}$, as $n\to\infty$. Combining all the results above, condition (ii) is established, and the proof of Lemma A.4 is completed. ■

LEMMA A.5. *If Assumptions 2.1–2.4 and 4.1–4.6 hold, then as* $n\to\infty$,

$$\begin{aligned}\widehat{\widetilde{\beta}}_2(v)-\widetilde{\beta}_2(v)=&\frac{1}{nh_1}\sum_{i=1}^n\int_{\underline{c}}^{\overline{c}}(2D_i(c)-1)K_1\left(\frac{X_i'\widetilde{\beta}(v)+\widetilde{\lambda}(c,v)}{h_1}\right)\left(\widetilde{X}_i+\frac{\alpha_{XO}(c,v)}{\alpha_{OO}(c,v)}\right)\varpi\\ &\times(c)\,dc\frac{1}{h_2}K_2\left(\frac{\widetilde{V}_i-v}{h_2}\right)t_\kappa\left(Z_i^c\right)[-\sigma(v)]^{-1}\\ &+\frac{1}{nh_1}\sum_{i=1}^n\int_{\underline{c}}^{\overline{c}}(2D_i(c)-1)K_1\left(\frac{X_i'\widetilde{\beta}(v)+\widetilde{\lambda}(c,v)}{h_1}\right)\left(\widetilde{X}_i+\frac{\alpha_{XO}(c,v)}{\alpha_{OO}(c,v)}\right)\varpi\\ &\times(c)\,dc\frac{1}{h_2^2}K_2'\left(\frac{\widetilde{V}_i-v}{h_2}\right)t_\kappa\left(Z_i^c\right)[-\sigma(v)]^{-1}\left(\widehat{\widetilde{V}}_i-\widetilde{V}_i\right)+o_p\left(n^{-\frac{1}{2}}\right),\end{aligned}$$

in which

$$\sigma(v)=\int_{\underline{c}}^{\overline{c}}\left\{\alpha_{XX}(c,v)-\alpha_{OO}(c,v)\left[\frac{\partial\widetilde{\lambda}(c,v,b)}{\partial b}\frac{\partial\widetilde{\lambda}(c,v,b)}{\partial b'}\right]_{b=\widetilde{\beta}(v)}\right\}\omega(c)\,dc$$

and

$$\alpha_{XX}(c,v)=\int\varphi\left(v,\Lambda(c,v)-\widetilde{\beta}_0(v),c,\widetilde{X}\right)\widetilde{X}\widetilde{X}'dF_{\widetilde{X}},$$

$$\alpha_{XO}(c,v)=\int\varphi\left(v,\Lambda(c,v)-\widetilde{\beta}_0(v),c,\widetilde{X}\right)\widetilde{X}dF_{\widetilde{X}},$$

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

$$\alpha_{OO}(c,v)=\int\varphi\left(v,\Lambda(c,v)-\widetilde{\beta}_0(v),c,\widetilde{X}\right)dF_{\widetilde{X}}.$$

Proof. Define the gradient of criterion function $\widehat{S}_n(b|v)$ as $\widehat{T}_n(b|v)$, where

$$\begin{aligned}&\widehat{T}_n(b|v)\\&=\int_{\underline{c}}^{\overline{c}}\frac{1}{nh_1h_2}\sum_{i=1}^{n}(2D_i(c)-1)K_1\left(\frac{X_i'b+\widehat{\widetilde{\lambda}}(c,v,b)}{h_1}\right)K_2\left(\frac{\widehat{\widetilde{V}}_i-v}{h_2}\right)\left(\widetilde{X}_i+\frac{\partial\widehat{\widetilde{\lambda}}(c,v,b)}{\partial b}\right)t_\kappa\left(Z_i^c\right)\varpi(c)\,dc\\&\overset{\dagger_{51}}{=}\int_{\underline{c}}^{\overline{c}}\frac{1}{nh_1h_2}\sum_{i=1}^{n}(2D_i(c)-1)K_1\left(\frac{X_i'b+\widehat{\widetilde{\lambda}}(c,v,b)}{h_1}\right)K_2\left(\frac{\widehat{\widetilde{V}}_i-v}{h_2}\right)\widetilde{X}_it_\kappa\left(Z_i^c\right)\varpi(c)\,dc,\end{aligned}$$

where $\dagger_{51}$ follows from Lemma A.3 that $\widehat{T}_n\left(\widehat{\widetilde{\lambda}}(c,v,b)\,|c,v,b\right)=0$, as well as

$$\frac{1}{nh_1h_2}\sum_{i=1}^{n}(2D_i(c)-1)K_1\left(\frac{X_i'b+\widehat{\widetilde{\lambda}}(c,v,b)}{h_1}\right)K_2\left(\frac{\widehat{\widetilde{V}}_i-v}{h_2}\right)t_\kappa\left(Z_i^c\right)=0. \tag{12}$$

According to $\widehat{T}_n(b|v)$, define the hessian of the criterion function $\widehat{S}_n(b|v)$ as $\widehat{H}_n(b|v)$, where

$$\begin{aligned}&\widehat{H}_n(b|v)\\&=\int_{\underline{c}}^{\overline{c}}\frac{1}{nh_1^2h_2}\sum_{i=1}^{n}(2D_i(c)-1)K_1'\left(\frac{X_i'b+\widehat{\widetilde{\lambda}}(c,v,b)}{h_1}\right)K_2\left(\frac{\widehat{\widetilde{V}}_i-v}{h_2}\right)\widetilde{X}_i\left(\widetilde{X}_i+\frac{\partial\widehat{\widetilde{\lambda}}(c,v,b)}{\partial b}\right)'t_\kappa\left(Z_i^c\right)\varpi(c)\,dc\\&\overset{\dagger_{52}}{=}\int_{\underline{c}}^{\overline{c}}\frac{1}{nh_1^2h_2}\sum_{i=1}^{n}(2D_i(c)-1)K_1'\left(\frac{X_i'b+\widehat{\widetilde{\lambda}}(c,v,b)}{h_1}\right)K_2\left(\frac{\widehat{\widetilde{V}}_i-v}{h_2}\right)\\&\times\left(\widetilde{X}_i\widetilde{X}_i'-\frac{\partial\widehat{\widetilde{\lambda}}(c,v,b)}{\partial b}\frac{\partial\widehat{\widetilde{\lambda}}(c,v,b)}{\partial b'}\right)t_\kappa\left(Z_i^c\right)\varpi(c)\,dc,\end{aligned}$$

where $\dagger_{52}$ follows from differentiating (12) with respect to b, that is,

$$\frac{1}{nh_1^2h_2}\sum_{i=1}^{n}(2D_i(c)-1)K_1'\left(\frac{X_i'b+\widehat{\widetilde{\lambda}}(c,v,b)}{h_1}\right)K_2\left(\frac{\widehat{\widetilde{V}}_i-v}{h_2}\right)\left(\widetilde{X}_i+\frac{\partial\widehat{\widetilde{\lambda}}(c,v,b)}{\partial b}\right)t_\kappa\left(Z_i^c\right)=0.$$

The Taylor expansion of $\widehat{\widetilde{\beta}}(v)$ about the true value $\widetilde{\beta}(v)$ gives that

$$\widehat{\widetilde{\beta}}_2(v)-\widetilde{\beta}_2(v)=\left[-\widehat{H}_n\left(\overline{\beta}(v)\,|v\right)\right]^{-1}\widehat{T}_n\left(\widetilde{\beta}(v)\,|v\right), \tag{13}$$

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

where $\overline{\beta}(v)$ lies between $\widehat{\widetilde{\beta}}(v)$ and $\widetilde{\beta}(v)$. First, we analyze $\widehat{T}_n\left(\widetilde{\beta}(v)|v\right)$, and the representation is given by

$$
\begin{aligned}
&\widehat{T}_n\left(\widetilde{\beta}(v)|v\right)\\
&=\int_{\underline{c}}^{\overline{c}}\frac{1}{nh_1h_2}\sum_{i=1}^{n}(2D_i(c)-1)K_1\left(\frac{X_i'\widetilde{\beta}(v)+\widehat{\widetilde{\lambda}}\left(c,v,\widetilde{\beta}(v)\right)}{h_1}\right)K_2\left(\frac{\widehat{\widetilde{V}}_i-v}{h_2}\right)\widetilde{X}_i t_\kappa\left(Z_i^c\right)\varpi(c)\,dc\\
&\overset{\dagger_{53}}{=}\int_{\underline{c}}^{\overline{c}}\frac{1}{nh_1h_2}\sum_{i=1}^{n}(2D_i(c)-1)K_1\left(\frac{X_i'\widetilde{\beta}(v)+\widetilde{\lambda}(c,v)}{h_1}\right)K_2\left(\frac{\widehat{\widetilde{V}}_i-v}{h_2}\right)\widetilde{X}_i t_\kappa\left(Z_i^c\right)\varpi(c)\,dc\\
&+\int_{\underline{c}}^{\overline{c}}\frac{1}{nh_1^2h_2}\sum_{i=1}^{n}(2D_i(c)-1)K_1'\left(\frac{X_i'\widetilde{\beta}(v)+\overline{\lambda}\left(c,v,\widetilde{\beta}(v)\right)}{h_1}\right)K_2\left(\frac{\widehat{\widetilde{V}}_i-v}{h_2}\right)\widetilde{X}_i t_\kappa\left(Z_i^c\right)\\
&\times\left(\widehat{\widetilde{\lambda}}\left(c,v,\widetilde{\beta}(v)\right)-\widetilde{\lambda}(c,v)\right)\varpi(c)\,dc\\
&\overset{\dagger_{54}}{=}\int_{\underline{c}}^{\overline{c}}\frac{1}{nh_1h_2}\sum_{i=1}^{n}(2D_i(c)-1)K_1\left(\frac{X_i'\widetilde{\beta}(v)+\widetilde{\lambda}(c,v)}{h_1}\right)K_2\left(\frac{\widehat{\widetilde{V}}_i-v}{h_2}\right)t_\kappa\left(Z_i^c\right)\\
&\times\left(\widetilde{X}_i+\frac{\alpha_{OO}^{-1}\left(c,v,\widetilde{\beta}(v)\right)}{nh_1^2h_2}\sum_{i=1}^{n}(2D_i(c)-1)K_1'\left(\frac{X_i'\widetilde{\beta}(v)+\overline{\lambda}\left(c,v,\widetilde{\beta}(v)\right)}{h_1}\right)K_2\left(\frac{\widehat{\widetilde{V}}_i-v}{h_2}\right)\widetilde{X}_i t_\kappa\left(Z_i^c\right)\right)\\
&\times\varpi(c)\,dc+o_p\left(n^{-\frac{1}{2}}\right)\\
&\overset{\dagger_{55}}{=}\int_{\underline{c}}^{\overline{c}}\frac{1}{nh_1h_2}\sum_{i=1}^{n}(2D_i(c)-1)K_1\left(\frac{X_i'\widetilde{\beta}(v)+\widetilde{\lambda}(c,v)}{h_1}\right)K_2\left(\frac{\widehat{\widetilde{V}}_i-v}{h_2}\right)t_\kappa\left(Z_i^c\right)\\
&\times\left(\widetilde{X}_i+\frac{\alpha_{XO}(c,v)}{\alpha_{OO}(c,v)}\right)\varpi(c)\,dc+o_p\left(n^{-\frac{1}{2}}\right),
\end{aligned}
$$

where $\dagger_{53}$ follows from the Taylor expansion of $\widehat{\widetilde{\lambda}}\left(c,v,\widetilde{\beta}(v)\right)$ around $\widetilde{\lambda}(c,v)$. Recall that $\widetilde{\lambda}\left(c,v,\widetilde{\beta}(v)\right)=\widetilde{\lambda}(c,v)=\widetilde{\beta}_0(v)-\Lambda(c,v)$. $\dagger_{54}$ follows from Lemma A.3 about the asymptotic expression for $\widehat{\widetilde{\lambda}}\left(c,v,\widetilde{\beta}(v)\right)-\widetilde{\lambda}(c,v)$. For $\dagger_{55}$, first note that

$$
\begin{aligned}
&\alpha_{OO}(c,v)=\alpha_{OO}\left(c,v,\widetilde{\beta}(v)\right)\\
&=\int\varphi\left(v,\Lambda(c,v)-\widetilde{\beta}_0(v),c,\widetilde{X}\right)dF_{\widetilde{X}}.
\end{aligned}
$$

After applying similar arguments in Lemma A.3 about $\widehat{H}_n\left(\overline{\lambda}(c,v,b)|c,v,b\right)$, we have

$$
\begin{aligned}
&\frac{1}{nh_1^2h_2}\sum_{i=1}^{n}(2D_i(c)-1)K_1'\left(\frac{X_i'\widetilde{\beta}(v)+\overline{\lambda}\left(c,v,\widetilde{\beta}(v)\right)}{h_1}\right)K_2\left(\frac{\widehat{\widetilde{V}}_i-v}{h_2}\right)\widetilde{X}_i t_\kappa\left(Z_i^c\right)\\
&=\int\varphi\left(v,\Lambda(c,v)-\widetilde{\beta}_0(v),c,\widetilde{X}\right)dF_{\widetilde{X}}+O_p\left(\frac{1}{\sqrt{n}h_0h_1^2h_2^2}\right)=\alpha_{XO}(c,v)+O_p\left(\frac{1}{\sqrt{n}h_0h_1^2h_2^2}\right).
\end{aligned}
$$

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

Together combining the result that

$$\frac{1}{nh_1h_2}\sum_{i=1}^{n}(2D_i(c)-1)K_1\left(\frac{X_i'\widetilde{\beta}(v)+\widetilde{\lambda}(c,v)}{h_1}\right)K_2\left(\frac{\widehat{\widetilde{V}}_i-v}{h_2}\right)t_\kappa\left(Z_i^c\right)$$
$$=\widehat{T}_n\left(\widetilde{\lambda}(c,v)\,|c,v,\widetilde{\beta}(v)\right)=O_p\left(\frac{1}{\sqrt{nh_0h_1h_2^2}}\right),$$

$\dagger_{55}$ is established. Following similar arguments in Lemma A.3 about $\widehat{T}_n\left(\widetilde{\lambda}(c,v,b)\,|c,v,b\right)$, we have that uniformly over $v\in\widetilde{V}_\kappa$, as $n\to\infty$,

$$\widehat{T}_n\left(\widetilde{\beta}(v)\,|v\right)=O_p\left(\frac{1}{\sqrt{nh_0h_1h_2^2}}\right). \tag{14}$$

Moreover, applying the Taylor expansion to $\widetilde{V}_i$ in each i, we can establish that

$$\begin{aligned}
&\widehat{T}_n\left(\widetilde{\beta}(v)\,|v\right)\\
&=\frac{1}{nh_1}\sum_{i=1}^{n}\int_{\underline{c}}^{\overline{c}}(2D_i(c)-1)K_1\left(\frac{X_i'\widetilde{\beta}(v)+\widetilde{\lambda}(c,v)}{h_1}\right)\left(\widetilde{X}_i+\frac{\alpha_{XO}(c,v)}{\alpha_{OO}(c,v)}\right)\varpi(c)\,dc\\
&\times\frac{1}{h_2}K_2\left(\frac{\widetilde{V}_i-v}{h_2}\right)t_\kappa\left(Z_i^c\right)\\
&+\frac{1}{nh_1}\sum_{i=1}^{n}\int_{\underline{c}}^{\overline{c}}(2D_i(c)-1)K_1\left(\frac{X_i'\widetilde{\beta}(v)+\widetilde{\lambda}(c,v)}{h_1}\right)\left(\widetilde{X}_i+\frac{\alpha_{XO}(c,v)}{\alpha_{OO}(c,v)}\right)\varpi(c)\,dc\\
&\times\frac{1}{h_2^2}K_2'\left(\frac{\widetilde{V}_i-v}{h_2}\right)t_\kappa\left(Z_i^c\right)\left(\widehat{\widetilde{V}}_i-\widetilde{V}_i\right)+o_p\left(n^{-\frac{1}{2}}\right),
\end{aligned}$$

where the second-order expansion follows from the fact that $O_p\left(\frac{1}{h_2^3}\left(\widehat{\widetilde{V}}_i-\widetilde{V}_i\right)^2\right)=O_p\left(\frac{1}{nh_0^2h_2^3}\right)=o_p\left(n^{-\frac{1}{2}}\right)$. Next, we analyze $\widehat{H}_n\left(\overline{\beta}(v)\,|v\right)$. According to the derivation of $\widehat{H}_n(b|v)$ above, we can establish the $\widehat{H}_n\left(\overline{\beta}(v)\,|v\right)$, and it is given by

$$\begin{aligned}
\widehat{H}_n\left(\overline{\beta}(v)\,|v\right)&=\int_{\underline{c}}^{\overline{c}}\frac{1}{nh_1^2h_2}\sum_{i=1}^{n}(2D_i(c)-1)K_1'\left(\frac{X_i'\overline{\beta}(v)+\widehat{\widetilde{\lambda}}\left(c,v,\overline{\beta}(v)\right)}{h_1}\right)K_2\left(\frac{\widehat{\widetilde{V}}_i-v}{h_2}\right)\\
&\times\left(\widetilde{X}_i\widetilde{X}_i'-\left[\frac{\partial\widehat{\widetilde{\lambda}}(c,v,b)}{\partial b}\frac{\partial\widehat{\widetilde{\lambda}}(c,v,b)}{\partial b'}\right]_{b=\overline{\beta}(v)}\right)t_\kappa\left(Z_i^c\right)\varpi(c)\,dc\\
&\overset{\dagger_{56}}{=}\int_{\underline{c}}^{\overline{c}}\frac{1}{nh_1^2h_2}\sum_{i=1}^{n}(2D_i(c)-1)K_1'\left(\frac{X_i'\overline{\beta}(v)+\widetilde{\lambda}\left(c,v,\overline{\beta}(v)\right)}{h_1}\right)K_2\left(\frac{\widehat{\widetilde{V}}_i-v}{h_2}\right)\\
&\times\left(\widetilde{X}_i\widetilde{X}_i'-\left[\frac{\partial\widetilde{\lambda}(c,v,b)}{\partial b}\frac{\partial\widetilde{\lambda}(c,v,b)}{\partial b'}\right]_{b=\overline{\beta}(v)}\right)t_\kappa\left(Z_i^c\right)\varpi(c)\,dc+o_p(1)
\end{aligned}$$

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

$$\overset{\dagger_{57}}{=}\int_{\underline{c}}^{\overline{c}}\frac{1}{nh_1^2h_2}\sum_{i=1}^{n}\left(2D_i(c)-1\right)K_1'\left(\frac{X_i'\widetilde{\beta}(v)+\widetilde{\lambda}(c,v)}{h_1}\right)K_2\left(\frac{\widehat{\widetilde{V}}_i-v}{h_2}\right)$$
$$\times\left(\widetilde{X}_i\widetilde{X}_i'-\left[\frac{\partial\widetilde{\lambda}(c,v,b)}{\partial b}\frac{\partial\widetilde{\lambda}(c,v,b)}{\partial b'}\right]_{b=\widetilde{\beta}(v)}\right)t_\kappa\left(Z_i^c\right)\varpi(c)\,dc+o_p(1)$$
$$\overset{\dagger_{58}}{=}\sigma(v)+o_p(1),$$

where $\dagger_{56}$ follows from Lemma A.3 and $\dagger_{57}$ follows from Lemma A.4. For $\dagger_{58}$, applying the argument similar to $\dagger_{55}$, we have

$$\frac{1}{nh_1^2h_2}\sum_{i=1}^{n}\left(2D_i(c)-1\right)K_1'\left(\frac{X_i'\widetilde{\beta}(v)+\widetilde{\lambda}\left(c,v,\widetilde{\beta}(v)\right)}{h_1}\right)K_2\left(\frac{\widehat{\widetilde{V}}_i-v}{h_2}\right)\widetilde{X}_i\widetilde{X}_i't_\kappa\left(Z_i^c\right)$$
$$=\int\varphi\left(v,\Lambda(c,v)-\widetilde{\beta}_0(v),c,\widetilde{X}\right)\widetilde{X}\widetilde{X}'dF_{\widetilde{X}}+O_p\left(\frac{1}{\sqrt{n}h_0h_1^2h_2^2}\right)$$
$$=\alpha_{XX}(c,v)+o_p(1),$$

which makes that

$$\sigma(v)=\int_{\underline{c}}^{\overline{c}}\left\{\alpha_{XX}(c,v)-\alpha_{OO}(c,v)\left[\frac{\partial\widetilde{\lambda}(c,v,b)}{\partial b}\frac{\partial\widetilde{\lambda}(c,v,b)}{\partial b'}\right]_{b=\widetilde{\beta}(v)}\right\}\varpi(c)\,dc.$$

In summary, we establish

$$\widehat{\widetilde{\beta}}_2(v)-\widetilde{\beta}_2(v)=\frac{1}{nh_1}\sum_{i=1}^{n}\int_{\underline{c}}^{\overline{c}}\left(2D_i(c)-1\right)K_1\left(\frac{X_i'\widetilde{\beta}(v)+\widetilde{\lambda}(c,v)}{h_1}\right)\left(\widetilde{X}_i+\frac{\alpha_{XO}(c,v)}{\alpha_{OO}(c,v)}\right)\varpi$$
$$\times(c)\,dc\frac{1}{h_2}K_2\left(\frac{\widetilde{V}_i-v}{h_2}\right)t_\kappa\left(Z_i^c\right)[-\sigma(v)]^{-1}$$
$$+\frac{1}{nh_1}\sum_{i=1}^{n}\int_{\underline{c}}^{\overline{c}}\left(2D_i(c)-1\right)K_1\left(\frac{X_i'\widetilde{\beta}(v)+\widetilde{\lambda}(c,v)}{h_1}\right)\left(\widetilde{X}_i+\frac{\alpha_{XO}(c,v)}{\alpha_{OO}(c,v)}\right)\varpi$$
$$\times(c)\,dc\frac{1}{h_2^2}K_2'\left(\frac{\widetilde{V}_i-v}{h_2}\right)t_\kappa\left(Z_i^c\right)[-\sigma(v)]^{-1}\left(\widehat{\widetilde{V}}_i-\widetilde{V}_i\right)+o_p\left(n^{-\frac{1}{2}}\right).$$

■

LEMMA A.6. *If Assumptions 2.1–2.4 and 4.1–4.6 hold, then as* $n\to\infty$,

$$\widehat{\widetilde{\beta}}_2\left(\widehat{\widetilde{V}}_i\right)-\widehat{\widetilde{\beta}}_2\left(\widetilde{V}_i\right)=\frac{1}{n}\sum_{j=1}^{n}\delta\left(\widetilde{V}_i\right)\frac{\mathbf{1}\left\{\mathfrak{X}_j\le\mathfrak{X}_i\right\}-\widetilde{V}_i}{f\left(Z_i^c,Z_i^d\right)}\frac{1}{h_0}K_0\left(\frac{Z_j^c-Z_i^c}{h_0}\right)\mathbf{1}\left\{Z_j^d=Z_i^d\right\}+o_p\left(n^{-\frac{1}{2}}\right),$$

where $\delta(v)$ *is the partial derivative of* $\widetilde{\beta}_2(v)$ *with respect to* v.

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

Proof. Let $\widehat{\delta}(v)$ as the partial derivative of $\widehat{\widetilde{\beta}}_2(v)$ with respect to v. We can show that

$$\begin{aligned}\widehat{\widetilde{\beta}}_2\left(\widehat{\widetilde{V}}_i\right)-\widehat{\widetilde{\beta}}_2\left(\widetilde{V}_i\right) =&\widehat{\delta}\left(\overline{V}_i\right)\left(\widehat{\widetilde{V}}_i-\widetilde{V}_i\right)\\ =&\underbrace{\delta\left(\widetilde{V}_i\right)\left(\widehat{\widetilde{V}}_i-\widetilde{V}_i\right)}_{\dagger_{61}}+\underbrace{\left[\widehat{\delta}\left(\overline{V}_i\right)-\delta\left(\overline{V}_i\right)\right]\left(\widehat{\widetilde{V}}_i-\widetilde{V}_i\right)}_{\dagger_{62}}+\underbrace{\left[\delta\left(\overline{V}_i\right)-\delta\left(\widetilde{V}_i\right)\right]\left(\widehat{\widetilde{V}}_i-\widetilde{V}_i\right)}_{\dagger_{63}}.\end{aligned}$$

We will show that both $\dagger_{62}$ and $\dagger_{63}$ are of order $o_p\left(n^{-\frac{1}{2}}\right)$ uniformly over $v\in\widetilde{V}_\kappa$, as $n\to\infty$. First, consider $\dagger_{63}$. Let $\gamma(v)$ denote the first-order derivative of $\delta(v)$ with respect to v. Taking the Taylor expansion of $\delta\left(\overline{V}_i\right)$ about $\widetilde{V}_i$, we have

$$\left[\delta\left(\overline{V}_i\right)-\delta\left(\widetilde{V}_i\right)\right]\left(\widehat{\widetilde{V}}_i-\widetilde{V}_i\right)=\gamma\left(\overline{\overline{V}}_i\right)\left(\overline{V}_i-\widetilde{V}_i\right)\left(\widehat{\widetilde{V}}_i-\widetilde{V}_i\right),$$

where $\overline{\overline{V}}_i$ lies between $\overline{V}_i$ and $\widetilde{V}_i$. Because $\gamma(\cdot)$ is a continuous function on the compact set, then $\gamma(v)$ is uniformly bounded. It immediately follows from Lemma A.1 that $\dagger_{63}$ is bounded by $O_p\left(\frac{1}{nh_0^2}\right)=o_p\left(n^{-\frac{1}{2}}\right)$ provided $h_0^2\gg n^{-\frac{1}{2}}$.

Next, we analyze $\dagger_{62}$. Let $\widehat{\gamma}(v)$ denote the first-order derivative of $\widehat{\delta}(v)$ with respect to v. By taking the Taylor expansion of $\widehat{\delta}\left(\overline{V}_i\right)-\delta\left(\overline{V}_i\right)$ to $\widetilde{V}_i$, we get that

$$\left[\widehat{\delta}\left(\overline{V}_i\right)-\delta\left(\overline{V}_i\right)\right]\left(\widehat{\widetilde{V}}_i-\widetilde{V}_i\right)=\underbrace{\left[\widehat{\delta}\left(\widetilde{V}_i\right)-\delta\left(\widetilde{V}_i\right)\right]\left(\widehat{\widetilde{V}}_i-\widetilde{V}_i\right)}_{\dagger_{64}}+\underbrace{\left[\widehat{\gamma}\left(\overline{\overline{V}}_i\right)-\gamma\left(\overline{\overline{V}}_i\right)\right]\left(\overline{V}_i-\widetilde{V}_i\right)\left(\widehat{\widetilde{V}}_i-\widetilde{V}_i\right)}_{\dagger_{65}},$$

where $\overline{\overline{V}}_i$ lies between $\overline{V}_i$ and $\widetilde{V}_i$. Start with $\dagger_{65}$. Applying the argument similar to $\dagger_{63}$, $\dagger_{65}$ is bounded by $O_p\left(\frac{1}{nh_0^2}\right)=o_p\left(n^{-\frac{1}{2}}\right)$. Now, consider $\dagger_{64}$. Differentiating both sides of (13) with respect to v gives that

$$\begin{aligned}\widehat{\delta}(v)-\delta(v)=&\frac{\partial}{\partial v}\left\{\left[-\widehat{H}_n\left(\overline{\beta}(v)\,|v\right)\right]^{-1}\widehat{T}_n\left(\widetilde{\beta}(v)\,|v\right)\right\}\\ =&\underbrace{\left[-\widehat{H}_n\left(\overline{\beta}(v)\,|v\right)\right]^{-1}\left[\frac{\partial}{\partial v}\widehat{T}_n\left(\widetilde{\beta}(v)\,|v\right)\right]}_{\dagger_{66}}\\ &+\underbrace{\frac{\partial}{\partial v}\left\{\left[-\widehat{H}_n\left(\overline{\beta}(v)\,|v\right)\right]^{-1}\right\}\widehat{T}_n\left(\widetilde{\beta}(v)\,|v\right)}_{\dagger_{67}}.\end{aligned}$$

Start with $\dagger_{66}$. As argued in Lemma A.5, we can derive that $\left[-\widehat{H}_n\left(\overline{\beta}(v)\,|v\right)\right]^{-1}=O(1)$. From (14), we know that $\widehat{T}_n\left(\widetilde{\beta}(v)\,|v\right)=O_p\left(\frac{1}{\sqrt{n}h_0h_2^2}\right)$ uniformly over $v\in\widetilde{V}_\kappa$, as $n\to\infty$. Applying the argument similar to Lemma A.3, we know that $\frac{\partial}{\partial v}\widehat{T}_n\left(\widetilde{\beta}(v)\,|v\right)=O_p\left(\frac{1}{\sqrt{n}h_0h_2^3}\right)$ uniformly over $v\in\widetilde{V}_\kappa$, as $n\to\infty$. Then, combining $\left(\widehat{\widetilde{V}}_i-\widetilde{V}_i\right)$ in

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

$\dagger_{64}$, we know the contribution for $\dagger_{64}$ to $\dagger_{62}$ is $O_p\left(\frac{1}{nh_0^2h_2^3}\right) = o_p\left(n^{-\frac{1}{2}}\right)$ provided $h_0^2h_2^3 \gg n^{-\frac{1}{2}}$. Next, for $\dagger_{67}$. Because $\left[-\widehat{H}_n\left(\overline{\beta}\left(v\right)|v\right)\right]^{-1}$ is continuously differentiable in v, $\frac{\partial}{\partial v}\left\{\left[-\widehat{H}_n\left(\overline{\beta}\left(v\right)|v\right)\right]^{-1}\right\}$ is continuous in v on the compact set and so is bounded. Then, we can derive the contribution for $\dagger_{67}$ combined with $\left(\widehat{\widetilde{V}}_i - \widetilde{V}_i\right)$ in $\dagger_{64}$ to $\dagger_{62}$ is $O_p\left(\frac{1}{nh_0^2h_2^2}\right) = o_p\left(n^{-\frac{1}{2}}\right)$ provided $h_0^2h_2^2 \gg n^{-\frac{1}{2}}$. Thus, $\dagger_{62} = o_p\left(n^{-\frac{1}{2}}\right)$ uniformly over $v \in \widetilde{V}_\kappa$, as $n \to \infty$.

Now, let us discuss $\dagger_{61}$. Recall that

$$\widehat{\widetilde{V}}_i - \widetilde{V}_i = \frac{1}{nh_0}\sum_{j=1}^{n}\frac{\mathbf{1}\left\{\mathfrak{X}_j \le \mathfrak{X}_i\right\} - \widetilde{V}_i}{f\left(Z_i^c, Z_i^d\right)}K_0\left(\frac{Z_j^c - Z_i^c}{h_0}\right)\mathbf{1}\left\{Z_j^d = Z_i^d\right\} + o_p\left(n^{-\frac{1}{2}}\right).$$

Substitute the above representation into $\delta\left(\widetilde{V}_i\right)\left(\widehat{\widetilde{V}}_i - \widetilde{V}_i\right)$, where

$$\begin{aligned}&\delta(\widetilde{V}_i)(\widehat{\widetilde{V}}_i - \widetilde{V}_i)\\ &= \frac{1}{n}\sum_{j=1}^{n}\delta\left(\widetilde{V}_i\right)\frac{\mathbf{1}\left\{\mathfrak{X}_j \le \mathfrak{X}_i\right\} - \widetilde{V}_i}{f\left(Z_i^c, Z_i^d\right)}\frac{1}{h_0}K_0\left(\frac{Z_j^c - Z_i^c}{h_0}\right)\mathbf{1}\left\{Z_j^d = Z_i^d\right\} + o_p\left(n^{-\frac{1}{2}}\right).\end{aligned}$$

Combing the results above, we can show that

$$\begin{aligned}\widehat{\delta}\left(\overline{V}_i\right)\left(\widehat{\widetilde{V}}_i - \widetilde{V}_i\right) &= \delta\left(\widetilde{V}_i\right)\left(\widehat{\widetilde{V}}_i - \widetilde{V}_i\right) + o_p\left(n^{-\frac{1}{2}}\right)\\ &= \frac{1}{n}\sum_{j=1}^{n}\delta\left(\widetilde{V}_i\right)\frac{\mathbf{1}\left\{\mathfrak{X}_j \le \mathfrak{X}_i\right\} - \widetilde{V}_i}{f\left(Z_i^c, Z_i^d\right)}\frac{1}{h_0}K_0\left(\frac{Z_j^c - Z_i^c}{h_0}\right)\mathbf{1}\left\{Z_j^d = Z_i^d\right\}\\ &\quad + o_p\left(n^{-\frac{1}{2}}\right).\end{aligned}$$

■

LEMMA A.7. *If Assumptions 2.1–2.4 and 4.1–4.6 hold, then as* $n \to \infty$,

$$\begin{aligned}&\widehat{\widetilde{\lambda}}\left(c, v\right) - \widetilde{\lambda}\left(c, v\right)\\ &= \frac{1}{nh_1h_2\alpha_{OO}\left(c, v\right)}\sum_{i=1}^{n}\left(2D_i\left(c\right) - 1\right)K_1\left(\frac{X_i'\widetilde{\beta}\left(v\right) + \widetilde{\lambda}\left(c, v\right)}{h_1}\right)K_2\left(\frac{\widehat{\widetilde{V}}_i - v}{h_2}\right)t_\kappa\left(Z_i^c\right)\\ &\quad + o_p\left(n^{-\frac{1}{2}}\right).\end{aligned}$$

Proof. Recall that $\widehat{\widetilde{\lambda}}\left(c, v\right) = \underset{a}{\operatorname{argmax}}\widehat{S}_n\left(a|c, v\right)$, where

$$\widehat{S}_n\left(a|c, v\right) = \frac{1}{nh_2}\sum_{i=1}^{n}\left(2D_i\left(c\right) - 1\right)G_1\left(\frac{X_i'\widehat{\widetilde{\beta}}\left(v\right) + a}{h_1}\right)K_2\left(\frac{\widehat{\widetilde{V}}_i - v}{h_2}\right)t_\kappa\left(Z_i^c\right).$$

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

Applying similar arguments in Horowitz (1992, Thm. 1), we can show $\widehat{\widetilde{\lambda}}(c,v)$ that maximizes the smoothed maximum score criterion function above is a consistent estimator for $\widetilde{\lambda}(c,v) = \widetilde{\beta}_0(v) - \Lambda(c,v)$, for any $c \in (\underline{c}, \bar{c}), v \in \widetilde{V}_\kappa$. Thus, we will establish the asymptotic representation for $\widehat{\widetilde{\lambda}}(c,v) - \widetilde{\lambda}(c,v)$. The first-order condition for $\widehat{S}_n(a|c,v)$ gives that

$$\frac{1}{nh_1h_2}\sum_{i=1}^{n}(2D_i(c)-1)K_1\left(\frac{X_i'\widehat{\widetilde{\beta}}(v)+\widehat{\widetilde{\lambda}}(c,v)}{h_1}\right)K_2\left(\frac{\widehat{\widetilde{V}}_i-v}{h_2}\right)t_\kappa\left(Z_i^c\right)=0.$$

The Taylor expansion of $\widehat{\widetilde{\lambda}}(c,v)$ about $\widetilde{\lambda}(c,v)$ gives that

$$\begin{aligned}
&\widehat{\widetilde{\lambda}}(c,v)-\widetilde{\lambda}(c,v)\\
&\quad=\left[-\frac{1}{nh_1^2h_2}\sum_{i=1}^{n}(2D_i(c)-1)K_1'\left(\frac{X_i'\widehat{\widetilde{\beta}}(v)+\bar{\lambda}(c,v)}{h_1}\right)K_2\left(\frac{\widehat{\widetilde{V}}_i-v}{h_2}\right)t_\kappa\left(Z_i^c\right)\right]^{-1}\\
&\qquad\times\frac{1}{nh_1h_2}\sum_{i=1}^{n}(2D_i(c)-1)K_1\left(\frac{X_i'\widehat{\widetilde{\beta}}(v)+\widetilde{\lambda}(c,v)}{h_1}\right)K_2\left(\frac{\widehat{\widetilde{V}}_i-v}{h_2}\right)t_\kappa\left(Z_i^c\right),
\end{aligned}$$

where $\bar{\lambda}(c,v)$ lies between $\widehat{\widetilde{\lambda}}(c,v)$ and $\widetilde{\lambda}(c,v)$. After applying the same arguments in the proof of proposition 3.2 in Zhang (2016) and identical steps to those in Lemma A.5, we have

$$\begin{aligned}
&\widehat{\widetilde{\lambda}}(c,v)-\widetilde{\lambda}(c,v)\\
&\quad=\frac{1}{nh_1h_2\alpha_{OO}(c,v)}\sum_{i=1}^{n}(2D_i(c)-1)K_1\left(\frac{X_i'\widetilde{\beta}(v)+\widetilde{\lambda}(c,v)}{h_1}\right)K_2\left(\frac{\widehat{\widetilde{V}}_i-v}{h_2}\right)t_\kappa\left(Z_i^c\right)\\
&\qquad+o_p\left(n^{-\frac{1}{2}}\right).
\end{aligned}$$

■

LEMMA A.8. *If Assumptions 2.1–2.4 and 4.1–4.6 hold, then as* $n \to \infty$,

$$\widehat{\widetilde{\lambda}}\left(c,\widehat{\widetilde{V}}_i\right)-\widehat{\widetilde{\lambda}}\left(c,\widetilde{V}_i\right)=\frac{1}{n}\sum_{j=1}^{n}\delta_\lambda\left(c,\widetilde{V}_i\right)\frac{\mathbf{1}\left\{\mathfrak{X}_j\le\mathfrak{X}_i\right\}-\widetilde{V}_i}{f\left(Z_i^c,Z_i^d\right)}\frac{1}{h_0}K_0\left(\frac{Z_j^c-Z_i^c}{h_0}\right)\mathbf{1}\left\{Z_j^d=Z_i^d\right\}+o_p\left(n^{-\frac{1}{2}}\right),$$

where $\delta_\lambda(c,v)$ *is the first-order derivative of* $\lambda(c,v)$ *with respect to* v.

Proof. Similar arguments to those in Lemma A.6 show that

$$\widehat{\widetilde{\lambda}}\left(c,\widehat{\widetilde{V}}_i\right)-\widehat{\widetilde{\lambda}}\left(c,\widetilde{V}_i\right)=\delta_\lambda\left(c,\widetilde{V}_i\right)\left(\widehat{\widetilde{V}}_i-\widetilde{V}_i\right)+o_p\left(n^{-\frac{1}{2}}\right).$$

Combining the result in Lemma A.1, the proof is completed. ■

Proof of Theorem 4.1. Recall that $\widetilde{\beta}(v) = \left(1, \widetilde{\beta}_2'(v)\right)'$, $X = (X_1, \widetilde{X}')'$, and $\widetilde{\lambda}(c,v) = \widetilde{\beta}_0(v) - \Lambda(c,v)$. To save space, we suppress the trimming function $t_\kappa(\mathfrak{X}_i)t_\kappa\left(Z_i^c\right)$. Thus,

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

$\widehat{\vartheta}_\kappa$ can be written as

$$\widehat{\vartheta}_\kappa = \frac{1}{n}\sum_{i=1}^{n}\int_{\underline{c}}^{\overline{c}} \frac{1}{h_3} K_3\left(\frac{X_i'\widehat{\widetilde{\beta}}\left(\widehat{\widetilde{V}}_i\right) + \widehat{\widetilde{\lambda}}\left(c, \widehat{\widetilde{V}}_i\right)}{h_3}\right) dc \times \widehat{\widetilde{\beta}}\left(\widehat{\widetilde{V}}_i\right).$$

Furthermore, $\widehat{\vartheta}_\kappa - \vartheta_\kappa$ can be decomposed into

$$\widehat{\vartheta}_\kappa - \vartheta_\kappa = L_1 + L_2 + L_3 + L_4,$$

where

$$L_1 = \frac{1}{n}\sum_{i=1}^{n}\int_{\underline{c}}^{\overline{c}} \frac{1}{h_3} K_3\left(\frac{X_i'\widetilde{\beta}\left(\widetilde{V}_i\right) + \widetilde{\lambda}\left(c, \widetilde{V}_i\right)}{h_3}\right) dc \times \widetilde{\beta}\left(\widetilde{V}_i\right) - \vartheta_\kappa,$$

$$L_2 = \frac{1}{n}\sum_{i=1}^{n}\int_{\underline{c}}^{\overline{c}} \frac{1}{h_3} K_3\left(\frac{X_i'\widetilde{\beta}\left(\widetilde{V}_i\right) + \widetilde{\lambda}\left(c, \widetilde{V}_i\right)}{h_3}\right) dc \times \left[\widehat{\widetilde{\beta}}\left(\widehat{\widetilde{V}}_i\right) - \widetilde{\beta}\left(\widetilde{V}_i\right)\right],$$

$$L_3 = \frac{1}{n}\sum_{i=1}^{n}\int_{\underline{c}}^{\overline{c}} \frac{1}{h_3}\left[K_3\left(\frac{X_i'\widehat{\widetilde{\beta}}\left(\widehat{\widetilde{V}}_i\right) + \widetilde{\lambda}\left(c, \widetilde{V}_i\right)}{h_3}\right) - K_3\left(\frac{X_i'\widetilde{\beta}\left(\widetilde{V}_i\right) + \widetilde{\lambda}\left(c, \widetilde{V}_i\right)}{h_3}\right)\right] dc \times \widehat{\widetilde{\beta}}\left(\widehat{\widetilde{V}}_i\right),$$

$$L_4 = \frac{1}{n}\sum_{i=1}^{n}\int_{\underline{c}}^{\overline{c}} \frac{1}{h_3}\left[K_3\left(\frac{X_i'\widehat{\widetilde{\beta}}\left(\widehat{\widetilde{V}}_i\right) + \widehat{\widetilde{\lambda}}\left(c, \widehat{\widetilde{V}}_i\right)}{h_3}\right) - K_3\left(\frac{X_i'\widehat{\widetilde{\beta}}\left(\widetilde{V}_i\right) + \widetilde{\lambda}\left(c, \widetilde{V}_i\right)}{h_3}\right)\right] dc \times \widehat{\widetilde{\beta}}\left(\widehat{\widetilde{V}}_i\right).$$

First, we consider L_1. Note that

$$\begin{aligned}
L_1 &= \frac{1}{n}\sum_{i=1}^{n}\int_{\underline{c}}^{\overline{c}} \frac{1}{h_3} K_3\left(\frac{\widetilde{\beta}_0(\widetilde{V}_i) + X_i'\widetilde{\beta}\left(\widetilde{V}_i\right) - \Lambda\left(c, \widetilde{V}_i\right)}{h_3}\right) dc \times \widetilde{\beta}\left(\widetilde{V}_i\right) - \vartheta_\kappa \\
&= \frac{1}{n}\sum_{i=1}^{n}\int_{\underline{c}}^{\overline{c}} \frac{1}{h_3} K_3\left(t_3\right) d\Lambda^{-1}\left(\widetilde{\beta}_0\left(\widetilde{V}_i\right) + X_i'\widetilde{\beta}\left(\widetilde{V}_i\right) - t_3 h_3, \widetilde{V}_i\right) \times \widetilde{\beta}\left(\widetilde{V}_i\right) - \vartheta_\kappa \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\widetilde{\beta}\left(\widetilde{V}_i\right)}{\Lambda'\left(\Lambda^{-1}\left(\widetilde{\beta}_0\left(\widetilde{V}_i\right) + X_i'\widetilde{\beta}\left(\widetilde{V}_i\right), \widetilde{V}_i\right), \widetilde{V}_i\right)} - \vartheta_\kappa\right] + O_p\left(h_3^{\kappa_3}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} f_{1n}(\xi_i) + o_p(n^{-\frac{1}{2}}).
\end{aligned}$$

The first equation holds by the definition that $\widetilde{\lambda}(c, v) = \widetilde{\beta}_0(v) - \Lambda(c, v)$. Recall that $\Lambda'(\cdot,\cdot)$ denotes the first-order derivative, and $\Lambda^{-1}(\cdot,\cdot)$ means the inverse of $\Lambda(\cdot,\cdot)$ with respect to the first variable. The second equation follows from the change of variables $t_3 h_3 = \widetilde{\beta}_0\left(\widetilde{V}_i\right) + X_i'\widetilde{\beta}\left(\widetilde{V}_i\right) - \Lambda(c, \widetilde{V}_i)$. The third equation follows from Assumption 4.4.

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

Next, consider L_2. Let $\Pi(\widetilde{\beta}_0(\widetilde{V}_i)+X_i'\widetilde{\beta}(\widetilde{V}_i),\widetilde{V}_i)=1/\Lambda'(\Lambda^{-1}(\widetilde{\beta}_0(\widetilde{V}_i)+X_i'\widetilde{\beta}(\widetilde{V}_i),\widetilde{V}_i),\widetilde{V}_i)$. Similarly, after a change of variable and the Taylor expansion, we have

$$
\begin{aligned}
L_2 &= \frac{1}{n}\sum_{i=1}^{n}\Pi\left(\widetilde{\beta}_0\left(\widetilde{V}_i\right)+X_i'\widetilde{\beta}\left(\widetilde{V}_i\right),\widetilde{V}_i\right)\left[\widehat{\widetilde{\beta}}\left(\widehat{\widetilde{V}}_i\right)-\widetilde{\beta}\left(\widetilde{V}_i\right)\right]+o_p(n^{-\frac{1}{2}}) \\
&= \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} 0 \\ \Pi\left(\widetilde{\beta}_0\left(\widetilde{V}_i\right)+X_i'\widetilde{\beta}\left(\widetilde{V}_i\right),\widetilde{V}_i\right)\left[\widehat{\widetilde{\beta}}_2\left(\widehat{\widetilde{V}}_i\right)-\widetilde{\beta}_2\left(\widetilde{V}_i\right)\right] \end{pmatrix}+o_p(n^{-\frac{1}{2}}).
\end{aligned}
$$

The function $\Pi(\cdot,\cdot)$ can be viewed as a weighting function, varying in each $\left(X_i,\widetilde{V}_i\right)$. Plugging the representation of $\widehat{\widetilde{\beta}}_2\left(\widehat{\widetilde{V}}_i\right)-\widetilde{\beta}_2\left(\widetilde{V}_i\right)$ into L_2, we have

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\Pi\left(\widetilde{\beta}_0\left(\widetilde{V}_i\right)+X_i'\widetilde{\beta}\left(\widetilde{V}_i\right),\widetilde{V}_i\right)\left[\widehat{\widetilde{\beta}}_2\left(\widehat{\widetilde{V}}_i\right)-\widetilde{\beta}_2\left(\widetilde{V}_i\right)\right] \\
=&\frac{1}{n(n-1)}\sum_{i\neq j}\mathcal{F}_{2n}(\xi_i.\xi_j)+\frac{1}{n(n-1)(n-2)}\sum_{i\neq j\neq k}\mathcal{F}_{3n}\left(\xi_i,\xi_j,\xi_k\right) \\
+&\frac{1}{n(n-1)}\sum_{i\neq j}\mathcal{F}_{4n}(\xi_i.\xi_j)+o_p(n^{-\frac{1}{2}}),
\end{aligned}
$$

where

$$
\begin{aligned}
\mathcal{F}_{2n}(\xi_i.\xi_j) &= \int_{\underline{c}}^{\overline{c}}(2D_j(c)-1)\left(\widetilde{X}_j+\frac{\alpha_{XO}\left(c,\widetilde{V}_i\right)}{\alpha_{OO}\left(c,\widetilde{V}_i\right)}\right)\frac{1}{h_1}K_1\left(\frac{X_j'\widetilde{\beta}\left(\widetilde{V}_i\right)+\widetilde{\lambda}\left(c,\widetilde{V}_i\right)}{h_1}\right)\varpi(c)dc \\
&\quad\times\Pi\left(\widetilde{\beta}_0\left(\widetilde{V}_i\right)+X_i'\widetilde{\beta}\left(\widetilde{V}_i\right),\widetilde{V}_i\right)\frac{1}{h_2}K_2\left(\frac{\widetilde{V}_j-\widetilde{V}_i}{h_2}\right)\left[-\sigma\left(\widetilde{V}_i\right)\right]^{-1}, \\
\mathcal{F}_{3n}\left(\xi_i,\xi_j,\xi_k\right) &= \int_{\underline{c}}^{\overline{c}}(2D_j(c)-1)\left(\widetilde{X}_j+\frac{\alpha_{XO}\left(c,\widetilde{V}_i\right)}{\alpha_{OO}\left(c,\widetilde{V}_i\right)}\right)\frac{1}{h_1}K_1\left(\frac{X_j'\widetilde{\beta}\left(\widetilde{V}_i\right)+\widetilde{\lambda}\left(c,\widetilde{V}_i\right)}{h_1}\right)\varpi(c)dc \\
&\quad\times\Pi\left(\widetilde{\beta}_0\left(\widetilde{V}_i\right)+X_i'\widetilde{\beta}\left(\widetilde{V}_i\right),\widetilde{V}_i\right)\frac{1}{h_2^2}K_2'\left(\frac{\widetilde{V}_j-\widetilde{V}_i}{h_2}\right)\left[-\sigma\left(\widetilde{V}_i\right)\right]^{-1} \\
&\quad\times\frac{\mathbf{1}\left\{\mathfrak{X}_k\leq\mathfrak{X}_j\right\}-\widetilde{V}\left(\mathfrak{X}_j,Z_j\right)}{f\left(Z_j^c,Z_j^d\right)}\frac{1}{h_0}K_0\left(\frac{Z_k^c-Z_j^c}{h_0}\right)\mathbf{1}\left\{Z_k^d=Z_j^d\right\},
\end{aligned}
$$

$$
\mathcal{F}_{4n}(\xi_i,\xi_j)=\Pi\left(\widetilde{\beta}_0\left(\widetilde{V}_i\right)+X_i'\widetilde{\beta}\left(\widetilde{V}_i\right),\widetilde{V}_i\right)\delta\left(\widetilde{V}_i\right)\times\frac{\mathbf{1}\left\{\mathfrak{X}_j\leq\mathfrak{X}_i\right\}-\widetilde{V}_i}{f\left(Z_i^c,Z_i^d\right)}\frac{1}{h_0}K_0\left(\frac{Z_j^c-Z_i^c}{h_0}\right)\mathbf{1}\left\{Z_j^d=Z_i^d\right\}.
$$

For $\mathcal{F}_{2n}(\xi_i.\xi_j)$, according to Powell, Stock, and Stoker (1989), the condition sufficient for U-statistic projection is $\mathbb{E}||\mathcal{F}_{2n}\left(\xi_i,\xi_j\right)||^2=o(n)$. Notice that the only term in $\mathcal{F}_{2n}$ that becomes unbounded when $n\to\infty$ is $\frac{1}{h_1}K_1\left(\frac{X_j'\widetilde{\beta}(\widetilde{V}_i)+\widetilde{\lambda}(c,\widetilde{V}_i)}{h_1}\right)\frac{1}{h_2}K_2\left(\frac{\widetilde{V}_j-\widetilde{V}_i}{h_2}\right)$, which has the order $O\left(h_1^{-1}h_2^{-1}\right)$. Thus, the result following from the U-statistic projection is

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

given by

$$\frac{1}{n(n-1)}\sum_{i\neq j}\mathcal{F}_{2n}\left(\xi_i,\xi_j\right)=\frac{1}{n}\sum_{j=1}^{n}\mathbb{E}\left[\mathcal{F}_{2n}\left(\xi_i,\xi_j\right)|\xi_j\right]+o_p\left(n^{-\frac{1}{2}}\right).$$

In evaluating $\mathbb{E}\left[\mathcal{F}_{2n}(\xi_i.\xi_j)|\xi_j\right]$, averaging over ξ_i will be done in three steps: (i) average over $\widetilde{V}_i$ given X_i, (ii) average over X_i, and (iii) solve the integral of c. Let us make a change of variable $th_2=\widetilde{V}_i-\widetilde{V}_j$. After applying step (i), and the Taylor expansion of $\widetilde{V}_j$, the integral is equal to

$$\int_{\underline{c}}^{\overline{c}}\left(2D_j(c)-1\right)\left(\widetilde{X}_j+\frac{\alpha_{XO}\left(c,\widetilde{V}_j\right)}{\alpha_{OO}\left(c,\widetilde{V}_j\right)}\right)\frac{1}{h_1}K_1\left(\frac{X_j'\widetilde{\beta}\left(\widetilde{V}_j\right)+\widetilde{\lambda}\left(c,\widetilde{V}_j\right)}{h_1}\right)\varpi(c)dc$$
$$\times\Pi\left(\widetilde{\beta}_0\left(\widetilde{V}_j\right)+X_i'\widetilde{\beta}\left(\widetilde{V}_j\right),\widetilde{V}_j\right)f\left(\widetilde{V}_j|X_i\right)\left[-\sigma\left(\widetilde{V}_j\right)\right]^{-1},$$

plus a term with order $o_p(n^{-\frac{1}{2}})$, as long as $\frac{h_2^K}{h_1^K}$ is of order $o_p(n^{-\frac{1}{2}})$, where $f(\cdot|\cdot)$ denotes the density function of $\widetilde{V}$, given X. After applying step (ii), the integral is equal to

$$\int_{\underline{c}}^{\overline{c}}\left(2D_j(c)-1\right)\left(\widetilde{X}_j+\frac{\alpha_{XO}\left(c,\widetilde{V}_j\right)}{\alpha_{OO}\left(c,\widetilde{V}_j\right)}\right)\frac{1}{h_1}K_1\left(\frac{X_j'\widetilde{\beta}\left(\widetilde{V}_j\right)+\widetilde{\lambda}\left(c,\widetilde{V}_j\right)}{h_1}\right)\varpi(c)dc$$
$$\times\mathbb{E}_X\left[\Pi\left(\widetilde{\beta}_0\left(\widetilde{V}_j\right)+X'\widetilde{\beta}\left(\widetilde{V}_j\right),\widetilde{V}_j\right)f\left(\widetilde{V}_j|X\right)\right]\left[-\sigma\left(\widetilde{V}_j\right)\right]^{-1},$$

where $\mathbb{E}_X$ means the expectation with respect to X. Let us make a change of variable $th_1=X_j'\widetilde{\beta}\left(\widetilde{V}_j\right)+\widetilde{\beta}_0\left(\widetilde{V}_j\right)-\Lambda\left(c,\widetilde{V}_j\right)$. After applying step (iii), and Newey and McFadden (1994, Thm. 8.1), the integral is equal to

$$\left(2\times\mathbf{1}\left\{\widetilde{\varepsilon}\left(\widetilde{V}_j\right)<0\right\}-1\right)\Gamma\left(\widetilde{X}_j,X_j'\widetilde{\beta}\left(\widetilde{V}_j\right)+\widetilde{\beta}_0\left(\widetilde{V}_j\right),\widetilde{V}_j\right)$$
$$\times\mathbb{E}_X\left[\Pi\left(\widetilde{\beta}_0\left(\widetilde{V}_j\right)+X'\widetilde{\beta}\left(\widetilde{V}_j\right),\widetilde{V}_j\right)f\left(\widetilde{V}_j|X\right)\right]\left[-\sigma\left(\widetilde{V}_j\right)\right]^{-1},$$

plus a term with order $o_p(n^{-\frac{1}{2}})$, where

$$\Gamma\left(\widetilde{X}_j,X_j'\widetilde{\beta}\left(\widetilde{V}_j\right)+\widetilde{\beta}_0\left(\widetilde{V}_j\right),\widetilde{V}_j\right)$$
$$=\left(\widetilde{X}_j+\frac{\alpha_{XO}\left(\Lambda^{-1}\left(X_j'\widetilde{\beta}\left(\widetilde{V}_j\right)+\widetilde{\beta}_0\left(\widetilde{V}_j\right),\widetilde{V}_j\right),\widetilde{V}_j\right)}{\alpha_{OO}\left(\Lambda^{-1}\left(X_j'\widetilde{\beta}\left(\widetilde{V}_j\right)+\widetilde{\beta}_0\left(\widetilde{V}_j\right),\widetilde{V}_j\right),\widetilde{V}_j\right)}\right)\frac{\varpi\left(\Lambda^{-1}\left(X_j'\widetilde{\beta}\left(\widetilde{V}_j\right)+\widetilde{\beta}_0\left(\widetilde{V}_j\right),\widetilde{V}_j\right)\right)}{\Lambda'\left(\Lambda^{-1}\left(X_j'\widetilde{\beta}\left(\widetilde{V}_j\right)+\widetilde{\beta}_0\left(\widetilde{V}_j\right),\widetilde{V}_j\right)\right)}.$$

Thus, we derive that

$$\frac{1}{n(n-1)}\sum_{i\neq j}\mathcal{F}_{2n}(\xi_i.\xi_j)=\frac{1}{n}\sum_{i=1}^{n}f_{2n}(\xi_i)+o_p(n^{-\frac{1}{2}}),$$

where

$$f_{2n}(\xi_i)=\left(2\times\mathbf{1}\left\{\widetilde{\varepsilon}\left(\widetilde{V}_i\right)<0\right\}-1\right)\Gamma\left(\widetilde{X}_i,X_i'\widetilde{\beta}\left(\widetilde{V}_i\right)+\widetilde{\beta}_0\left(\widetilde{V}_i\right),\widetilde{V}_i\right)$$
$$\times\mathbb{E}_X\left[\Pi\left(\widetilde{\beta}_0\left(\widetilde{V}_i\right)+X'\widetilde{\beta}\left(\widetilde{V}_i\right),\widetilde{V}_i\right)f\left(\widetilde{V}_i|X\right)\right]\left[-\sigma\left(\widetilde{V}_i\right)\right]^{-1}.$$

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

Similarly, the remaining terms can be derived as

$$\frac{1}{n(n-1)(n-2)}\sum_{i\neq j\neq k}\mathcal{F}_{3n}(\xi_i,\xi_j,\xi_k)=\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\left[\mathcal{F}_{3n}(\xi_i.\xi_j,\xi_k)|\xi_k\right]+o_p(n^{-\frac{1}{2}}),$$

and

$$\frac{1}{n(n-1)}\sum_{i\neq j}\mathcal{F}_{4n}(\xi_i.\xi_j)=\frac{1}{n}\sum_{j=1}^{n}\mathbb{E}\left[\mathcal{F}_{4n}(\xi_i.\xi_j)|\xi_j\right]+o_p(n^{-\frac{1}{2}}).$$

Recall that

$$\begin{aligned}\mathcal{F}_{3n}(\xi_i,\xi_j,\xi_k)=&\int_{\underline{c}}^{\bar{c}}(2D_j(c)-1)\left(\widetilde{X}_j+\frac{\alpha_{XO}(c,\widetilde{V}_i)}{\alpha_{OO}(c,\widetilde{V}_i)}\right)\frac{1}{h_1}K_1\left(\frac{X_j'\widetilde{\beta}(\widetilde{V}_i)+\widetilde{\lambda}(c,\widetilde{V}_i)}{h_1}\right)\varpi(c)dc\\&\times\Pi\left(\widetilde{\beta}_0(\widetilde{V}_i)+X_i'\widetilde{\beta}(\widetilde{V}_i),\widetilde{V}_i\right)\frac{1}{h_2^2}K_2'\left(\frac{\widetilde{V}_j-\widetilde{V}_i}{h_2}\right)\left[-\sigma(\widetilde{V}_i)\right]^{-1}\\&\times\frac{\mathbf{1}\{\mathfrak{X}_k\leq\mathfrak{X}_j\}-\widetilde{V}(\mathfrak{X}_j,Z_j)}{f\left(Z_j^c,Z_j^d\right)}\frac{1}{h_0}K_0\left(\frac{Z_k^c-Z_j^c}{h_0}\right)\mathbf{1}\left\{Z_k^d=Z_j^d\right\}.\end{aligned}$$

In evaluating $\mathbb{E}\left[\mathcal{F}_{3n}(\xi_i.\xi_j,\xi_k)|\xi_k\right]$, we first fix ξ_j and average over ξ_i. Averaging over ξ_i comprises two steps: (i) average over $\widetilde{V}_i$ given X_i and (ii) average over X_i. In applying step (i), we need a change of variables $th_2=\widetilde{V}_j-\widetilde{V}_i$. Then, applying the Taylor expansion of $\widetilde{V}_j$, the integral is equal to

$$\begin{aligned}&\int_{\underline{c}}^{\bar{c}}(2D_j(c)-1)\left[\Upsilon_1'(X_i,\widetilde{V}_j,\widetilde{X}_j,c)\frac{1}{h_1}K_1\left(\frac{X_j'\widetilde{\beta}(\widetilde{V}_j)+\widetilde{\lambda}(c,\widetilde{V}_j)}{h_1}\right)\right.\\&\left.+\Upsilon_1(X_i,\widetilde{V}_j,\widetilde{X}_j,c)\frac{1}{h_1^2}K_1'\left(\frac{X_j'\widetilde{\beta}(\widetilde{V}_j)+\widetilde{\lambda}(c,\widetilde{V}_j)}{h_1}\right)\left(\widetilde{X}_j'\delta(\widetilde{V}_j)+\delta_\lambda(c,\widetilde{V}_j)\right)\right]\varpi(c)dc\\&\times\frac{\mathbf{1}\{\mathfrak{X}_k\leq\mathfrak{X}_j\}-\widetilde{V}(\mathfrak{X}_j,Z_j)}{f\left(Z_j^c,Z_j^d\right)}\frac{1}{h_0}K_0\left(\frac{Z_k^c-Z_j^c}{h_0}\right)\mathbf{1}\left\{Z_k^d=Z_j^d\right\},\end{aligned}$$

plus a term with order $o_p(n^{-\frac{1}{2}})$, where $\Upsilon_1'(X_i,\widetilde{V}_i,\widetilde{X}_j,c)$ denotes the first-order derivative with respect to $\widetilde{V}_i$, $\delta_\lambda(c,v)$ denotes the first-order derivative with respect to v,

$$\Upsilon_1(X_i,\widetilde{V}_i,\widetilde{X}_j,c)=\Pi\left(\widetilde{\beta}_0(\widetilde{V}_i)+X_i'\widetilde{\beta}(\widetilde{V}_i),\widetilde{V}_i\right)\left[-\sigma(\widetilde{V}_i)\right]^{-1}f(\widetilde{V}_i|X_i)\left(\widetilde{X}_j+\frac{\alpha_{XO}(c,\widetilde{V}_i)}{\alpha_{OO}(c,\widetilde{V}_i)}\right),$$

and by making use of $\int K_2'(s)ds=0$, $\int K_2'(s)sds=-1$. Then, we average over X_i and get

$$\begin{aligned}&\int_{\underline{c}}^{\bar{c}}(2D_j(c)-1)\left[\mathbb{E}_{X_i}\left[\Upsilon_1'(X_i,\widetilde{V}_j,\widetilde{X}_j,c)\right]\frac{1}{h_1}K_1\left(\frac{X_j'\widetilde{\beta}(\widetilde{V}_j)+\widetilde{\lambda}(c,\widetilde{V}_j)}{h_1}\right)\right.\\&\left.+\mathbb{E}_{X_i}\left[\Upsilon_1(X_i,\widetilde{V}_j,\widetilde{X}_j,c)\right]\frac{1}{h_1^2}K_1'\left(\frac{X_j'\widetilde{\beta}(\widetilde{V}_j)+\widetilde{\lambda}(c,\widetilde{V}_j)}{h_1}\right)\left(\widetilde{X}_j'\delta(\widetilde{V}_j)+\delta_\lambda(c,\widetilde{V}_j)\right)\right]\varpi(c)dc\end{aligned}$$

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

$$\times \frac{\mathbf{1}\{\mathfrak{X}_k \le \mathfrak{X}_j\} - \widetilde{V}(\mathfrak{X}_j, Z_j)}{f\left(Z_j^c, Z_j^d\right)} \frac{1}{h_0} K_0\left(\frac{Z_k^c - Z_j^c}{h_0}\right) \mathbf{1}\left\{Z_k^d = Z_j^d\right\}.$$

Averaging over ξ_j will be done in four steps: (i) average over $D_j(c)$ given $\widetilde{V}_j$ and Z_j; (ii) solve the integral of c; (iii) average over Z_j given X_j; and (iv) average over X_j. After applying step (i), the integral above is equal to

$$\int_{\underline{c}}^{\overline{c}} \mu\left(\widetilde{V}_j, Z_j, c\right) \left[\mathbb{E}_{X_i}\left[\Upsilon_1'\left(X_i, \widetilde{V}_j, \widetilde{X}_j, c\right)\right] \frac{1}{h_1} K_1\left(\frac{X_j'\widetilde{\beta}\left(\widetilde{V}_j\right) + \widetilde{\lambda}\left(c, \widetilde{V}_j\right)}{h_1}\right)\right.$$
$$\left. + \mathbb{E}_{X_i}\left[\Upsilon_1\left(X_i, \widetilde{V}_j, \widetilde{X}_j, c\right)\right] \frac{1}{h_1^2} K_1'\left(\frac{X_j'\widetilde{\beta}\left(\widetilde{V}_j\right) + \widetilde{\lambda}\left(c, \widetilde{V}_j\right)}{h_1}\right)\left(\widetilde{X}_j'\delta\left(\widetilde{V}_j\right) + \delta_\lambda\left(c, \widetilde{V}_j\right)\right)\right] \varpi(c) dc$$
$$\times \frac{\mathbf{1}\{\mathfrak{X}_k \le \mathfrak{X}_j\} - \widetilde{V}(\mathfrak{X}_j, Z_j)}{f\left(Z_j^c, Z_j^d\right)} \frac{1}{h_0} K_0\left(\frac{Z_k^c - Z_j^c}{h_0}\right) \mathbf{1}\left\{Z_k^d = Z_j^d\right\},$$

where $\mu\left(\widetilde{V}, Z, c\right) = 2 \times Pr\left(\widetilde{\varepsilon}\left(\widetilde{V}\right) < X'\widetilde{\beta}\left(\widetilde{V}\right) + \widetilde{\lambda}\left(c, \widetilde{V}\right) | \widetilde{V}, Z\right) - 1$. Let us make a change of variable $th_1 = X_j'\widetilde{\beta}\left(\widetilde{V}_j\right) + \widetilde{\beta}_0\left(\widetilde{V}_j\right) - \Lambda\left(c, \widetilde{V}_j\right)$. By applying step (ii), and taking the Taylor expansion, the integral is equal to

$$\Upsilon_2\left(\widetilde{V}_j, Z_j, \Lambda^{-1}\left(X_j'\widetilde{\beta}\left(\widetilde{V}_j\right) + \widetilde{\beta}_0\left(\widetilde{V}_j\right), \widetilde{V}_j\right)\right) \Pi\left(\widetilde{\beta}_0\left(\widetilde{V}_j\right) + X_j'\widetilde{\beta}\left(\widetilde{V}_j\right), \widetilde{V}_j\right)$$
$$\times \frac{\mathbf{1}\{\mathfrak{X}_k \le \mathfrak{X}_j\} - \widetilde{V}(\mathfrak{X}_j, Z_j)}{f\left(Z_j^c, Z_j^d\right)} \frac{1}{h_0} K_0\left(\frac{Z_k^c - Z_j^c}{h_0}\right) \mathbf{1}\left\{Z_k^d = Z_j^d\right\},$$

plus a term with order $o_p(n^{-\frac{1}{2}})$, where $\mu\left(\widetilde{V}, Z, \Lambda^{-1}\left(X'\widetilde{\beta}\left(\widetilde{V}\right) + \widetilde{\beta}_0\left(\widetilde{V}\right), \widetilde{V}\right)\right) = 0$, by the definition of $\mu\left(\widetilde{V}, Z, c\right)$ and Assumption 2.3. Furthermore,

$$\Upsilon_2\left(\widetilde{V}, Z, c\right)$$
$$= \frac{\partial}{\partial c}\left[\mu\left(\widetilde{V}, Z, c\right) \mathbb{E}_{X_i}\left[\Upsilon_1\left(X_i, \widetilde{V}, \widetilde{X}, c\right)\right] \left(\widetilde{X}'\delta\left(\widetilde{V}\right) + \delta_\lambda\left(c, \widetilde{V}\right)\right) \Pi\left(\Lambda\left(c, \widetilde{V}\right), \widetilde{V}\right) \varpi(c)\right].$$

Note that as long as given X and Z, $\widetilde{V}$ is determined. Let us make a change of variable $th_0 = Z_j^c - Z_k^c$. After applying step (iii), and the Taylor expansion of Z_k^c, the integral is equal to

$$\Upsilon_2\left(\widetilde{V}_{jk}, Z_k, \Lambda^{-1}\left(X_j'\widetilde{\beta}\left(\widetilde{V}_{jk}\right) + \widetilde{\beta}_0\left(\widetilde{V}_{jk}\right), \widetilde{V}_{jk}\right)\right) \Pi\left(\widetilde{\beta}_0\left(\widetilde{V}_{jk}\right) + X_j'\widetilde{\beta}\left(\widetilde{V}_{jk}\right), \widetilde{V}_{jk}\right)$$
$$\times \left[\mathbf{1}\{\mathfrak{X}_k \le \mathfrak{X}_j\} - \widetilde{V}(\mathfrak{X}_j, Z_k)\right] \frac{f\left(Z_k^c, Z_k^d | X_j\right)}{f\left(Z_k^c, Z_k^d\right)},$$

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

plus a term with order $o_p(n^{-\frac{1}{2}})$, where $f(\cdot,\cdot|\cdot)$ denotes the density of $\left(Z_k^c, Z_k^d\right)$ conditional on X, and $\widetilde{V}_{jk}$ denotes $\widetilde{V}(\mathfrak{X}_j, Z_k)$. Finally, after applying step (iv), the integral is equal to

$$\mathbb{E}_{X_j}\Bigg[\Upsilon_2\left(\widetilde{V}_{jk}, Z_k, \Lambda^{-1}\left(X_j'\widetilde{\beta}\left(\widetilde{V}_{jk}\right)+\widetilde{\beta}_0\left(\widetilde{V}_{jk}\right), \widetilde{V}_{jk}\right)\right)\Pi\left(\widetilde{\beta}_0\left(\widetilde{V}_{jk}\right)+X_j'\widetilde{\beta}\left(\widetilde{V}_{jk}\right), \widetilde{V}_{jk}\right)$$
$$\times\left[\mathbf{1}\left\{\mathfrak{X}_k \le \mathfrak{X}_j\right\}-\widetilde{V}\left(\mathfrak{X}_j, Z_k\right)\right]\frac{f\left(Z_k^c, Z_k^d|X_j\right)}{f\left(Z_k^c, Z_k^d\right)}\Big|\xi_k\Bigg].$$

Thus, we get that

$$\frac{1}{n(n-1)(n-2)}\sum_{i\neq j\neq k}\mathcal{F}_{3n}\left(\xi_i,\xi_j,\xi_k\right)=\frac{1}{n}\sum_{i=1}^{n}f_{3n}(\xi_i)+o_p(n^{-\frac{1}{2}}),$$

where

$$f_{3n}(\xi_i)=\mathbb{E}_{X_j}\Bigg[\Upsilon_2\left(\widetilde{V}_{ji}, Z_i, \Lambda^{-1}\left(X_j'\widetilde{\beta}\left(\widetilde{V}_{ji}\right)+\widetilde{\beta}_0\left(\widetilde{V}_{ji}\right), \widetilde{V}_{ji}\right)\right)\Pi\left(\widetilde{\beta}_0\left(\widetilde{V}_{ji}\right)+X_j'\widetilde{\beta}\left(\widetilde{V}_{ji}\right), \widetilde{V}_{ji}\right)$$
$$\times\left[\mathbf{1}\left\{\mathfrak{X}_i \le \mathfrak{X}_j\right\}-\widetilde{V}\left(\mathfrak{X}_j, Z_i\right)\right]\frac{f\left(Z_i^c, Z_i^d|X_j\right)}{f\left(Z_i^c, Z_i^d\right)}\Big|\xi_i\Bigg],$$

and $\mathbb{E}_{X_j}$ denotes the expectation with respect to X_j, and $\widetilde{V}_{ji}=\widetilde{V}\left(\mathfrak{X}_j, Z_i\right)$. Recall that

$$\mathcal{F}_{4n}(\xi_i,\xi_j)=\Pi\left(\widetilde{\beta}_0\left(\widetilde{V}_i\right)+X_i'\widetilde{\beta}\left(\widetilde{V}_i\right), \widetilde{V}_i\right)\delta\left(\widetilde{V}_i\right)\times\frac{\mathbf{1}\left\{\mathfrak{X}_j \le \mathfrak{X}_i\right\}-\widetilde{V}_i}{f\left(Z_i^c, Z_i^d\right)}\frac{1}{h_0}K_0\left(\frac{Z_j^c-Z_i^c}{h_0}\right)\mathbf{1}\left\{Z_j^d=Z_i^d\right\}.$$

In evaluating $\mathbb{E}\left[\mathcal{F}_{4n}(\xi_i.\xi_j)|\xi_j\right]$, we first average over Z_i given X_i, and then average over X_i. First, making a change of variable $th_0=Z_i^c-Z_j^c$, and taking the Taylor expansion of Z_j^c, the integral is equal to

$$\Pi\left(\widetilde{\beta}_0\left(\widetilde{V}_{ij}\right)+X_i'\widetilde{\beta}\left(\widetilde{V}_{ij}\right), \widetilde{V}_{ij}\right)\delta\left(\widetilde{V}_{ij}\right)\left[\mathbf{1}\left\{\mathfrak{X}_j \le \mathfrak{X}_i\right\}-\widetilde{V}_{ij}\right]\frac{f(Z_j^c, Z_j^d|X_i)}{f\left(Z_j^c, Z_j^d\right)},$$

plus a term with order $o_p(n^{-\frac{1}{2}})$. Then, after averaging over X_i, we have

$$\frac{1}{n(n-1)}\sum_{i\neq j}\mathcal{F}_{4n}(\xi_i.\xi_j)=\frac{1}{n}\sum_{i=1}^{n}f_{4n}(\xi_i)+o_p(n^{-\frac{1}{2}}),$$

where

$$f_{4n}(\xi_i)=\mathbb{E}_{X_j}\left[\Pi\left(\widetilde{\beta}_0\left(\widetilde{V}_{ji}\right)+X_j'\widetilde{\beta}\left(\widetilde{V}_{ji}\right), \widetilde{V}_{ji}\right)\delta\left(\widetilde{V}_{ji}\right)\left[\mathbf{1}\left\{\mathfrak{X}_i \le \mathfrak{X}_j\right\}-\widetilde{V}_{ji}\right]\frac{f(Z_i^c, Z_i^d|X_j)}{f\left(Z_i^c, Z_i^d\right)}\Big|\xi_i\right].$$

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

Thus, we can derive that

$$L_2 = \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} 0 \\ \Pi\left(\widetilde{\beta}_0\left(\widetilde{V}_i\right)+X_i'\widetilde{\beta}\left(\widetilde{V}_i\right),\widetilde{V}_i\right)\left[\widehat{\widetilde{\beta}}_2\left(\widehat{\widetilde{V}}_i\right)-\widetilde{\beta}_2\left(\widetilde{V}_i\right)\right] \end{pmatrix}+o_p(n^{-\frac{1}{2}})$$
$$= \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} 0 \\ f_{2n}(\xi_i)+f_{3n}(\xi_i)+f_{4n}(\xi_i) \end{pmatrix}+o_p(n^{-\frac{1}{2}})$$
$$= \frac{1}{n}\sum_{i=1}^{n} f_{5n}(\xi_i)+o_p(n^{-\frac{1}{2}}).$$

For L_3, by taking the Taylor expansion of $\widetilde{\beta}_2\left(\widetilde{V}_i\right)$, we have

$$L_3 = \frac{1}{n}\sum_{i=1}^{n}\int_{\underline{c}}^{\overline{c}}\frac{1}{h_3^2}K_3'\left(\frac{X_i'\overline{\beta}\left(\overline{V}_i\right)+\widetilde{\lambda}\left(c,\widetilde{V}_i\right)}{h_3}\right)\widetilde{X}_i'\left[\widehat{\widetilde{\beta}}_2\left(\widehat{\widetilde{V}}_i\right)-\widetilde{\beta}_2\left(\widetilde{V}_i\right)\right]dc\times\left[\widetilde{\beta}\left(\widetilde{V}_i\right)+\left(\widehat{\widetilde{\beta}}\left(\widehat{\widetilde{V}}_i\right)-\widetilde{\beta}\left(\widetilde{V}_i\right)\right)\right]$$
$$= \frac{1}{n}\sum_{i=1}^{n}\int_{\underline{c}}^{\overline{c}}\frac{1}{h_3^2}K_3'\left(\frac{X_i'\widetilde{\beta}\left(\widetilde{V}_i\right)+\widetilde{\lambda}\left(c,\widetilde{V}_i\right)}{h_3}\right)dc\times\widetilde{X}_i'\left[\widehat{\widetilde{\beta}}_2\left(\widehat{\widetilde{V}}_i\right)-\widetilde{\beta}_2\left(\widetilde{V}_i\right)\right]\widetilde{\beta}\left(\widetilde{V}_i\right)+o_p(n^{-\frac{1}{2}})$$
$$= \frac{1}{n}\sum_{i=1}^{n}\Pi'\left(\widetilde{\beta}_0\left(\widetilde{V}_i\right)+X_i'\widetilde{\beta}\left(\widetilde{V}_i\right),\widetilde{V}_i\right)\widetilde{X}_i'\left[\widehat{\widetilde{\beta}}_2\left(\widehat{\widetilde{V}}_i\right)-\widetilde{\beta}_2\left(\widetilde{V}_i\right)\right]\widetilde{\beta}\left(\widetilde{V}_i\right)+o_p(n^{-\frac{1}{2}}),$$

where $\overline{\beta}(\cdot)$ lies between $\widetilde{\beta}(\cdot)$ and $\widehat{\widetilde{\beta}}(\cdot)$, $\overline{V}_i$ lies between $\widetilde{V}_i$ and $\widehat{\widetilde{V}}_i$, and $\Pi'(\cdot,\cdot)$ denote the first-order derivative with respect to the first variable. The second equality follows from the fact that the convergence rate of $\frac{1}{h_3^2}\left[\widehat{\widetilde{\beta}}_2\left(\widehat{\widetilde{V}}_i\right)-\widetilde{\beta}_2\left(\widetilde{V}_i\right)\right]\times\left[\widehat{\widetilde{\beta}}\left(\widehat{\widetilde{V}}_i\right)-\widetilde{\beta}\left(\widetilde{V}_i\right)\right]'$ is faster than $n^{-\frac{1}{2}}$. The third equality holds by a change of variables $t_3h_3=\widetilde{\beta}_0(U_i)+X_i'\widetilde{\beta}\left(\widetilde{V}_i\right)-\Lambda\left(c,\widetilde{V}_i\right)$, and making use of $\int K_3'(s)ds=0$, $\int K_3'(s)sds=-1$. Plug in the asymptotic representation of $\widehat{\widetilde{\beta}}_2\left(\widehat{\widetilde{V}}_i\right)-\widetilde{\beta}_2\left(\widetilde{V}_i\right)$, and it gives

$$L_3 = \frac{1}{n(n-1)}\sum_{i\neq j}\mathcal{F}_{6n}(\xi_i.\xi_j)+\frac{1}{n(n-1)(n-2)}\sum_{i\neq j\neq k}\mathcal{F}_{7n}\left(\xi_i,\xi_j,\xi_k\right)$$
$$+\frac{1}{n(n-1)}\sum_{i\neq j}\mathcal{F}_{8n}(\xi_i.\xi_j)+o_p(n^{-\frac{1}{2}}),$$

where

$$\mathcal{F}_{6n}(\xi_i.\xi_j)=\widetilde{\beta}\left(\widetilde{V}_i\right)\times\Pi'\left(\widetilde{\beta}_0\left(\widetilde{V}_i\right)+X_i'\widetilde{\beta}\left(\widetilde{V}_i\right),\widetilde{V}_i\right)\frac{1}{h_2}K_2\left(\frac{\widetilde{V}_j-\widetilde{V}_i}{h_2}\right)\left[-\sigma\left(\widetilde{V}_i\right)\right]^{-1}\widetilde{X}_i'$$
$$\times\int_{\underline{c}}^{\overline{c}}\left(2D_j(c)-1\right)\left(\widetilde{X}_j+\frac{\alpha_{XO}\left(c,\widetilde{V}_i\right)}{\alpha_{OO}\left(c,\widetilde{V}_i\right)}\right)\frac{1}{h_1}K_1\left(\frac{X_j'\widetilde{\beta}\left(\widetilde{V}_i\right)+\widetilde{\lambda}\left(c,\widetilde{V}_i\right)}{h_1}\right)\varpi(c)dc,$$

$$\mathcal{F}_{7n}\left(\xi_i,\xi_j,\xi_k\right)=\widetilde{\beta}\left(\widetilde{V}_i\right)\times\Pi'\left(\widetilde{\beta}_0\left(\widetilde{V}_i\right)+X_i'\widetilde{\beta}\left(\widetilde{V}_i\right),\widetilde{V}_i\right)\frac{1}{h_2^2}K_2'\left(\frac{\widetilde{V}_j-\widetilde{V}_i}{h_2}\right)\left[-\sigma\left(\widetilde{V}_i\right)\right]^{-1}\widetilde{X}_i'$$
$$\times\int_{\underline{c}}^{\overline{c}}\left(2D_j(c)-1\right)\left(\widetilde{X}_j+\frac{\alpha_{XO}\left(c,\widetilde{V}_i\right)}{\alpha_{OO}\left(c,\widetilde{V}_i\right)}\right)\frac{1}{h_1}K_1\left(\frac{X_j'\widetilde{\beta}\left(\widetilde{V}_i\right)+\widetilde{\lambda}\left(c,\widetilde{V}_i\right)}{h_1}\right)\varpi(c)dc$$

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

$$\times \frac{\mathbf{1}\{\mathfrak{X}_k \le \mathfrak{X}_j\} - \widetilde{V}(\mathfrak{X}_j, Z_j)}{f\left(Z_j^c, Z_j^d\right)} \frac{1}{h_0} K_0\left(\frac{Z_k^c - Z_j^c}{h_0}\right) \mathbf{1}\left\{Z_k^d = Z_j^d\right\},$$

$$\mathcal{F}_{8n}(\xi_i, \xi_j) = \widetilde{\beta}(\widetilde{V}_i) \times \Pi'\left(\widetilde{\beta}_0(\widetilde{V}_i) + X_i'\widetilde{\beta}(\widetilde{V}_i), \widetilde{V}_i\right) \widetilde{X}_i' \delta(\widetilde{V}_i)$$
$$\times \frac{\mathbf{1}\{\mathfrak{X}_j \le \mathfrak{X}_i\} - \widetilde{V}_i}{f\left(Z_i^c, Z_i^d\right)} \frac{1}{h_0} K_0\left(\frac{Z_j^c - Z_i^c}{h_0}\right) \mathbf{1}\left\{Z_j^d = Z_i^d\right\}.$$

By an identical argument,

$$L_3 = \frac{1}{n}\sum_{i=1}^n f_{6n}(\xi_i) + \frac{1}{n}\sum_{i=1}^n f_{7n}(\xi_i) + \frac{1}{n}\sum_{i=1}^n f_{8n}(\xi_i) + o_p(n^{-\frac{1}{2}}),$$

where these terms above are given by

$$f_{6n}(\xi_i) = \widetilde{\beta}(\widetilde{V}_i) \times \left(2 \times \mathbf{1}\{\widetilde{\varepsilon}(\widetilde{V}_i) < 0\} - 1\right)\left[-\sigma(\widetilde{V}_i)\right]^{-1}$$
$$\times \mathbb{E}_X\left[\Pi'\left(\widetilde{\beta}_0(\widetilde{V}_i) + X'\widetilde{\beta}(\widetilde{V}_i), \widetilde{V}_i\right) f(\widetilde{V}_i|X)\widetilde{X}'\right] \Gamma\left(\widetilde{X}_i, X_i'\widetilde{\beta}(\widetilde{V}_i) + \widetilde{\beta}_0(\widetilde{V}_i), \widetilde{V}_i\right),$$

$$f_{7n}(\xi_i) = \mathbb{E}_{X_j}\Bigg[\Upsilon_4\left(\widetilde{V}_{ji}, Z_i, \Lambda^{-1}\left(X_j'\widetilde{\beta}(\widetilde{V}_{ji}) + \widetilde{\beta}_0(\widetilde{V}_{ji}), \widetilde{V}_{ji}\right)\right) \Pi\left(\widetilde{\beta}_0(\widetilde{V}_{ji}) + X_j'\widetilde{\beta}(\widetilde{V}_{ji}), \widetilde{V}_{ji}\right)$$
$$\times \left[\mathbf{1}\{\mathfrak{X}_i \le \mathfrak{X}_j\} - \widetilde{V}(\mathfrak{X}_j, Z_i)\right] \frac{f\left(Z_i^c, Z_i^d|X_j\right)}{f\left(Z_i^c, Z_i^d\right)} \Big| \xi_i\Bigg],$$

$$f_{8n}(\xi_i) = \mathbb{E}_{X_j}\left[\widetilde{\beta}(\widetilde{V}_{ji}) \times \Pi'\left(\widetilde{\beta}_0(\widetilde{V}_{ji}) + X_j'\widetilde{\beta}(\widetilde{V}_{ji}), \widetilde{V}_{ji}\right) \widetilde{X}_j' \delta(\widetilde{V}_{ji}) \left[\mathbf{1}\{\mathfrak{X}_i \le \mathfrak{X}_j\} - \widetilde{V}_{ji}\right] \frac{f(Z_i^c, Z_i^d|X_j)}{f\left(Z_i^c, Z_i^d\right)} \Big| \xi_i\right],$$

and

$$\Upsilon_3(X_i, \widetilde{V}_i, \widetilde{X}_j, c) = \widetilde{\beta}(\widetilde{V}_i) \Pi'\left(\widetilde{\beta}_0(\widetilde{V}_i) + X_i'\widetilde{\beta}(\widetilde{V}_i), \widetilde{V}_i\right)\left[-\sigma(\widetilde{V}_i)\right]^{-1} f(\widetilde{V}_i|X_i) \widetilde{X}_i' \left(\widetilde{X}_j + \frac{\alpha_{XO}(c, \widetilde{V}_i)}{\alpha_{OO}(c, \widetilde{V}_i)}\right),$$

$$\Upsilon_4(\widetilde{V}, Z, c) = \frac{\partial}{\partial c}\left[\mu(\widetilde{V}, Z, c) \mathbb{E}_{X_i}\left[\Upsilon_3(X_i, \widetilde{V}, \widetilde{X}, c)\right]\left(X'\delta(\widetilde{V}) + \delta_\lambda(c, \widetilde{V})\right) \Pi'\left(\Lambda(c, \widetilde{V}), \widetilde{V}\right) \varpi(c)\right].$$

Similarly, we can derive that

$$L_4 = \frac{1}{n}\sum_{i=1}^n \int_{\underline{c}}^{\bar{c}} \frac{1}{h_3^2} K_3'\left(\frac{X_i'\widetilde{\beta}(\widetilde{V}_i) + \widetilde{\lambda}(c, \widetilde{V}_i)}{h_3}\right)\left[\widehat{\widetilde{\lambda}}\left(c, \widehat{\widetilde{V}}_i\right) - \widetilde{\lambda}(c, \widetilde{V}_i)\right] dc \times \widetilde{\beta}(\widetilde{V}_i) + o_p(n^{-\frac{1}{2}}).$$

Applying Lemmas A.7 and A.8, it suffices to show that

$$L_4 = \frac{1}{n(n-1)}\sum_{i \ne j} \mathcal{F}_{9n}(\xi_i, \xi_j) + \frac{1}{n(n-1)(n-2)}\sum_{i \ne j \ne k} \mathcal{F}_{10n}(\xi_i, \xi_j, \xi_k)$$
$$+ \frac{1}{n(n-1)}\sum_{i \ne j} \mathcal{F}_{11n}(\xi_i, \xi_j) + o_p(n^{-\frac{1}{2}}),$$

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

where

$$\mathcal{F}_{9n}(\xi_i.\xi_j) = \int_{\underline{c}}^{\overline{c}} \frac{1}{h_3^2} K_3' \left(\frac{X_i' \widetilde{\beta}(\widetilde{V}_i) + \widetilde{\lambda}(c, \widetilde{V}_i)}{h_3} \right) \frac{(2D_j(c) - 1)}{\alpha_{OO}(c, \widetilde{V}_i)} \frac{1}{h_1} K_1 \left(\frac{X_j' \widetilde{\beta}(\widetilde{V}_i) + \widetilde{\lambda}(c, \widetilde{V}_i)}{h_1} \right) dc$$
$$\times \frac{1}{h_2} K_2 \left(\frac{\widetilde{V}_j - \widetilde{V}_i}{h_2} \right) \widetilde{\beta}(\widetilde{V}_i),$$

$$\mathcal{F}_{10n}(\xi_i, \xi_j, \xi_k) = \int_{\underline{c}}^{\overline{c}} \frac{1}{h_3^2} K_3' \left(\frac{X_i' \widetilde{\beta}(\widetilde{V}_i) + \widetilde{\lambda}(c, \widetilde{V}_i)}{h_3} \right) \frac{(2D_j(c) - 1)}{\alpha_{OO}(c, \widetilde{V}_i)} \frac{1}{h_1} K_1 \left(\frac{X_j' \widetilde{\beta}(\widetilde{V}_i) + \widetilde{\lambda}(c, \widetilde{V}_i)}{h_1} \right) dc$$
$$\times \frac{1}{h_2^2} K_2' \left(\frac{\widetilde{V}_j - \widetilde{V}_i}{h_2} \right) \widetilde{\beta}(\widetilde{V}_i) \frac{\mathbf{1}\{\mathfrak{X}_k \le \mathfrak{X}_j\} - \widetilde{V}(\mathfrak{X}_j, Z_j)}{f(Z_j^c, Z_j^d)} \frac{1}{h_0} K_0 \left(\frac{Z_k^c - Z_j^c}{h_0} \right) \mathbf{1}\{Z_k^d = Z_j^d\},$$

$$\mathcal{F}_{11n}(\xi_i.\xi_j) = \int_{\underline{c}}^{\overline{c}} \frac{1}{h_3^2} K_3' \left(\frac{X_i' \widetilde{\beta}(\widetilde{V}_i) + \widetilde{\lambda}(c, \widetilde{V}_i)}{h_3} \right) \delta_\lambda(c, \widetilde{V}_i) dc \times \widetilde{\beta}(\widetilde{V}_i)$$
$$\times \frac{\mathbf{1}\{\mathfrak{X}_j \le \mathfrak{X}_i\} - \widetilde{V}(\mathfrak{X}_i, Z_i)}{f(Z_i^c, Z_i^d)} \frac{1}{h_0} K_0 \left(\frac{Z_j^c - Z_i^c}{h_0} \right) \mathbf{1}\{Z_j^d = Z_i^d\}.$$

Recall that $\delta_\lambda(c, v)$ is the first-order derivative of $\lambda(c, v)$ with respect to v. Applying the argument similar to the preceding one, we have

$$\frac{1}{n(n-1)} \sum_{i \neq j} \mathcal{F}_{9n}(\xi_i.\xi_j) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[\mathcal{F}_{9n}(\xi_i.\xi_j) | \xi_j \right] + o_p(n^{-\frac{1}{2}}),$$

$$\frac{1}{n(n-1)(n-2)} \sum_{i \neq j \neq k} \mathcal{F}_{10n}(\xi_i, \xi_j, \xi_k) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[\mathcal{F}_{10n}(\xi_i, \xi_j, \xi_k) | \xi_k \right] + o_p(n^{-\frac{1}{2}}),$$

$$\frac{1}{n(n-1)} \sum_{i \neq j} \mathcal{F}_{11n}(\xi_i.\xi_j) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[\mathcal{F}_{11n}(\xi_i.\xi_j) | \xi_j \right] + o_p(n^{-\frac{1}{2}}).$$

In evaluating $\mathbb{E}\left[\mathcal{F}_{9n}(\xi_i.\xi_j)|\xi_j\right]$, averaging over ξ_i will be done in four steps: (i) average over X_{1i} given $\widetilde{X}_i$ and $\widetilde{V}_i$; (ii) average over $\widetilde{V}_i$ given $\widetilde{X}_i$; (iii) solve the integral of c; and (iv) average over $\widetilde{X}_i$. Let us make a change of variable $th_3 = X_{1i} + \widetilde{X}_i' \widetilde{\beta}_2(\widetilde{V}_i) + \widetilde{\lambda}(c, \widetilde{V}_i)$. After applying step (i) and taking the Taylor expansion of $-\widetilde{X}_i' \widetilde{\beta}_2(\widetilde{V}_i) - \widetilde{\lambda}(c, \widetilde{V}_i)$, the integral is equal to

$$\int_{\underline{c}}^{\overline{c}} \left[-f'\left(-\widetilde{X}_i' \widetilde{\beta}_2(\widetilde{V}_i) - \widetilde{\lambda}(c, \widetilde{V}_i) | \widetilde{X}_i, \widetilde{V}_i \right) \right] \frac{(2D_j(c) - 1)}{\alpha_{OO}(c, \widetilde{V}_i)} \frac{1}{h_1} K_1 \left(\frac{X_j' \widetilde{\beta}(\widetilde{V}_i) + \widetilde{\lambda}(c, \widetilde{V}_i)}{h_1} \right)$$
$$\times \frac{1}{h_2} K_2 \left(\frac{\widetilde{V}_j - \widetilde{V}_i}{h_2} \right) \widetilde{\beta}(\widetilde{V}_i) dc,$$

plus a term with order $o_p(n^{-\frac{1}{2}})$, where $f'(\cdot|\cdot,\cdot)$ is the derivative of the density function of X_1 given $(\widetilde{X}, \widetilde{V})$, with respect to the first argument. Let us make a change of variable $th_2 = \widetilde{V}_i - \widetilde{V}_j$. After applying step (ii) and the Taylor expansion of $\widetilde{V}_j$, the integral is equal to

$$\int_{\underline{c}}^{\overline{c}} \left[-f'\left(-\widetilde{X}_i' \widetilde{\beta}_2(\widetilde{V}_j) - \widetilde{\lambda}(c, \widetilde{V}_j) | \widetilde{X}_i, \widetilde{V}_j \right) \right] \frac{(2D_j(c) - 1)}{\alpha_{OO}(c, \widetilde{V}_j)} \frac{1}{h_1} K_1 \left(\frac{X_j' \widetilde{\beta}(\widetilde{V}_j) + \widetilde{\lambda}(c, \widetilde{V}_j)}{h_1} \right) f(\widetilde{V}_j | \widetilde{X}_i) \widetilde{\beta}(\widetilde{V}_j) dc,$$

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

plus a term with order $o_p(n^{-\frac{1}{2}})$, as long as $\frac{h_2^{\kappa_2}}{h_1^{\kappa_2}}$ is of order $o_p(n^{-\frac{1}{2}})$. Let us make a change of variables $th_1 = X_j'\widetilde{\beta}(\widetilde{V}_j) + \widetilde{\beta}_0(\widetilde{V}_j) - \Lambda(c, \widetilde{V}_j)$. After applying step (iii), and Newey and McFadden (1994, Thm. 8.1), the integral is equal to

$$\left[-f'\left(X_{1j} + (\widetilde{X}_j - \widetilde{X}_i)'\widetilde{\beta}_2(\widetilde{V}_j)\middle|\widetilde{X}_i, \widetilde{V}_j\right)\right]\frac{2\times\mathbf{1}\{\widetilde{\varepsilon}(\widetilde{V}_j) < 0\} - 1}{\alpha_{OO}\left(\Lambda^{-1}\left(X_j'\widetilde{\beta}(\widetilde{V}_j) + \widetilde{\beta}_0(\widetilde{V}_j), \widetilde{V}_j\right), \widetilde{V}_j\right)}$$
$$\times\Pi\left(\widetilde{\beta}_0(\widetilde{V}_j) + X_j'\widetilde{\beta}(\widetilde{V}_j), \widetilde{V}_j\right)f(\widetilde{V}_j|\widetilde{X}_i)\,\widetilde{\beta}(\widetilde{V}_j),$$

plus a term with order $o_p(n^{-\frac{1}{2}})$. Finally, after applying step (iv), the integral is equal to

$$\mathbb{E}_{\widetilde{X}}\left[-f'\left(X_{1j} + (\widetilde{X}_j - \widetilde{X})'\widetilde{\beta}_2(\widetilde{V}_j)\middle|\widetilde{X}, \widetilde{V}_j\right)f(\widetilde{V}_j|\widetilde{X})\right]$$
$$\times\frac{2\times\mathbf{1}\{\widetilde{\varepsilon}(\widetilde{V}_j) < 0\} - 1}{\alpha_{OO}\left(\Lambda^{-1}\left(X_j'\widetilde{\beta}(\widetilde{V}_j) + \widetilde{\beta}_0(\widetilde{V}_j), \widetilde{V}_j\right), \widetilde{V}_j\right)}\Pi\left(\widetilde{\beta}_0(\widetilde{V}_j) + X_j'\widetilde{\beta}(\widetilde{V}_j), \widetilde{V}_j\right)\widetilde{\beta}(\widetilde{V}_j),$$

where $\mathbb{E}_{\widetilde{X}}$ means the expectation with respect to $\widetilde{X}$, and $f(\cdot)$ denotes the density function of $\widetilde{X}$. Thus, we show that

$$\frac{1}{n(n-1)}\sum_{i\neq j}\mathcal{F}_{9n}(\xi_i.\xi_j) = \frac{1}{n}\sum_{i=1}^{n} f_{9n}(\xi_i) + o_p(n^{-\frac{1}{2}}),$$

where

$$f_{9n}(\xi_i) = \mathbb{E}_{\widetilde{X}}\left[-f'\left(X_{1i} + (\widetilde{X}_i - \widetilde{X})'\widetilde{\beta}_2(\widetilde{V}_i)\middle|\widetilde{X}, \widetilde{V}_i\right)f(\widetilde{V}_i|\widetilde{X})\right]$$
$$\times\frac{2\times\mathbf{1}\{\widetilde{\varepsilon}(\widetilde{V}_i) < 0\} - 1}{\alpha_{OO}\left(\Lambda^{-1}\left(X_i'\widetilde{\beta}(\widetilde{V}_i) + \widetilde{\beta}_0(\widetilde{V}_i), \widetilde{V}_i\right), \widetilde{V}_i\right)}\Pi\left(\widetilde{\beta}_0(\widetilde{V}_i) + X_i'\widetilde{\beta}(\widetilde{V}_i), \widetilde{V}_i\right)\widetilde{\beta}(\widetilde{V}_i).$$

Recall that

$$\mathcal{F}_{10n}(\xi_i, \xi_j, \xi_k) = \int_{\underline{c}}^{\overline{c}}\frac{1}{h_3^2}K_3'\left(\frac{X_i'\widetilde{\beta}(\widetilde{V}_i) + \widetilde{\lambda}(c, \widetilde{V}_i)}{h_3}\right)\frac{(2D_j(c) - 1)}{\alpha_{OO}(c, \widetilde{V}_i)}\frac{1}{h_1}K_1\left(\frac{X_j'\widetilde{\beta}(\widetilde{V}_i) + \widetilde{\lambda}(c, \widetilde{V}_i)}{h_1}\right)dc$$
$$\times\frac{1}{h_2^2}K_2'\left(\frac{\widetilde{V}_j - \widetilde{V}_i}{h_2}\right)\widetilde{\beta}(\widetilde{V}_i)\frac{\mathbf{1}\{\mathfrak{X}_k \leq \mathfrak{X}_j\} - \widetilde{V}(\mathfrak{X}_j, Z_j)}{f\left(Z_j^c, Z_j^d\right)}\frac{1}{h_0}K_0\left(\frac{Z_k^c - Z_j^c}{h_0}\right)\mathbf{1}\left\{Z_k^d = Z_j^d\right\}.$$

In evaluating $\mathbb{E}\left[\mathcal{F}_{10n}(\xi_i, \xi_j, \xi_k)|\xi_k\right]$, we first fix ξ_j and average over ξ_i. Similar to evaluating $\mathbb{E}\left[\mathcal{F}_{9n}(\xi_i, \xi_j)|\xi_j\right]$, averaging over ξ_i will be done in three steps: (i) average over X_{1i} given $\widetilde{X}_i$ and $\widetilde{V}_i$; (ii) average over $\widetilde{V}_i$ given $\widetilde{X}_i$; and (iii) average over $\widetilde{X}_i$. Let us make a change of variables $th_3 = X_{1i} + \widetilde{X}_i'\widetilde{\beta}_2(\widetilde{V}_i) + \widetilde{\lambda}(c, \widetilde{V}_i)$. After applying step (i), and taking the Taylor

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

expansion of $-\widetilde{X}_i'\widetilde{\beta}_2\left(\widetilde{V}_i\right)-\widetilde{\lambda}\left(c,\widetilde{V}_i\right)$, the integral is equal to

$$\int_{\underline{c}}^{\overline{c}}\left[-f'\left(-\widetilde{X}_i'\widetilde{\beta}_2\left(\widetilde{V}_i\right)-\widetilde{\lambda}\left(c,\widetilde{V}_i\right)|\widetilde{X}_i,\widetilde{V}_i\right)\right]\frac{\left(2D_j(c)-1\right)}{\alpha_{OO}\left(c,\widetilde{V}_i\right)}\frac{1}{h_1}K_1\left(\frac{X_j'\widetilde{\beta}\left(\widetilde{V}_i\right)+\widetilde{\lambda}\left(c,\widetilde{V}_i\right)}{h_1}\right)dc$$
$$\times\frac{1}{h_2^2}K_2'\left(\frac{\widetilde{V}_j-\widetilde{V}_i}{h_2}\right)\widetilde{\beta}\left(\widetilde{V}_i\right)\frac{\mathbf{1}\left\{\mathfrak{X}_k\le\mathfrak{X}_j\right\}-\widetilde{V}\left(\mathfrak{X}_j,Z_j\right)}{f\left(Z_j^c,Z_j^d\right)}\frac{1}{h_0}K_0\left(\frac{Z_k^c-Z_j^c}{h_0}\right)\mathbf{1}\left\{Z_k^d=Z_j^d\right\},$$

plus a term with order $o_p(n^{-\frac{1}{2}})$, where $f'(\cdot|\cdot,\cdot)$ is the derivative of the density function of X_1 given $\left(\widetilde{X},\widetilde{V}\right)$, with respect to the first argument. Let us make a change of variables $th_2=\widetilde{V}_j-\widetilde{V}_i$. After applying step (ii), and taking the Taylor expansion of $\widetilde{V}_j$, the integral is equal to

$$\int_{\underline{c}}^{\overline{c}}\left(2D_j(c)-1\right)\left[\Upsilon_5'\left(\widetilde{V}_j,\widetilde{X}_i,c\right)\frac{1}{h_1}K_1\left(\frac{X_j'\widetilde{\beta}\left(\widetilde{V}_j\right)+\widetilde{\lambda}\left(c,\widetilde{V}_j\right)}{h_1}\right)\right.$$
$$\left.+\Upsilon_5\left(\widetilde{V}_j,\widetilde{X}_i,c\right)\frac{1}{h_1^2}K_1'\left(\frac{X_j'\widetilde{\beta}\left(\widetilde{V}_j\right)+\widetilde{\lambda}\left(c,\widetilde{V}_j\right)}{h_1}\right)\left(\widetilde{X}_j'\delta\left(\widetilde{V}_j\right)+\delta_\lambda\left(c,\widetilde{V}_j\right)\right)\right]dc$$
$$\times\frac{\mathbf{1}\left\{\mathfrak{X}_k\le\mathfrak{X}_j\right\}-\widetilde{V}\left(\mathfrak{X}_j,Z_j\right)}{f\left(Z_j^c,Z_j^d\right)}\frac{1}{h_0}K_0\left(\frac{Z_k^c-Z_j^c}{h_0}\right)\mathbf{1}\left\{Z_k^d=Z_j^d\right\},$$

plus a term with order $o_p(n^{-\frac{1}{2}})$, where $\Upsilon_5'\left(\widetilde{V},\widetilde{X},c\right)$ denotes the first-order derivative with respect to $\widetilde{V}$, and

$$\Upsilon_5\left(\widetilde{V}_j,\widetilde{X}_i,c\right)=\left[-f'\left(-\widetilde{X}_i'\widetilde{\beta}_2\left(\widetilde{V}_i\right)-\widetilde{\lambda}\left(c,\widetilde{V}_i\right)|\widetilde{X}_i,\widetilde{V}_i\right)\right]\frac{\widetilde{\beta}\left(\widetilde{V}_j\right)f\left(\widetilde{V}_j|\widetilde{X}_i\right)}{\alpha_{OO}\left(c,\widetilde{V}_j\right)},$$

where $f(\cdot|\cdot)$ denotes the density function of $\widetilde{V}$ conditional on $\widetilde{X}$. After applying step (iii), the integral is equal to

$$\int_{\underline{c}}^{\overline{c}}\left(2D_j(c)-1\right)\left[\mathbb{E}_{\widetilde{X}}\left[\Upsilon_5'\left(\widetilde{V}_j,\widetilde{X},c\right)\right]\frac{1}{h_1}K_1\left(\frac{X_j'\widetilde{\beta}\left(\widetilde{V}_j\right)+\widetilde{\lambda}\left(c,\widetilde{V}_j\right)}{h_1}\right)\right.$$
$$\left.+\mathbb{E}_{\widetilde{X}}\left[\Upsilon_5\left(\widetilde{V}_j,\widetilde{X},c\right)\right]\frac{1}{h_1^2}K_1'\left(\frac{X_j'\widetilde{\beta}\left(\widetilde{V}_j\right)+\widetilde{\lambda}\left(c,\widetilde{V}_j\right)}{h_1}\right)\left(\widetilde{X}_j'\delta\left(\widetilde{V}_j\right)+\delta_\lambda\left(c,\widetilde{V}_j\right)\right)\right]dc$$
$$\times\frac{\mathbf{1}\left\{\mathfrak{X}_k\le\mathfrak{X}_j\right\}-\widetilde{V}\left(\mathfrak{X}_j,Z_j\right)}{f\left(Z_j^c,Z_j^d\right)}\frac{1}{h_0}K_0\left(\frac{Z_k^c-Z_j^c}{h_0}\right)\mathbf{1}\left\{Z_k^d=Z_j^d\right\}.$$

Next, we average over ξ_j in four steps: (i) average over $D_j(c)$ given $\widetilde{V}_j$ and Z_j; (ii) solve the integral of c; (iii) average over Z_j given X_j; and (iv) average over X_j. After applying step (i),

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

the integral is equal to

$$\int_{\underline{c}}^{\overline{c}} \mu\left(\widetilde{V}_j, Z_j, c\right)\left[\mathbb{E}_{\widetilde{X}}\left[\Upsilon_5'\left(\widetilde{V}_j, \widetilde{X}, c\right)\right]\frac{1}{h_1}K_1\left(\frac{X_j'\widetilde{\beta}\left(\widetilde{V}_j\right)+\widetilde{\lambda}\left(c,\widetilde{V}_j\right)}{h_1}\right)\right.$$
$$\left.+\mathbb{E}_{\widetilde{X}}\left[\Upsilon_5\left(\widetilde{V}_j, \widetilde{X}, c\right)\right]\frac{1}{h_1^2}K_1'\left(\frac{X_j'\widetilde{\beta}\left(\widetilde{V}_j\right)+\widetilde{\lambda}\left(c,\widetilde{V}_j\right)}{h_1}\right)\left(\widetilde{X}_j'\delta\left(\widetilde{V}_j\right)+\delta_\lambda\left(c,\widetilde{V}_j\right)\right)\right]dc$$
$$\times\frac{\mathbf{1}\left\{\mathfrak{X}_k \le \mathfrak{X}_j\right\}-\widetilde{V}\left(\mathfrak{X}_j, Z_j\right)}{f\left(Z_j^c, Z_j^d\right)}\frac{1}{h_0}K_0\left(\frac{Z_k^c - Z_j^c}{h_0}\right)\mathbf{1}\left\{Z_k^d = Z_j^d\right\},$$

where $\mu\left(\widetilde{V}, Z, c\right) = 2\times Pr\left(\widetilde{\varepsilon}\left(\widetilde{V}\right) < X'\widetilde{\beta}\left(\widetilde{V}\right)+\widetilde{\lambda}\left(c,\widetilde{V}\right)\middle|\widetilde{V}, Z\right) - 1$. Let us make a change of variables $th_1 = X_j'\widetilde{\beta}\left(\widetilde{V}_j\right)+\widetilde{\beta}_0\left(\widetilde{V}_j\right)-\Lambda\left(c,\widetilde{V}_j\right)$. After applying step (ii), and taking the Taylor expansion, the integral is equal to

$$\Upsilon_6\left(\widetilde{V}_j, Z_j, \Lambda^{-1}\left(X_j'\widetilde{\beta}\left(\widetilde{V}_j\right)+\widetilde{\beta}_0\left(\widetilde{V}_j\right), \widetilde{V}_j\right)\right)\Pi\left(\widetilde{\beta}_0\left(\widetilde{V}_j\right)+X_j'\widetilde{\beta}\left(\widetilde{V}_j\right), \widetilde{V}_j\right)$$
$$\times\frac{\mathbf{1}\left\{\mathfrak{X}_k \le \mathfrak{X}_j\right\}-\widetilde{V}\left(\mathfrak{X}_j, Z_j\right)}{f\left(Z_j^c, Z_j^d\right)}\frac{1}{h_0}K_0\left(\frac{Z_k^c - Z_j^c}{h_0}\right)\mathbf{1}\left\{Z_k^d = Z_j^d\right\},$$

plus a term with order $o_p(n^{-\frac{1}{2}})$, where

$$\Upsilon_6\left(\widetilde{V}, Z, c\right) = \frac{\partial}{\partial c}\left[\mu\left(\widetilde{V}, Z, c\right)\mathbb{E}_{\widetilde{X}}\left[\Upsilon_5\left(\widetilde{V}, \widetilde{X}, c\right)\right]\left(\widetilde{X}'\delta\left(\widetilde{V}\right)+\delta_\lambda\left(c,\widetilde{V}\right)\right)\Pi\left(\Lambda\left(c,\widetilde{V}\right), \widetilde{V}\right)\right].$$

Let us make a change of variable $th_0 = Z_j^c - Z_k^c$. After applying step (iii), and taking the Taylor expansion of Z_k^c, the integral is equal to

$$\Upsilon_6\left(\widetilde{V}_{jk}, Z_k, \Lambda^{-1}\left(X_j'\widetilde{\beta}\left(\widetilde{V}_{jk}\right)+\widetilde{\beta}_0\left(\widetilde{V}_{jk}\right), \widetilde{V}_{jk}\right)\right)\Pi\left(\widetilde{\beta}_0\left(\widetilde{V}_{jk}\right)+X_j'\widetilde{\beta}\left(\widetilde{V}_{jk}\right), \widetilde{V}_{jk}\right)$$
$$\times\left[\mathbf{1}\left\{\mathfrak{X}_k \le \mathfrak{X}_j\right\}-\widetilde{V}\left(\mathfrak{X}_j, Z_k\right)\right]\frac{f\left(Z_k^c, Z_k^d\middle|X_j\right)}{f\left(Z_k^c, Z_k^d\right)},$$

plus a term with order $o_p(n^{-\frac{1}{2}})$, where $f(\cdot,\cdot|\cdot)$ denotes the density of $\left(Z_k^c, Z_k^d\right)$ conditional on X, and $\widetilde{V}_{jk}$ denotes $\widetilde{V}(\mathfrak{X}_j, Z_k)$. Finally, after applying step (iv), the integral is equal to

$$\mathbb{E}_{X_j}\left[\Upsilon_6\left(\widetilde{V}_{jk}, Z_k, \Lambda^{-1}\left(X_j'\widetilde{\beta}\left(\widetilde{V}_{jk}\right)+\widetilde{\beta}_0\left(\widetilde{V}_{jk}\right), \widetilde{V}_{jk}\right)\right)\Pi\left(\widetilde{\beta}_0\left(\widetilde{V}_{jk}\right)+X_j'\widetilde{\beta}\left(\widetilde{V}_{jk}\right), \widetilde{V}_{jk}\right)\right.$$
$$\left.\times\left[\mathbf{1}\left\{\mathfrak{X}_k \le \mathfrak{X}_j\right\}-\widetilde{V}\left(\mathfrak{X}_j, Z_k\right)\right]\frac{f\left(Z_k^c, Z_k^d\middle|X_j\right)}{f\left(Z_k^c, Z_k^d\right)}\middle|\xi_k\right].$$

Thus, we get that

$$\frac{1}{n(n-1)(n-2)}\sum_{i\neq j\neq k}\mathcal{F}_{10n}\left(\xi_i, \xi_j, \xi_k\right) = \frac{1}{n}\sum_{i=1}^{n} f_{10n}(\xi_i) + o_p(n^{-\frac{1}{2}}),$$

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

where $f_{10n}(\xi_i)$ is given by

$$f_{10n}(\xi_i) = \mathbb{E}_{X_j}\Bigg[\Upsilon_6\left(\widetilde{V}_{ji}, Z_i, \Lambda^{-1}\left(X_j'\widetilde{\beta}\left(\widetilde{V}_{ji}\right) + \widetilde{\beta}_0\left(\widetilde{V}_{ji}\right), \widetilde{V}_{ji}\right)\right)\Pi\left(\widetilde{\beta}_0\left(\widetilde{V}_{ji}\right) + X_j'\widetilde{\beta}\left(\widetilde{V}_{ji}\right), \widetilde{V}_{ji}\right)$$
$$\times\left[\mathbf{1}\left\{\mathfrak{X}_i \le \mathfrak{X}_j\right\} - \widetilde{V}\left(\mathfrak{X}_j, Z_i\right)\right]\frac{f\left(Z_i^c, Z_i^d|X_j\right)}{f\left(Z_i^c, Z_i^d\right)}\Big|\xi_i\Bigg].$$

Recall that

$$\mathcal{F}_{11n}(\xi_i.\xi_j) = \int_{\underline{c}}^{\overline{c}}\frac{1}{h_3^2}K_3'\left(\frac{X_i'\widetilde{\beta}\left(\widetilde{V}_i\right) + \widetilde{\lambda}\left(c, \widetilde{V}_i\right)}{h_3}\right)\delta_\lambda\left(c, \widetilde{V}_i\right)dc \times \widetilde{\beta}\left(\widetilde{V}_i\right)$$
$$\times\frac{\mathbf{1}\left\{\mathfrak{X}_j \le \mathfrak{X}_i\right\} - \widetilde{V}\left(\mathfrak{X}_i, Z_i\right)}{f\left(Z_i^c, Z_i^d\right)}\frac{1}{h_0}K_0\left(\frac{Z_j^c - Z_i^c}{h_0}\right)\mathbf{1}\left\{Z_j^d = Z_i^d\right\}.$$

In evaluating $\mathbb{E}\left[\mathcal{F}_{11n}\left(\xi_i, \xi_j\right)|\xi_j\right]$, averaging over ξ_i will be done in three steps: (i) solve the integral of c; (ii) average over Z_i given X_i; and (iii) average over X_i. Let us make a change of variables $th_3 = X_i'\widetilde{\beta}\left(\widetilde{V}_i\right) + \widetilde{\beta}_0\left(\widetilde{V}_i\right) - \Lambda\left(c, \widetilde{V}_i\right)$. After applying step (i), and the Taylor expansion to $X_i'\widetilde{\beta}\left(\widetilde{V}_i\right) + \widetilde{\beta}_0\left(\widetilde{V}_i\right)$, the integral is equal to

$$\Upsilon_7\left(X_i'\widetilde{\beta}\left(\widetilde{V}_i\right) + \widetilde{\beta}_0\left(\widetilde{V}_i\right), \widetilde{V}_i\right)\widetilde{\beta}\left(\widetilde{V}_i\right)\frac{\mathbf{1}\left\{\mathfrak{X}_j \le \mathfrak{X}_i\right\} - \widetilde{V}\left(\mathfrak{X}_i, Z_i\right)}{f\left(Z_i^c, Z_i^d\right)}\frac{1}{h_0}K_0\left(\frac{Z_j^c - Z_i^c}{h_0}\right)\mathbf{1}\left\{Z_j^d = Z_i^d\right\},$$

plus a term with order $o_p(n^{-\frac{1}{2}})$, where

$$\Upsilon_7\left(c, \widetilde{V}\right) = \frac{\partial}{\partial c}\left[\Pi\left(c, \widetilde{V}\right)\delta_\lambda\left(\Lambda^{-1}\left(c, \widetilde{V}\right), \widetilde{V}\right)\right].$$

Let us make a change of variable $th_0 = Z_i^c - Z_j^c$. After applying step (ii), and the Taylor expansion to Z_j^c, the integral is equal to

$$\Upsilon_7\left(X_i'\widetilde{\beta}\left(\widetilde{V}_{ij}\right) + \widetilde{\beta}_0\left(\widetilde{V}_{ij}\right), \widetilde{V}_{ij}\right)\widetilde{\beta}\left(\widetilde{V}_{ij}\right)\left[\mathbf{1}\left\{\mathfrak{X}_j \le \mathfrak{X}_i\right\} - \widetilde{V}\left(\mathfrak{X}_i, Z_j\right)\right]\frac{f\left(Z_j^c, Z_j^d|X_i\right)}{f\left(Z_j^c, Z_j^d\right)},$$

plus a term with order $o_p(n^{-\frac{1}{2}})$, where $f(\cdot, \cdot|\cdot)$ denotes the density of $\left(Z_k^c, Z_k^d\right)$ conditional on X, and $\widetilde{V}_{ij}$ denotes $\widetilde{V}(\mathfrak{X}_i, Z_j)$. Finally, after applying step (iv), the integral is equal to

$$\mathbb{E}_{X_i}\left[\Upsilon_7\left(X_i'\widetilde{\beta}\left(\widetilde{V}_{ij}\right) + \widetilde{\beta}_0\left(\widetilde{V}_{ij}\right), \widetilde{V}_{ij}\right)\widetilde{\beta}\left(\widetilde{V}_{ij}\right)\left[\mathbf{1}\left\{\mathfrak{X}_j \le \mathfrak{X}_i\right\} - \widetilde{V}\left(\mathfrak{X}_i, Z_j\right)\right]\frac{f\left(Z_j^c, Z_j^d|X_i\right)}{f\left(Z_j^c, Z_j^d\right)}\Big|\xi_j\right].$$

Thus, we get that

$$\frac{1}{n(n-1)}\sum_{i\neq j}\mathcal{F}_{11n}(\xi_i.\xi_j) = \frac{1}{n}\sum_{i=1}^{n}f_{11n}(\xi_i) + o_p(n^{-\frac{1}{2}}),$$

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

where

$$f_{11n}(\xi_i) = \mathbb{E}_{X_j}\left[\Upsilon_7\left(X_j'\widetilde{\beta}\left(\widetilde{V}_{ji}\right) + \widetilde{\beta}_0\left(\widetilde{V}_{ji}\right), \widetilde{V}_{ji}\right)\widetilde{\beta}\left(\widetilde{V}_{ji}\right)\left[\mathbf{1}\left\{\mathfrak{X}_i \leq \mathfrak{X}_j\right\} - \widetilde{V}\left(\mathfrak{X}_j, Z_i\right)\right]\frac{f\left(Z_i^c, Z_i^d | X_j\right)}{f\left(Z_i^c, Z_i^d\right)}\middle| \xi_i\right].$$

Thus, we can derive that

$$L_4 = \frac{1}{n}\sum_{i=1}^{n} f_{9n}(\xi_i) + \frac{1}{n}\sum_{i=1}^{n} f_{10n}(\xi_i) + \frac{1}{n}\sum_{i=1}^{n} f_{11n}(\xi_i) + o_p(n^{-\frac{1}{2}}).$$

To summarize, we have that

$$\begin{aligned}\widehat{\vartheta}_\kappa - \vartheta_\kappa =& \frac{1}{n}\sum_{i=1}^{n} f_{1n}(\xi_i) + \frac{1}{n}\sum_{i=1}^{n} f_{5n}(\xi_i) + \frac{1}{n}\sum_{i=1}^{n} f_{6n}(\xi_i) + \frac{1}{n}\sum_{i=1}^{n} f_{7n}(\xi_i) \\ &+ \frac{1}{n}\sum_{i=1}^{n} f_{8n}(\xi_i) + \frac{1}{n}\sum_{i=1}^{n} f_{9n}(\xi_i) + \frac{1}{n}\sum_{i=1}^{n} f_{10n}(\xi_i) + \frac{1}{n}\sum_{i=1}^{n} f_{11n}(\xi_i) + o_p(n^{-\frac{1}{2}}).\end{aligned}$$

The desired result follows from applying the Central Limit Theorem

$$\sqrt{n}\left(\widehat{\vartheta}_\kappa - \vartheta_\kappa\right) \sim \mathcal{N}\left(0, \Sigma_\vartheta\right),$$

where $\Sigma_\vartheta = \mathbb{E}\left[f_\vartheta(\xi_i) f_\vartheta(\xi_i)'\right]$ with $f_\vartheta(\xi_i) = \frac{1}{n}\sum_{i=1}^{n} f_{1n}(\xi_i) + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=5}^{11} f_{jn}(\xi_i)$. ■

B. Validity of the Bootstrap

In practice, inference based on combining an asymptotic normality result (Theorem 4.1) with an estimate of the asymptotic variance can be overly complicated. In such cases, resampling techniques like the ordinary nonparametric bootstrap can be used to obtain confidence regions for the parameters of interest. Using results from Chen et al. (2003; hereafter CLV), we can establish the validity of this approach. This section is structured as follows. First, we describe the general result for the asymptotic validity of the bootstrap in CLV. Applying this general result, we state the sufficient conditions for the asymptotic validity of the bootstrap in our case, and finally summarize the discussion in Theorem B.1.

Let Θ denote a finite dimensional parameter set and $\mathcal{H}$ an infinite dimensional parameter set. CLV assumes there exists a nonrandom measurable vector-valued function $\mathcal{M} : \mathbb{R}^K \times \mathcal{H} \to \mathbb{R}^L$, with $L \geq K$, such that the moment condition

$$\mathcal{M}(\theta, \mathfrak{h}(\cdot, \theta)) = 0$$

holds, where $\theta \in \Theta$ and $\mathfrak{h}(\cdot, \theta) \in \mathcal{H}$ are the true unknown finite and infinite dimensional parameters. Let $\widehat{\mathfrak{h}}(\cdot, \theta)$ denote an initial nonparametric estimator of $\mathfrak{h}(\cdot, \theta)$, and define $\widehat{\theta}$ as the estimator of θ, by approximately solving the sample minimization problem:

$$\widehat{\theta} = \min_{\theta \in \Theta} \left\| \mathcal{M}_n\left(\theta, \widehat{\mathfrak{h}}(\cdot, \theta)\right) \right\|.$$

They denote $\widehat{\mathfrak{h}}^*(\cdot, \theta)$ and $\widehat{\theta}^*$ as the estimators of $\mathfrak{h}(\cdot, \theta)$ and θ, respectively, based on the bootstrap data. The validity of the bootstrap procedure in CLV relies on two key conditions. First, the convergence rate of the bootstrap estimator $\widehat{\mathfrak{h}}^*(\cdot, \theta)$ should be faster than $n^{-1/4}$ (Assumption 2.4(B) in Theorem B in CLV). Second, the recentered sample moment function

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

can be represented as a smooth linear functional of $\widehat{\mathfrak{h}}^*(\cdot,\theta)-\widehat{\mathfrak{h}}(\cdot,\theta)$ (Assumption 2.6(B) in Theorem B in CLV).

Our estimator can be viewed as a special case of CLV. The parameter of interest ϑ_κ satisfies the moment condition

$$M(\mathfrak{h}(\cdot))-\vartheta_\kappa=0,$$

where $\mathfrak{h}(\cdot)=\left(\widetilde{\beta}_0(\cdot),\widetilde{\beta}_2(\cdot),\Lambda(\cdot,\cdot),\widetilde{V}(\cdot,\cdot)\right)$, and

$$M(\mathfrak{h})=\mathbb{E}\left[\frac{1}{\Lambda'\left(\Lambda^{-1}\left(\widetilde{\beta}_0\left(\widetilde{V}\right)+X_1+\widetilde{X}'\widetilde{\beta}_2\left(\widetilde{V}\right),\widetilde{V}\right),\widetilde{V}\right)}\times t_\kappa(\mathfrak{X})t_\kappa\left(Z^c\right)\right].$$

We can follow similar steps to those in CLV to show the validity of the bootstrap for our estimator. Let $\widehat{\mathfrak{h}}(\cdot)=\left(\widehat{\widetilde{\beta}}_0(\cdot),\widehat{\widetilde{\beta}}_2(\cdot),\widehat{\Lambda}(\cdot,\cdot),\widehat{\widetilde{V}}(\cdot,\cdot)\right)$, and let $M_n(\mathfrak{h})$ represent the sample average of $M(\mathfrak{h})$. Then, $M_n\left(\widehat{\mathfrak{h}}\right)=\widehat{\vartheta}_\kappa$. Let $\left\{\xi_i^*\right\}_{i=1}^n$ be drawn with replacement from the original sample $\{\xi_i\}_{i=1}^n=\{Y_i,X_i,Z_i\}_{i=1}^n$. Let $\widehat{\mathfrak{h}}^*,\widehat{\vartheta}_\kappa^*$ be the estimators of $\mathfrak{h}(\cdot),\vartheta_\kappa$, and $M_n^*(\mathfrak{h})$ be the sample average of $M(\mathfrak{h})$, but all based on the bootstrap data. We have $M_n^*\left(\widehat{\mathfrak{h}}^*\right)=\widehat{\vartheta}_\kappa^*$. General sufficient conditions for the asymptotic validity of the bootstrap were studied by CLV. We specialize these assumptions to our case, as summarized below:

Assumption B.1. $\vartheta_\kappa\in\operatorname{int}(\Theta)$ satisfies $M(\mathfrak{h})-\vartheta_\kappa=0$, and $\widehat{\vartheta}_\kappa\to^{a.s.}\vartheta_\kappa$.

Assumption B.2. (i) $\widehat{\mathfrak{h}}\in\mathcal{H}$ almost surely, and $\sup_{c\in(\underline{c},\bar{c}),x\in\mathfrak{X}_\kappa,z\in\mathcal{Z}_\kappa}|\widehat{\mathfrak{h}}-\mathfrak{h}|=o_{a.s.}(n^{-1/4})$. (ii) $\widehat{\mathfrak{h}}^*\in\mathcal{H}$ with P^* probability tending to one, and $\sup_{c\in(\underline{c},\bar{c}),x\in\mathfrak{X}_\kappa,z\in\mathcal{Z}_\kappa}\left|\widehat{\mathfrak{h}}^*-\widehat{\mathfrak{h}}\right|=o_{p^*}\left(n^{-1/4}\right)$.

Assumption B.3.

$$\sqrt{n}\left\{M_n^*\left(\widehat{\mathfrak{h}}\right)-M_n\left(\widehat{\mathfrak{h}}\right)+\Psi\left(\widehat{\mathfrak{h}}\right)\left[\widehat{\mathfrak{h}}^*-\widehat{\mathfrak{h}}\right]\right\}\to^d\mathcal{N}(0,V),$$

in P^*-probability, where $\Psi\left(\widehat{\mathfrak{h}}\right)\left[\widehat{\mathfrak{h}}^*-\widehat{\mathfrak{h}}\right]$ is the pathwise derivative of $M\left(\widehat{\mathfrak{h}}\right)$ at $\widehat{\mathfrak{h}}$ in the direction $\left[\widehat{\mathfrak{h}}^*-\widehat{\mathfrak{h}}\right]$, such that

$$\Psi\left(\widehat{\mathfrak{h}}\right)\left[\widehat{\mathfrak{h}}^*-\widehat{\mathfrak{h}}\right]=\lim_{\tau\to0}\left[M\left(\widehat{\mathfrak{h}}+\tau\left(\widehat{\mathfrak{h}}^*-\widehat{\mathfrak{h}}\right)\right)-M\left(\widehat{\mathfrak{h}}\right)\right]/\tau.$$

Let $\mathcal{H}_{\delta_n}=\left\{t\in\mathcal{H}:\sup_{c\in(\underline{c},\bar{c}),x\in\mathfrak{X}_\kappa,z\in\mathcal{Z}_\kappa}|t-\mathfrak{h}|\le\delta_n\right\}$. $\|A\|=(tr(A'A))^{1/2}$ for any matrix A.

Assumption B.4. For all $\bar{\mathfrak{h}}\in\mathcal{H}_{\delta_n}$ with a positive sequence $\delta_n=o(1)$, (i) the pathwise derivative $\Psi\left(\bar{\mathfrak{h}}\right)\left[\bar{\bar{\mathfrak{h}}}-\bar{\mathfrak{h}}\right]$ exists in all directions $\left[\bar{\bar{\mathfrak{h}}}-\bar{\mathfrak{h}}\right]\in\mathcal{H}$; (ii) For all $\bar{\bar{\mathfrak{h}}}\in\mathcal{H}_{\delta_n}$ and

$$\left\|M\left(\bar{\bar{\mathfrak{h}}}\right)-M\left(\bar{\mathfrak{h}}\right)-\Psi\left(\bar{\mathfrak{h}}\right)\left[\bar{\bar{\mathfrak{h}}}-\bar{\mathfrak{h}}\right]\right\|\le C\sup_{c\in(\underline{c},\bar{c}),x\in\mathfrak{X}_\kappa,z\in\mathcal{Z}_\kappa}\left|\bar{\bar{\mathfrak{h}}}-\bar{\mathfrak{h}}\right|^2$$

with C being a positive constant.

Assumption B.5. (i) $\sup_{(\bar{\mathfrak{h}}-\mathfrak{h})\in\mathcal{H}_{\delta_n}}\left|M_n\left(\bar{\mathfrak{h}}\right)-M\left(\bar{\mathfrak{h}}\right)-\left(M_n(\mathfrak{h})-\vartheta_\kappa\right)\right|=o_{a.s.}\left(n^{-1/2}\right)$ for any positive sequence $\delta_n=o(1)$. (ii) $\sup_{(\bar{\mathfrak{h}}-\mathfrak{h})\in\mathcal{H}_{\delta_n}}|M_n^*(\bar{\mathfrak{h}})-M_n(\bar{\mathfrak{h}})-(M_n^*(\mathfrak{h})-M_n(\mathfrak{h}))|=o_{p^*}(n^{-1/2})$ for any positive sequence $\delta_n=o(1)$.

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

Remark B.1. Assumption B.1 requires that $\widehat{\vartheta}_\kappa$ is consistent, which is easily satisfied under the assumptions of Theorem 4.1. Assumption B.2(i) can be verified under the same assumptions as those in Theorem 4.1. In particular, it can be obtained by applying similar reasoning to that in Lemmas A.1 and A.5–A.8, and the U-statistic projection procedure. Moreover, Assumption B.2(ii) can be verified under the same assumptions implying Assumption B.2(i) for a variety of kernel density and regression estimators. See, for example, Hall (1991). Note that we have shown, by the proof of Theorem 4.1, that

$$\sqrt{n}\left\{M_n(\mathfrak{h}) - \vartheta_\kappa + \Psi(\mathfrak{h})\left[\widehat{\mathfrak{h}} - \mathfrak{h}\right]\right\} \to^d \mathcal{N}(0, \Sigma_\vartheta). \tag{B.1}$$

Using the discussion after Theorem B in CLV, and by applying a result from Gine and Zinn (1990), Assumption B.3 can be verified in the same way we establish (B.1), with $V = \Sigma_\vartheta$. Assumption B.4 can be verified under the continuous differentiability property of kernel functions in Assumption 4.4. Theorem 3 in CLV gives some sufficient conditions to verify Assumption B.5. The smoothness conditions in Assumptions 4.3 and 4.5 imply that the moment function $M(\mathfrak{h}(\cdot)) - \vartheta_\kappa$ is sufficiently continuous in $\mathfrak{h}$ and ϑ_κ, ensuring that the conditions in Theorem 3 in CLV hold.

THEOREM A.1. *Suppose that Assumptions B.1–B.5 hold, then*

$$\sqrt{n}\left(\widehat{\vartheta}_\kappa^* - \widehat{\vartheta}_\kappa\right) \to^d \mathcal{N}(0, V)$$

in P^*-probability.

Proof. Denote $\nu_n^*(\mathfrak{h}) = \sqrt{n}\left(M_n^*(\mathfrak{h}) - M_n(\mathfrak{h})\right)$ and $\nu_n(\mathfrak{h}) = \sqrt{n}(M_n(\mathfrak{h}) - M(\mathfrak{h}))$. By the triangle inequality and Assumption B.5, we have, for each positive sequence $\delta_n = o(1)$,

$$\begin{aligned} &\sup_{\left(\bar{\mathfrak{h}}, \underline{\mathfrak{h}}\right) \in \mathcal{H}_{\delta_n} \times \mathcal{H}_{\delta_n}} \left|\nu_n^*\left(\bar{\mathfrak{h}}\right) - \nu_n^*\left(\underline{\mathfrak{h}}\right)\right| = o_{p^*}(1), \\ &\sup_{\left(\bar{\mathfrak{h}}, \underline{\mathfrak{h}}\right) \in \mathcal{H}_{\delta_n} \times \mathcal{H}_{\delta_n}} \left|\nu_n\left(\bar{\mathfrak{h}}\right) - \nu_n\left(\underline{\mathfrak{h}}\right)\right| = o_{a.s.}(1). \end{aligned} \tag{B.2}$$

Applying similar arguments to those in the proof of Theorem B in CLV, we can approximate $\widehat{\vartheta}_\kappa^* - \widehat{\vartheta}_\kappa$ with a linear function $\mathcal{L}_n^* = M_n^*\left(\widehat{\mathfrak{h}}\right) - M_n\left(\widehat{\mathfrak{h}}\right) + \Psi\left(\widehat{\mathfrak{h}}\right)\left[\widehat{\mathfrak{h}}^* - \widehat{\mathfrak{h}}\right]$ plus an error $o_{p^*}\left(n^{-1/2}\right)$. By adding and subtracting the triangle inequality,

$$\begin{aligned} &\left\|\widehat{\vartheta}_\kappa^* - \widehat{\vartheta}_\kappa - \mathcal{L}_n^*\right\| = \left\|M_n^*\left(\widehat{\mathfrak{h}}^*\right) - M_n\left(\widehat{\mathfrak{h}}\right) - \mathcal{L}_n^*\right\| \\ =&\Big\|M\left(\widehat{\mathfrak{h}}^*\right) - M\left(\widehat{\mathfrak{h}}\right) - \Psi\left(\widehat{\mathfrak{h}}\right)\left[\widehat{\mathfrak{h}}^* - \widehat{\mathfrak{h}}\right] + \left[M_n\left(\widehat{\mathfrak{h}}^*\right) - M\left(\widehat{\mathfrak{h}}^*\right) - M_n\left(\widehat{\mathfrak{h}}\right) + M\left(\widehat{\mathfrak{h}}\right)\right] \\ &+\left[M_n^*\left(\widehat{\mathfrak{h}}^*\right) - M_n\left(\widehat{\mathfrak{h}}^*\right) - M_n^*\left(\widehat{\mathfrak{h}}\right) + M_n\left(\widehat{\mathfrak{h}}\right)\right]\Big\| \\ \le& C \sup_{c,x,z} \left|\widehat{\mathfrak{h}}^* - \widehat{\mathfrak{h}}\right|^2 + n^{-1/2}\left\|\nu_n\left(\widehat{\mathfrak{h}}^*\right) - \nu_n\left(\widehat{\mathfrak{h}}\right)\right\| + n^{-1/2}\left\|\nu_n^*\left(\widehat{\mathfrak{h}}^*\right) - \nu_n^*\left(\widehat{\mathfrak{h}}\right)\right\| \\ =&o_{p^*}\left(n^{-1/2}\right), \end{aligned}$$

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

where the order of the error $o_{p^*}\left(n^{-1/2}\right)$ follows from Assumptions B.2 and B.4 and the stochastic equicontinuity properties (B.2). It follows that

$$\sqrt{n}\left(\widehat{\vartheta}_\kappa^* - \widehat{\vartheta}_\kappa\right) = \sqrt{n}\left\{M_n^*\left(\widehat{\mathfrak{h}}\right) - M_n\left(\widehat{\mathfrak{h}}\right) + \Psi\left(\widehat{\mathfrak{h}}\right)\left[\widehat{\mathfrak{h}}^* - \widehat{\mathfrak{h}}\right]\right\} + o_{p^*}(1).$$

The desired result then follows from Assumption B.3. ■

REFERENCES

Abrevaya, J. (1999) Rank estimation of a transformation model with observed truncation. *Econometrics Journal* 2(2), 292–305.

Abrevaya, J. (2003) Pairwise-difference rank estimation of the transformation model. *Journal of Business and Economic Statistics* 21(3), 437–447.

Asparouhova, E., R. Golanski, K. Kasprzyk, & S. T. Asparouhov (2002) Rank estimators for a transformation model. *Econometric Theory* 18(5), 1099–1120.

Blundell, R., X. Chen, & D. Kristensen (2007) Semi-nonparametric IV estimation of shape-invariant Engel curves. *Econometrica* 75(6), 1613–1669.

Box, G. E. P. & D. R. Cox (1964) An analysis of transformations. *Journal of the Royal Statistical Society* 26(2), 211–252.

Buchinsky, M. (1995) Quantile regression, Box–Cox transformation model, and the U.S. wage structure, 1963–1987. *Journal of Econometrics* 65(1), 109–154.

Carroll, R. J., Fan, J., Gijbels, I. & Wand, M. P. (1997) Generalized partially linear single-index models. *Journal of the American Statistical Association* 92(438), 477–489.

Cavanagh, C. & R. P. Sherman (1998) Rank estimators for monotonic index models. *Journal of Econometrics* 84(2), 351–381.

Chen, K., Z. Jin, & Z. Ying (2002) Semiparametric analysis of transformation models with censored data. *Biometrika* 89(3), 659–668.

Chen, S. (2002) Distribution-free estimation of the Box–Cox regression model with censoring. *Econometric Theory* 28(3), 680–695.

Chen, S. (2010) An integrated maximum score estimator for a generalized censored quantile regression model. *Journal of Econometrics* 155(1), 90–98.

Chen, X., O. Linton & I. Van Keilegom (2003) Estimation of semiparametric models when the criterion function is not smooth. *Econometrica* 71(5), 1591–1608.

Chernozhukov, V., I. Fernández-Val, & B. Melly (2013) Inference on counterfactual distributions. *Econometrica* 81(6), 2205–2268.

Chesher, A. (2003) Identification in nonseparable models. *Econometrica* 71(5), 1405–1441.

Chiappori, P. A., I. Komunjer & D. Kristensen (2015) Nonparametric identification and estimation of transformation models. *Journal of Econometrics* 188(1), 22–39.

Feve, F. & J. P. Florens (2010) The practice of non-parametric estimation by solving inverse problems: The example of transformation models. *Econometrics Journal* 13(3), S1–S27.

Florens, J. P., J. J. Heckman, C. Meghir & E. Vytlacil (2008) Identification of treatment effects using control functions in model with continuous, endogenous treatment and heterogeneous effects. *Econometrica* 76, 1191–1206.

Florens, J. P. & S. Sokullu (2017) Nonparametric estimation of semiparametric transformation models. *Econometric Theory* 33(4), 35.

Gine, E. & J. Zinn (1990) Bootstrapping general empirical measures. *The Annals of Probability* 18(2), 851–869.

Hall, P. (1991) On convergence rates of suprema. *Probability Theory and Related Fields* 89(4), 447–455.

Heckman, J. & E. Vytlacil (1998) Instrumental variables methods for the correlated random coefficient model: Estimating the average rate of return to schooling when the return is correlated with schooling. *The Journal of Human Resources* 33(4), 974–987.

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

Heckman, J. & E. Vytlacil (2005) Structural equations, treatment effects, and econometric policy evaluation. *Econometrica* 73(3), 669–738.
Heckman, J. & E. Vytlacil (2007) Econometric evaluation of social programs, Part I: Causal models, structural models and econometric policy evaluation. In J. J. Heckman & E. E. Leamer (eds.), *Handbook of Econometrics*, vol. 6, pp. 4779–4874. Elsevier.
Hoderlein, S. & R. Sherman (2015) Identification and estimation in a correlated random coefficients binary response model. *Journal of Econometrics* 188, 135–149.
Horowitz, J. L. (1992) A smoothed maximum score estimator for the binary response model. *Econometrica* 60(3), 505–531.
Horowitz, J. L. (1996) Semiparametric estimation of a regression model with an unknown transformation of the dependent variable. *Econometrica* 64, 103–137.
Horowitz, J. L. (1998) *Semiparametric Methods in Econometrics*. Springer, New York.
Horowitz, J. L. (2001) Nonparametric estimation of a generalized additive model with an unknown link function. *Econometrica* 69(2), 499–513.
Imbens, G. & W. Newey (2009) Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica* 77(5), 1481–1512.
Jacho-Chávez, D., A. Lewbel & O. Linton (2010) Identification and nonparametric estimation of a transformed additively separable model. *Journal of Econometrics* 156(2), 392–407.
Khan, S., & E. Tamer (2007) Partial rank estimation of duration models with general forms of censoring. *Journal of Econometrics* 136(1), 251–280.
Lewbel, A. & K. Pendakur (2017) Unobserved preference heterogeneity in demand using generalized random coefficients. *Journal of Political Economy* 125(4), 1100–1148.
Li, Q. & J. S. Racine (2007) *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.
Linton, O., S. Sperlich, & I. Van Keilegom (2008) Estimation of a semiparametric transformation model. *Annals of Statistics* 36(2), 686–718.
Manski, C. F. (1975) Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics* 3(3), 205–228.
Manski, C. F. (1985) Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator. *Journal of Econometrics* 27(3), 313–333.
Masten, M. & A. Torgovitsky (2016) Identification of instrumental variable correlated random coefficients models. *Review of Economics and Statistics* 98(5), 1001–1005.
Moon, J.M. (2013) Sieve Extremum Estimation of Transformation Models. Technical report, UCSD, Working paper.
Newey, W. K. & McFadden, D. (1994) Large sample estimation and hypothesis testing. In Engle, R. & McFadden, D. (eds.), *Handbook of Econometrics*, vol. 4. Elsevier Science.
Powell, J. L., J. H. Stock, & T. M. Stoker (1989) Semiparametric estimation of index coefficients. *Econometrica* 57(6), 1403–1430.
Sherman, R. P. (1994a) U-processes in the analysis of a generalized semiparametric regression estimator. *Econometric Theory* 10(2), 372–395.
Sherman, R. P. (1994b) Maximal inequalities for degenerate U-processes with applications to optimization estimators. *Annals of Statistics* 22(1), 439–459.
Shin, Y. (2010) Local rank estimation of transformation models with functional coefficients. *Econometric Theory* 26(6), 1807–1819.
Vanhems, A. & I. Van Keilegom (2018) Estimation of a semiparametric transformation model in the presence of endogeneity. *Econometric Theory* 35(1), 73–110.
Wooldridge, J. (1997) On two stage least squares estimation of the average treatment effect in a random coefficient model. *Economics Letters* 56, 129–133.
Wooldridge, J. (2003) Further results on instrumental variables estimation of average treatment effects in the correlated random coefficient model. *Economics Letters* 79, 185–191.
Wooldridge, J. (2008) Instrumental variables estimation of the average treatment effect in the correlated random coefficient model. In D. Millimet, J. Smith, & E. Vytlacil (eds.), *Advances in Econometrics:*

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press

Modeling and Evaluating Treatment Effects in Econometrics, vol. 21, pp. 93–116. Emerald Group Publishing Limited.

Xu, X. & L. F. Lee (2015) A spatial autoregressive model with a nonlinear transformation of the dependent variable. *Journal of Econometrics* 186(1), 1–18.

Zhang, Z. (2016) Semiparametric estimation of partially linear transformation models under conditional quantile restriction. *Econometric Theory* 32(2), 458–497.

https://doi.org/10.1017/S0266466621000281 Published online by Cambridge University Press