Journal of Statistical Planning and Inference 208 (2020) 94–118

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Estimation of a partially linear additive model with generated covariates[☆]

Xin Geng[a], Carlos Martins-Filho[b,*], Feng Yao[c,d]

[a] School of Finance, Nankai University, Tianjin 300350, PR China
[b] Department of Economics, University of Colorado, Boulder, CO 80309-0256, USA
[c] Department of Economics, West Virginia University, Morgantown, WV 26505, USA
[d] School of Economics and Trade, Guangdong University of Foreign Studies, Guangzhou, Guangdong 510006, PR China

ARTICLE INFO

Article history:
Received 24 October 2018
Received in revised form 16 October 2019
Accepted 5 February 2020
Available online 17 February 2020

JEL classification:
C14
C36

MSC:
primary 62G07
62G08
62G20

Keywords:
Partially linear regression
Additive nonparametric regression
Generated covariates

ABSTRACT

We propose kernel-based estimators for *both* the parametric and nonparametric components of a partially linear additive regression model where a subset of the covariates entering the nonparametric component are generated by the estimation of an auxiliary nonparametric regression. Both estimators are shown to be asymptotically normally distributed. The estimator for the finite dimensional parameter is shown to converge at the parametric $\sqrt{n}$ rate and the estimator for the infinite dimensional parameter converges at a slower nonparametric rate that, as usual, depends on the rate of decay of the bandwidths and the dimensionality of the underlying regression. A small Monte Carlo study is conducted to shed light on the finite sample performance of our estimators and to contrast them with those of estimators available in the extant literature.

1. Introduction

The estimation of partially linear regression models has been the subject of a large literature since the seminal work of Heckman (1986) and Robinson (1988). See, e.g., Chamberlain (1992), Cuzick (1992), Linton (1995), Fan et al. (1998), Fan and Li (1999) and Juhl and Xiao (2005) among many in this literature. A number of papers (Newey et al., 1999; Li and Wooldridge, 2000; Pinkse, 2000; Fan and Li, 2003; Manzan and Zerom, 2005; Su and Ullah, 2008; Yu et al., 2011; Martins-Filho and Yao, 2012) have considered the estimation and asymptotic efficiency gains that may result from knowledge that the nonparametric component of a partially linear regression model has a partial or fully additive structure. Generally speaking, this additivity can emerge directly from primitive assumptions (Li and Wooldridge, 2000; Manzan and Zerom, 2005) or from the specification of systems of regressions that lead to additivity of the regression of interest. A frequently

☆ We would like to thank an associate editor, a referee and seminar and conference participants at the European, China, and Asian Meetings of the Econometric Society, 3rd Conference of the International Society for Nonparametric Statistics, University of Colorado at Boulder, and IFPRI, for their helpful comments and suggestions. Xin Geng's research was supported by the Fundamental Research Funds for the Central Universities 63192145. All remaining errors are ours.
* Corresponding author.
E-mail addresses: xin.geng@nankai.edu.cn (X. Geng), carlos.martins@colorado.edu (C. Martins-Filho), feng.yao@mail.wvu.edu, 201470006@oamail.gdufs.edu.cn (F. Yao).

https://doi.org/10.1016/j.jspi.2020.02.002
0378-3758/

occurring example of the latter case is econometric models that include "endogenous" covariates, where identification and estimation result from the specification of control functions (Newey et al., 1999; Pinkse, 2000; Su and Ullah, 2008; Martins-Filho and Yao, 2012; Ozabaci et al., 2014). Hence, consider the following partially linear triangular simultaneous equations model

$$Y = \beta_0 + X_2'\beta + m(X_1, Z_1) + \varepsilon, \tag{1}$$

$$X = \Pi(Z) + U, \tag{2}$$

where Y is a scalar regressand, $Z_1 \in \mathbb{R}^{D_{11}}$ is a subvector of $Z = (Z_1', Z_2')' \in \mathbb{R}^{D_1}$ with $D_1 = D_{11} + D_{12}$, X_1 and X_2 are non-overlapping subvectors of $X \in \mathbb{R}^{D_2}$ of dimensions D_{21} and D_{22} with $D_2 = D_{21} + D_{22}$, $U \in \mathbb{R}^{D_2}$, $\mathrm{E}(U|Z) = 0$ and $\mathrm{E}(\epsilon|Z, U) = \mathrm{E}(\epsilon|U) \equiv g(U)$. Note that by subtracting $E(Y|Z, U)$ from Eq. (1) and defining $v = Y - \mathrm{E}(Y|Z, U)$ we obtain

$$Y = \beta_0 + X_2'\beta + m(X_1, Z_1) + g(U) + v. \tag{3}$$

In this paper, our primary goal is to propose and study the asymptotic properties of estimators for *both* the finite dimensional parameter $(\beta_0, \beta')'$ and the infinite dimensional parameter $m(\cdot)$ in Eq. (3). A complicating factor in the estimation of this model is that some of the covariates appearing in the additive nonparametric component are not observed (U) and must be generated (estimated) by the auxiliary regression (2).

The asymptotic properties of a general class of estimators of the finite dimensional parameter of semiparametric models with generated covariates have been recently studied by Mammen et al. (2016). They establish consistency and $\sqrt{n}$ asymptotic normality of a parametric estimator that is obtained by a three-step estimation procedure. The first step involves nonparametric estimation (generation) of the unobserved covariates. The second step involves nonparametric estimation of a nuisance regression that may depend on the finite dimensional parameter (profiling), and the third step involves the minimization of a GMM-type objective function based on the first two steps. Although the class of semiparametric models they consider includes the partially linear additive nonparametric regression with generated regressors we study in this paper, their estimation procedure fails to incorporate the additivity that is present in our model. In contrast, the estimators we propose for *both* the finite and infinite dimensional parameters make full use of the additive structure of the nonparametric component. In this regard, our paper is more closely related to Newey et al. (1999) or Su and Ullah (2008), although the moment conditions we use in motivating and deriving our estimators are different from those used in these papers.

The moment conditions we use to motivate our estimators are similar, but critically different, from those employed by Manzan and Zerom (2005) to estimate a model that is identical to our Eq. (3) but where the covariates U are observed. Hence, the asymptotic normality we obtain for our estimator of the finite dimensional parameter can be viewed as an extension of their main result to the case of a partially linear regression model with generated regressors. In fact, as explained in Section 2 of this paper, we have found critical flaws in the proof of the main theorem in Manzan and Zerom (2005), casting doubt on the asymptotic normality of their estimator.[1] Thus, as a special case of our result we obtain the asymptotic normality of *our* estimator when U is observed. It also reveals that a proof for the main theorem in Manzan and Zerom (2005) may not exist for reasons that are related to those that prevented (Martins-Filho and Yao, 2012) from giving an asymptotic characterization for their estimator.

Our proposed estimators are kernel based and relatively easy to implement since they do not require any numerical optimization or iterated procedures. As will be shown, they are consistent and asymptotically normally distributed, with the estimator for the finite dimensional parameter converging at the parametric $\sqrt{n}$ rate, and the nonparametric estimator converging at the expected nonparametric rate that is a function of the rate of decay of the bandwidth and the dimensionality of the underlying regressions.

Newey et al. (1999) proposed series estimators (power and splines) for a model where $\beta_0 = 0$ and the partially linear structure in (3) is generically modeled as $g_0(X, Z_1)$.[2] Otherwise, their model is identical to ours. Given that our partially linear structure is a restriction on g_0, their estimation method can be adapted to the model described by (1) and (2) (see Section 6 of their paper). In Section 3 of this paper, we contrast the additional assumptions they make to characterize some of the asymptotic behavior of their estimators with those we make to obtain similar results. Martins-Filho and Yao (2012) proposed kernel-based estimators for $(\beta_0, \beta')'$ and $m(\cdot)$, but although their estimators appear to have good finite-sample properties, they have failed to provide a characterization of their asymptotic behavior. In fact, our theoretical work suggests that their estimators cannot be shown to be asymptotically normally distributed under standard parametric and nonparametric normalizations (see details in Section 2).

Although the estimation procedure we consider is conceptually simple and fairly easy to implement, its asymptotic characterization is non-trivial, requiring repeated analysis of U-Statistics of high degree. This has been greatly facilitated by results in Yao and Martins-Filho (2015), which are used frequently in our proofs. The ancillary results required to obtain our theorems are, to our knowledge, novel and can be used in other contexts where generated regressors are encountered in various types of two stage kernel-based estimators.

1 Their results have been used in Manzan and Zerom (2010), Ozabaci and Henderson (2015) and featured in Li and Racine (2007) and Henderson and Parmeter (2015).

2 Ozabaci et al. (2014) also considered a model similar to that in Newey et al. (1999), but in their formulation $\Pi(Z)$, $g_0(X, Z_1)$, and $g(U)$ are all additive nonparametric functions of each of their arguments.

The rest of this paper is organized as follows. Section 2 describes the model in greater detail, considers identification and the moment conditions used in estimation, and provides a detailed algorithm for estimation. Section 3 gives asymptotic characterizations for our estimators and the assumptions we use to obtain our results. Where appropriate, we contrast our assumptions with those in Newey et al. (1999). Section 4 contains a small Monte Carlo study that sheds some light on the finite sample performance of our estimators and contrasts them to the series estimator proposed by Newey et al. (1999). Section 5 concludes. Supporting lemmas and the proofs of all theorems are given in Appendix A. An online appendix (OA) provides the proofs for the lemmas and details on the order of the U-Statistics appearing in the proofs of the theorems.

2. Moment conditions, identification and estimation

2.1. Moment conditions

We start by deriving a collection of conditional moments that emerge from Eqs. (2) and (3) and are the bases for the estimators we propose in Section 2.2. First, as usually assumed in the additive nonparametric literature (see, *inter alia*, Linton and Härdle, 1996; Kim et al., 1999; Manzan and Zerom, 2005; Martins-Filho and Yang, 2007), we put $\mathrm{E}\left(m(X_1, Z_1)\right) = \mathrm{E}\left(g(U)\right) = 0$, since each component in an additive nonparametric model can only be identified up to an additive constant.[3]

Using a suitable "instrument" function, we now obtain moment conditions that motivate our estimator for β_0 and β. For simplicity, in what follows, we put $W = (X_1', Z_1')'$. As in Kim et al. (1999), we define an "instrument" function $\eta = \eta(W, U) \equiv \frac{f_W(W) f_U(U)}{f_{WU}(W,U)}$, where f_W is the marginal density of W, f_U the marginal density of U and f_{WU} the joint density of W and U. Note that $\mathrm{E}\big(\eta|W\big) = 1$, $\mathrm{E}\big(\eta|U\big) = 1$, $\mathrm{E}\big(\eta g(U)|W\big) = 0$, and $\mathrm{E}\big(\eta m(W)|U\big) = 0$. By pre-multiplying both sides of Eq. (3) by η, and taking conditional expectations given W and U, respectively, we have

$$\mathrm{E}\left(\eta(Y - X_2'\beta - \beta_0) \mid W\right) = m(W), \qquad \mathrm{E}\left(\eta(Y - X_2'\beta - \beta_0) \mid U\right) \; = \; g(U). \tag{4}$$

It is apparent that if β_0 and β were known, and U were observed, $m(W)$ and $g(U)$ could be estimated based on the moment conditions (4) using an estimated $\hat{\eta}$ constructed with nonparametric density estimators of f_W, f_U, and f_{WU} evaluated at all data points. To address the fact that β_0 and β are unknown, note that $m(W)$ and $g(U)$ can be expressed as conditional expectations containing β and β_0 in (4). Substituting them back into (3) and rearranging, with $\beta_0 = \mathrm{E}(\eta(Y - X_2'\beta))$, we have

$$Y^* \; = \; X_2^{*\prime}\,\beta \; + \; v, \tag{5}$$

where $Y^* \equiv Y - \mathrm{E}(\eta Y|W) - \mathrm{E}(\eta Y|U) + \mathrm{E}(\eta Y)$, and $X_2^* \equiv X_2 - \mathrm{E}(\eta X_2|W) - \mathrm{E}(\eta X_2|U) + \mathrm{E}(\eta X_2)$.

Estimation of β could proceed by exploring the fact that $\mathrm{E}(X_2^* v) = 0$. In fact, the last equation provides a class of moments that can be used to estimate β. This follows since pre-multiplying by any arbitrary measurable function $L \equiv L(Z, U)$, we have $\mathrm{E}(L X_2^* v) = 0$, with instruments $L X_2^*$. It seems natural to put $L = 1$, as in Manzan and Zerom (2005), since X_2^* is already in the space spanned by all "instrumental variables" $\{Z, U\}$. However, to obtain the $\sqrt{n}$-consistency of the estimator for β, we find it essential to have the instruments orthogonal to each of the nonparametric regressors, specifically, $\mathrm{E}(L X_2^*|W) = 0$ and $\mathrm{E}(L X_2^*|U) = 0$. This requirement arises due to the additive structure of the nonparametric components and their separable estimation. As such, we set $L = \eta(W, U)$ since it satisfies $\mathrm{E}(L X_2^*|W) = 0$ and $\mathrm{E}(L X_2^*|U) = 0$.[4]

Failure to recognize the critical role of setting $L = \eta$ is at the heart of the problems we have encountered in the proof of asymptotic normality of the estimator of β proposed by Manzan and Zerom (2005).[5] Martins-Filho and Yao (2012) also failed to suggest, or understand, the role of L in obtaining asymptotic properties of the kernel-based estimators for this model. In fact, a more careful investigation of the consequences of choosing such a normalizing function in establishing the asymptotic properties of estimators of β_0, β, and m remains an open and important topic of study, as it also has a direct impact on the structure of the variances of their asymptotic distributions.

Hence, we consider the moment condition

$$\mathrm{E}\left(\eta X_2^*(Y^* - X_2^{*\prime}\,\beta)\right) \; = \; 0. \tag{6}$$

We denote the additive components in Y^*, X_2^*, and their corresponding error terms by $m_1(W) \equiv \mathrm{E}(\eta Y|W)$, $m_2(W) \equiv \mathrm{E}(\eta X_2|W)$, $m_3(W) \equiv \mathrm{E}(\eta|W)$, $g_1(U) \equiv \mathrm{E}(\eta Y|U)$, $g_2(U) \equiv \mathrm{E}(\eta X_2|U)$, $g_3(U) \equiv \mathrm{E}(\eta|U)$, $\mu_1 \equiv \mathrm{E}(\eta Y)$, $\mu_2 \equiv \mathrm{E}(\eta X_2)$,

[3] See Schick (1986). As in Robinson (1988), $\mathrm{E}\left(m(X_1, Z_1)\right) = 0$ can be relaxed by setting $\beta_0 = 0$.

[4] Robinson (1988) does not encounter this issue since there is only one nonparametric regression involved in the moment condition explored in estimation and the instruments $X - E(X|Z)$ (in his notation) are orthogonal to the nonparametric regressor Z. However, in other partially linear additive models, such as those for quantile regression, similar requirements may be needed. See, e.g., Cheng and Zerom (2015).

[5] Their difficulties occur primarily in the proof of their Lemma 2. First, their $\tilde{g}_j^k$ for $k \neq j$ does not equal to zero. Second, for $\mathrm{E}(S_{g_1 - \tilde{g}_1})$, we find it impossible to approximate it with the term following the $\sim$ sign on the second line of page 320. Third, the proof neglects the order of the error induced by using a nonparametric density estimator, which is dominating in terms of magnitude. In other words, the lemma suggests that the nonparametric estimators they use have a better than parametric convergence rate $o_p(n^{-1/2})$, which cannot be true.

$v_{m1} \equiv \eta Y - m_1(W)$, $v_{m2} \equiv \eta X_2 - m_2(W)$, $v_{m3} \equiv \eta - m_3(W)$, $v_{g1} \equiv \eta Y - g_1(U)$, $v_{g2} \equiv \eta X_2 - g_2(U)$, and $v_{g3} \equiv \eta - g_3(U)$. Given the moment condition associated with $m(W)$ in Eq. (4), we let $v_m \equiv \eta(Y - X_2'\beta - \beta_0) - m(W) = v_{m1} - v_{m2}'\beta - v_{m3}\beta_0$.

The moment condition (6) suggests an estimator of β obtained by inserting estimators of η, Y^*, and X_2^* prior to an application of a standard rule, such as the no-intercept ordinary least squares (OLS) method. Note that by (4), we have $m = m_1 - m_2'\beta - m_3\beta_0$, and $g = g_1 - g_2'\beta - g_3\beta_0$. Thus, to estimate Y^*, X_2^*, m, and g, we need only estimate each of their additive components separately. The main technical difficulty rests in the fact that U must be substituted by a generated regressor $\hat{U}$ in the estimation of all conditional moments involving U and $\eta(W,U)$. Kernel-based nonparametric regression estimators are employed throughout this paper, and for identification purposes, existence and non-singularity of $\Phi_0 \equiv \mathrm{E}\big(\eta X_2^* X_2^{*\prime}\big)$ are assumed.

2.2. Estimation

Consider a random sample $\{Y_i, X_i, Z_i\}_{i=1}^n$ where $(Y_i, X_i', Z_i')'$ is distributed as $(Y, X', Z')'$ for any $i = 1, \ldots, n$.

Based on the moment conditions given in Section 2.1, we now describe in detail our proposed estimation procedure. Since U_i is not observed, the first step in the estimation generates $\hat{U}_i$. We obtain a Nadaraya–Watson (NW) estimator for $\Pi(z)$ from (2), with the jth element defined as

$$\hat{\Pi}_j(z) = \underset{\theta}{\operatorname{argmin}} \frac{1}{nh_1^{D_1}} \sum_{t=1}^{n} (X_{t,j} - \theta)^2 K_1\left(\frac{Z_t - z}{h_1}\right) \text{ for } j = 1, \ldots, D_2,$$

where $X_{t,j}$ is the jth element of X_t, $h_1 > 0$ is the associated bandwidth, and $K_1 : \mathbb{R}^{D_1} \to \mathbb{R}$ is a multivariate kernel function. To associate the relevant subvector of $\Pi(Z_i)$ with X_{2i}, we define $\Pi(Z_i) \equiv (\Pi_1'(Z_i), \Pi_2'(Z_i))'$, where $\Pi_2(Z_i) \equiv (\Pi_{21}(Z_i), \ldots, \Pi_{2D_{22}}(Z_i))' = X_{2i} - U_{2i}$. $\Pi_1(Z_i)$ is defined similarly. Denote the estimates by $\hat{\Pi}(Z_i) = \big(\hat{\Pi}_1'(Z_i), \hat{\Pi}_2'(Z_i)\big)' \equiv \big(\hat{\Pi}_1(Z_i), \ldots, \hat{\Pi}_{D_2}(Z_i)\big)'$ and calculate the nonparametric residuals $\hat{U}_i \equiv \big(\hat{U}_{i1}, \ldots, \hat{U}_{iD_2}\big)'$, where $\hat{U}_{ij} \equiv X_{i,j} - \hat{\Pi}_j(Z_i)$, for $j = 1, \ldots, D_2$ and $i = 1, \ldots, n$.

In the second step, we estimate η_i (instrument functions) from Section 2.1 using W_t, and the generated regressors $\hat{U}_t$ obtained in the first step. We first obtain Rosenblatt–Parzen density estimators for f_U, f_W, and f_{WU}:

$$\hat{f}_{\hat{U}}(u) = \frac{1}{nh_2^{D_2}} \sum_{t=1}^{n} K_2\left(\frac{\hat{U}_t - u}{h_2}\right), \qquad \hat{f}_W(w) = \frac{1}{nh_3^{D_3}} \sum_{t=1}^{n} K_3\left(\frac{W_t - w}{h_3}\right),$$
$$\hat{f}_{W\hat{U}}(w,u) = \frac{1}{nh_4^{D_4}} \sum_{t=1}^{n} K_4\left(\frac{(W_t' \quad \hat{U}_t')' - (w' \quad u')'}{h_4}\right),$$

where $K_2 : \mathbb{R}^{D_2} \to \mathbb{R}$, $K_3 : \mathbb{R}^{D_3} \to \mathbb{R}$, and $K_4 : \mathbb{R}^{D_4} \to \mathbb{R}$ are multivariate kernel functions with $D_3 \equiv D_{11} + D_{21}$ and $D_4 \equiv D_2 + D_3$, and $h_i > 0$ is the associated bandwidth for $i = 2, 3, 4$. Similarly, denote the infeasible Rosenblatt–Parzen density estimators using the unobserved $\{U_t\}_{t=1}^n$ by $\hat{f}_U$ and $\hat{f}_{WU}$. Then, a natural estimator for η_i is $\hat{\eta}_i = \hat{\eta}(W_i, \hat{U}_i) \equiv \frac{\hat{f}_W(W_i)\hat{f}_{\hat{U}}(\hat{U}_i)}{\hat{f}_{W\hat{U}}(W_i, \hat{U}_i)}$.

In the third step we obtain the NW estimators for each conditional expectation in Y_i^* and X_{2i}^* as follows:

$$\begin{aligned} \hat{m}_1(w) &= \frac{1}{nh_3^{D_3}} \frac{1}{\hat{f}_W(w)} \sum_{t=1}^{n} K_3\left(\frac{W_t - w}{h_3}\right) \hat{\eta}_t Y_t, \qquad & \hat{m}_2(w) &= \frac{1}{nh_3^{D_3}} \frac{1}{\hat{f}_W(w)} \sum_{t=1}^{n} K_3\left(\frac{W_t - w}{h_3}\right) \hat{\eta}_t X_{2t}, \\ \hat{g}_1(u) &= \frac{1}{nh_2^{D_2}} \frac{1}{\hat{f}_{\hat{U}}(u)} \sum_{t=1}^{n} K_2\left(\frac{\hat{U}_t - u}{h_2}\right) \hat{\eta}_t Y_t, & \hat{g}_2(u) &= \frac{1}{nh_2^{D_2}} \frac{1}{\hat{f}_{\hat{U}}(u)} \sum_{t=1}^{n} K_2\left(\frac{\hat{U}_t - u}{h_2}\right) \hat{\eta}_t X_{2t}. \end{aligned} \tag{7}$$

Estimators of the unconditional expectations μ_1 and μ_2 are given by $\hat{\mu}_1 = \frac{1}{n}\sum_{t=1}^n \hat{\eta}_t Y_t$, and $\hat{\mu}_2 = \frac{1}{n}\sum_{t=1}^n \hat{\eta}_t X_{2t}$. Thus, we define estimators of Y_i^* and X_{2i}^* respectively as $\hat{Y}_i = Y_i - \hat{m}_1(W_i) - \hat{g}_1(\hat{U}_i) + \hat{\mu}_1$, $\hat{X}_{2i} = X_{2i} - \hat{m}_2(W_i) - \hat{g}_2(\hat{U}_i) + \hat{\mu}_2$, for $i = 1, \ldots, n$.

In the fourth step, using the estimators $\hat{\eta}_i$, $\hat{Y}_i$, and $\hat{X}_{2i}$ derived in the previous steps, instead of η_i, Y_i^*, and X_{2i}^* in (6), we have a feasible no-intercept OLS estimator of β,

$$\hat{\beta} = \left(\hat{\boldsymbol{X}}_2'\hat{\boldsymbol{\eta}}\hat{\boldsymbol{X}}_2\right)^{-1} \hat{\boldsymbol{X}}_2'\hat{\boldsymbol{\eta}}\hat{\boldsymbol{Y}}, \tag{8}$$

where $\hat{\boldsymbol{Y}} = (\hat{Y}_1, \ldots, \hat{Y}_n)'$, $\hat{\boldsymbol{X}}_2 = (\hat{X}_{21}, \ldots, \hat{X}_{2n})'$, and $\hat{\boldsymbol{\eta}} = \operatorname{diag}\{\hat{\eta}_i\}_{i=1}^n$. Given that $\beta_0 = \mathrm{E}(Y - X_2'\beta)$ and $\hat{\beta}$, an estimator of β_0 is $\hat{\beta}_0 = \bar{Y} - \bar{X}_2'\hat{\beta}$, where $\bar{Y} \equiv \frac{1}{n}\sum_{t=1}^n Y_t$, and $\bar{X}_2 \equiv \frac{1}{n}\sum_{t=1}^n X_{2t}$.

Finally, the last step provides an estimator for m. Given Eq. (4) and the estimators $\hat{\beta}_0$ and $\hat{\beta}$, we propose the following estimators for $m(W_i)$ and $g(U_i)$,

$$\hat{m}(w) = \hat{m}_1(w) - \hat{m}_2'(w)\hat{\beta} - \hat{m}_3(w)\hat{\beta}_0, \qquad \hat{g}(u) = \hat{g}_1(u) - \hat{g}_2'(u)\hat{\beta} - \hat{g}_3(u)\hat{\beta}_0, \tag{9}$$

where $\hat{m}_3(w)$ and $\hat{g}_3(u)$ are NW estimators for $m_3(w)$ and $g_3(u)$ defined similarly as $\hat{m}_1(w)$ and $\hat{g}_1(u)$ in (7) except that $\hat{\eta}_t$ is used, instead of $\hat{\eta}_t Y_t$, as regressand.

Our choice of NW estimators in the first and third steps of our estimation procedure was guided by the following considerations. First, NW estimators are widely used and easy to implement. Second, they have been the estimator of choice for the infinite dimensional parameter in a number of papers that study partially linear models (see, *inter alia*, Robinson, 1988; Fan and Li, 1999; Li and Wooldridge, 2002; Manzan and Zerom, 2005; and Juhl and Xiao, 2005). Third, whereas other nonparametric kernel-based regression estimators could be used, such as local-polynomial estimators (see Linton, 1995 and Su and Ullah, 2008), their asymptotic properties do not impact our main result in Theorem 3. Fourth, whereas the choice of NW estimators does not impact Theorem 3, it does impact the structure of the bias of the estimator $\hat{m}(\cdot)$ in Theorem 4. In particular, the design adaptability of the bias of local polynomial estimators (Fan, 1992) is not a property of NW estimators. However, given that in partially linear models $m(\cdot)$ is largely taken to be a nuisance parameter and given the theoretical convenience of dealing with NW estimators, we elect to use NW estimators throughout our estimation procedure.

3. Asymptotic characterizations of $\hat{\beta}$ and $\hat{m}(\cdot)$

In this section, we study the asymptotic properties of the estimators $\hat{\beta}$ and $\hat{m}(\cdot)$ defined in the previous section. We first establish the uniform convergence in probability rate of the Rosenblatt density estimator using estimated residuals $\{\hat{U}_i\}_{i=1}^n$. Second, we give the uniform convergence in probability rate of the NW estimator constructed using estimated residuals $\{\hat{U}_i\}_{i=1}^n$. Third, we establish $\sqrt{n}$ asymptotic normality of $\hat{\beta}-\beta$. Lastly, we use the asymptotic normality of $\sqrt{n}(\hat{\beta}-\beta)$ to establish the asymptotic distribution of $\hat{m}(\cdot)$ under suitable centering and normalization.

3.1. Assumptions

First we provide a list of general assumptions that will be adopted in our theorems and introduce notation. In what follows, C denotes a generic constant in $(0,\infty)$ that may vary from case to case. $k^{(j)}(x)$ denotes the jth-order derivative of $k(x)$ evaluated at x.

Assumption A1. The kernels K_i, for $i = 1, 2, 3, 4$, satisfy $K_i(x) = \prod_{j=1}^{D_i} k_i(x_j)$, where D_i is the corresponding dimension of K_i. k_i is symmetric about zero, 4-times continuously differentiable, and satisfies: (a) $\int k_i(x)\,dx = 1$; (b) $|k_i^{(j)}(x)||x|^{7+a} \to 0$ as $|x| \to \infty$, $j = 0, \ldots, 4$, for some $a > 0$; (c) k_i is a kernel of order s_i, i.e., $\int k_i(x)x^j\,dx = 0$ for $j = 1, \ldots, s_i - 1$, and $\int |k_i(x)||x|^{s_i}\,dx < C$. We let $s \equiv \max\{s_i\}_{i=1}^4$ and $\mu_{k_i,s_i} \equiv \int k_i(x)x^{s_i}\,dx$.

Our use of "higher-order" kernels is needed to attain suitable orders for the biases of our nonparametric estimators. Since global differentiability of the kernel functions is required in using Taylor's Theorem, in the following theorems, we will not consider kernels with compact support for theoretical purposes.[6] It is easy to construct kernels that satisfy the conditions in A1. For example, kernels of even order $s \geq 2$, can be defined as

$$k_s(x) = \sum_{j=0}^{\frac{1}{2}(s-2)} c_j x^{2j}\phi(x), \tag{10}$$

where $\phi(x) = (2\pi)^{-1/2}\exp(-\frac{1}{2}x^2)$ for suitably chosen c_j. In particular, given that we can evaluate the moments $m_{2j} = \int x^{2j}\phi(x)\,dx$, $0 \leq j \leq \frac{1}{2}(s-2)$, we choose $\{c_j\}_{j=0}^{\frac{1}{2}(s-2)}$ that satisfy the linear system of $s/2$ simultaneous equations $\sum_{j=0}^{\frac{1}{2}(s-2)} c_j m_{2(i+j)} = \delta_{i0}$, $0 \leq i \leq \frac{1}{2}(s-2)$, where δ_{i0} is Kronecker's delta. For example, $k_2(x) = \phi(x)$, $k_4(x) = \left(\frac{3}{2} - \frac{1}{2}x^2\right)\phi(x)$, and $k_6(x) = \left(\frac{15}{8} - \frac{5}{4}x^2 + \frac{1}{8}x^4\right)\phi(x)$. Note that these kernels are continuously differentiable of any order everywhere, and when multiplied by any polynomial function they are all uniformly bounded and absolutely integrable, as their tails decay exponentially. We show in Lemma 1 that product kernels satisfying A1 are locally Lipschitz continuous, which is necessary for Lemma 3.

Assumption A2. The components of the sequence $\{(Y_i, X_i', Z_i')\}_{i=1}^n$ of random vectors described in (2) and (3) are independent and identically distributed (IID). The density functions $f_W(w)$, $f_Z(z)$, $f_{WU}(w,u)$, $f_{UZ}(u,z)$, and $f_U(u)$ are uniformly bounded away from zero and infinity on arbitrary convex compact subsets of their domains. Here, $f_{UZ}(\cdot)$ is the joint density function of (U,Z).

The existence, boundedness properties and compactness of the support of the densities in Assumption A2 are common regularity conditions imposed to derive properties of kernel-based nonparametric estimators and largely overlap with Assumption 2 in Newey et al. (1999).

[6] This requirement is not essential, in practice, if we only consider estimated $\hat{U}$ taking values on a compact set.

Assumption A3. (i) $\mathrm{E}(m(W)) = \mathrm{E}(g(U)) = 0$, (ii) $\mathrm{E}(v^2|Z,U) = \sigma_v^2 < \infty$, $\mathrm{E}(U_j^2|Z) = \sigma_{U_j}^2 < \infty$, $\mathrm{E}(v_{m1}^2|W) = \sigma_{vm1}^2 < \infty$, $\mathrm{E}(v_{m2,j}^2|W) = \sigma_{vm2}^2 < \infty$, $\mathrm{E}(v_{g1}^2|U) = \sigma_{vg1}^2 < \infty$, $\mathrm{E}(v_{g2,j}^2|U) = \sigma_{vg2}^2 < \infty$, and (iii) the following Cramer's conditions: $\mathrm{E}|X_{2,j}|^p \le C^{p-2}p!\mathrm{E}|X_{2,j}|^2 < \infty$, $\mathrm{E}(|U_j|^p|Z) \le C^{p-2}p!\sigma_{U_j}^2$, for some $C > 0$, $p = 3, 4, \ldots$, and $j = 1, \ldots, D_2$.

A3(i) is assumed without loss of generality and is used in identification of the additive structure in Eq. (3). In A3(ii), it is not essential to assume that the second conditional moment of the error terms are independent of the conditioning variables; however, the boundedness of the second moment is crucial here, as in Assumptions 1 and 5 in Newey et al. (1999). Cramer's conditions in A3(iii) are imposed due to the use, in Lemma 2, of Bernstein's Inequality to establish the uniform order in probability of some specific averages. In particular, Lemma 2 is critical in handling the fact that U is estimated by $\hat{U}$, which is used in defining $\hat{f}_{\hat{U}}$, $\hat{f}_{W\hat{U}}$, and $\hat{\eta}$. If U were observed, Cramer's conditions could be relaxed.

Assumption A4. Let C^q denote the class of functions such that each of its elements: (i) is q-times partially continuously differentiable, and (ii) all their partial derivatives up to order q are uniformly bounded. For $d = 1, \ldots, D_2$, and $\ell = 1, 2, 3$, $\Pi_d(\cdot), f_{WU}(\cdot), f_{UZ}(\cdot), m(\cdot), g(\cdot), m_\ell(\cdot), g_\ell(\cdot) \in C^{s+1}$, where s is defined in Assumption A1.

Assumption A4 assumes smoothness of the regression functions and uniform bounds of their partial derivatives. This assumption, together with kernels of suitable order, as required in A1, gives desired orders for the biases. We note that in our Assumption A1 $s \equiv \max\{s_i\}_{i=1}^4$, and for convenience A4 requires all functions to be in C^{s+1}. This is sufficient for our theorems, but not necessary, expressing only the highest degree of smoothness needed. Depending on the context lower degrees of smoothness can be assumed.[7]

Assumption A5. Denote $L_{in} \equiv \left(\frac{\log n}{nh_i^{D_i}}\right)^{\frac{1}{2}} + h_i^{s_i}$, for $i = 1, \ldots, 4$, and $L_n = \sum_{i=2}^4 L_{in}$, where $h_i \to 0$ as $n \to \infty$ and satisfies:

(i) $h_1 = n^{-\delta}$, with $\frac{1}{2s_1} < \delta < \min_{\{i=2,4\}} \frac{D_i}{D_1(2s_i+D_i)}$;
(ii) for $i = 2, 4$, $h_i = n^{-\frac{1}{2s_i+D_i}}$, with $s_i \ge D_i/2 + 2$;
(iii) $h_3 = n^{-\frac{1}{2s_3+D_3}}$, with $\frac{1}{2} < \frac{s_3}{D_3} < \min_{\{i=2,4\}} \frac{s_i}{D_i}$.

Assumption A5 provides the order of all the bandwidths. The fact that using residual estimates $\{\hat{U}_i\}_{i=1}^n$, instead of $\{U_i\}_{i=1}^n$, has no impact on the first-order asymptotic properties of our estimator relies on undersmoothing in the first stage when regressing X on Z nonparametrically, and on $\Pi(z)$ being sufficiently smooth. For h_2, h_3, and h_4, the orders are chosen optimally by minimizing the mean squared error of traditional NW kernel estimators. The second inequality in A5(iii) implies that $L_{in}/L_{3n} \to 0$ for $i = 2, 4$ to ensure that using estimated densities for $f_U(\cdot)$ and $f_{WU}(\cdot)$ does not result in any asymptotic consequences in deriving the distribution of $\hat{m}(\cdot)$.

3.2. Theorems

By Theorem 2.6 in Li and Racine (2007), under A1–A5, for a compact subset $\mathcal{G}_Z \subset \mathbb{R}^{D_1}$, we have

$$\sup_{z\in\mathcal{G}_Z} \left|\hat{\Pi}(z) - \Pi(z)\right| = O_p(L_{1n}) \tag{11}$$

where $L_{1n} = \left(\frac{\log n}{nh_1^{D_1}}\right)^{1/2} + h_1^{s_1}$. This uniform convergence rate in probability of the NW estimator is used throughout this paper. Note that $\hat{f}_{\hat{U}}(\hat{U}_i)$ and $\hat{f}_{W\hat{U}}(W_i, \hat{U}_i)$ are used to approximate $f_U(U_i)$ and $f_{WU}(W_i, U_i)$ in η_i. By Theorem 1, we can show that the uniform convergence rate of $\hat{f}_{\hat{U}}(\hat{U}_i)$ to $f_U(U_i)$ using $\{\hat{U}_i\}_{i=1}^n$ is no different from that of the traditional Rosenblatt density estimator based on the unobserved $\{U_i\}_{i=1}^n$. A similar result holds for $\hat{f}_{W\hat{U}}(W_i, \hat{U}_i)$.

Theorem 1. *Under A1–A5, for arbitrary convex and compact subsets $\mathcal{G}_Z \subset \mathbb{R}^{D_1}$, $\mathcal{G}_U \subset \mathbb{R}^{D_2}$ and $\mathcal{G}_M \subset \mathbb{R}^{D_3}$, we have*

$$\begin{aligned} \sup_{u\in\mathcal{G}_U} \left|\hat{f}_{\hat{U}}(u) - f_U(u)\right| &= O_p(L_{2n}), \qquad \sup_{w\in\mathcal{G}_W} \left|\hat{f}_W(w) - f_W(w)\right| = O_p(L_{3n}), \\ \sup_{\{w,u\}\in\mathcal{G}_W\times\mathcal{G}_U} \left|\hat{f}_{W\hat{U}}(w,u) - f_{WU}(w,u)\right| &= O_p(L_{4n}), \end{aligned} \tag{12}$$

where $\mathcal{G}_W \times \mathcal{G}_U$ denotes the Cartesian product of sets $\mathcal{G}_W$ and $\mathcal{G}_U$, $L_{in} = \left(\frac{\log n}{nh_i^{D_i}}\right)^{1/2} + h_i^{s_i}$, for $i = 2, 3, 4$.

[7] For example, in Section 4, where specific data generating processes (DGP) are considered, it suffices to have $\Pi(\cdot) \in C^6$, $f_{WU}(\cdot) \in C^4$, $f_{UZ}(\cdot) \in C^5$, $m(\cdot) \in C^2$, and $g(\cdot) \in C^4$.

Note that in Theorem 1 we establish the uniform convergence rate of $\hat{f}_{\hat{U}}(u)$ and $\hat{f}_{W\hat{U}}(w,u)$ over $\mathscr{G}_U$ and $\mathscr{G}_W \times \mathscr{G}_U$, respectively. This is due to the fact that $\hat{U}_i$ is an estimated residual given by $\hat{U}_i = X_i - \hat{\Pi}(Z_i)$ and the uniform convergence rate of $\hat{\Pi}(Z_i)$ given in (11) is taken over a compact set $\mathscr{G}_Z$. Consequently, this implies that $|\hat{f}_{\hat{U}}(\hat{U}_i) - f_U(U_i)| = O_p(L_{2n})$ uniformly for any $\hat{U}_i, U_i \in \mathscr{G}_U$ as $|\hat{f}_{\hat{U}}(\hat{U}_i) - f_U(U_i)| \le |\hat{f}_{\hat{U}}(\hat{U}_i) - f_U(\hat{U}_i)| + |f_U(\hat{U}_i) - f_U(U_i)| \le O_p(L_{2n}) + C\|\hat{U}_i - U_i\|_E = O_p(L_{2n} + L_{1n}) = O_p(L_{2n})$. Similarly, we have $|\hat{f}_{W\hat{U}}(W_i, \hat{U}_i) - f_{WU}(W_i, U_i)| = O_p(L_{4n})$ uniformly. These results and A2 together imply that $\hat{\eta}_i - \eta_i = O_p(L_n)$ uniformly, where $L_n \equiv \sum_{i=2}^{4} L_{in}$, and consequently we have $\hat{\mu}_k - \mu_k = O_p(L_n)$ for $k = 1, 2$. With this result, we are ready to provide the uniform convergence rate of the estimators given in (7).

Theorem 2. *Under A1–A5, for arbitrary convex and compact subsets $\mathscr{G}_U$ and $\mathscr{G}_W$, for $k = 1, 2, 3$, we have,*

$$\sup_{u\in\mathscr{G}_U}\left|\hat{g}_k(u) - g_k(u)\right| = O_p\left(L_n + \frac{L_{1n}}{h_2}\right), \quad \sup_{w\in\mathscr{G}_W}\left|\hat{m}_k(w) - m_k(w)\right| = O_p\left(L_n\right). \tag{13}$$

The rates of uniform convergence in probability of $\hat{g}_k(\cdot)$ to $g_k(\cdot)$ and $\hat{m}_k(\cdot)$ to $m_k(\cdot)$, and by consequence, those of $\hat{g}(\cdot)$ to $g(\cdot)$ and $\hat{m}(\cdot)$ to $m(\cdot)$ depend fundamentally on the degree of smoothness of the functions appearing in A4 and the dimensions of the vectors X_i and Z_i. Given D_i for $i = 1, \ldots, 4$ and Assumption A5, it is possible to obtain the necessary smoothness in A4 that assures the results in Theorem 2. Furthermore, the given rate of convergence can be calculated as a function of n. Therefore, the same uniform convergence rate, $O_p(L_n + \frac{L_{1n}}{h_2})$, of the feasible estimator $\hat{g}_k(\hat{U}_i)$ to the true value $g_k(U_i)$ for any $\hat{U}_i, U_i \in \mathscr{G}_U$ follows immediately by Mean Value Theorem as $g(\cdot) \in C^1$ and $\|\hat{U}_i - U_i\|_E = O_p(L_{1n})$. Similarly, given Assumptions 3 and 8 in Newey et al. (1999), the rate of convergence in their Theorem 4.3 can also be calculated.[8] An important difference between our results and theirs is that, in our case, the rate is obtained taking into account the randomness of $\hat{U}_i$ and the estimation of $g(\cdot)$ ($\lambda(\cdot)$ in their notation), whereas they take $U = \bar{u}$ as fixed and the true $g(\cdot)$ to be known.

Note that the first term in the order of $|\hat{g}_k(u) - g_k(u)|$ is not new, as it is just a sum of uniform orders for different NW estimators. The h_2 in the denominator of the second term comes from a Taylor expansion of the kernel evaluated at the estimated residuals $\{\hat{U}_i\}_{i=1}^n$. With well chosen bandwidths in A5, it is essential to have that L_n^2, $\left(\frac{L_{1n}}{h_2}\right)^2 = o(n^{-1/2})$. This result will help establish the asymptotic distribution of $\hat{\beta}$:

$$\sqrt{n}(\hat{\beta} - \beta) = \left(\frac{1}{n}\hat{\boldsymbol{X}}_2'\hat{\boldsymbol{\eta}}\hat{\boldsymbol{X}}_2\right)^{-1}\frac{1}{\sqrt{n}}\hat{\boldsymbol{X}}_2'\hat{\boldsymbol{\eta}}(\hat{\boldsymbol{Y}} - \hat{\boldsymbol{X}}_2\beta). \tag{14}$$

As we can see in (14), there are two components that need to be studied to establish the asymptotic properties of $\sqrt{n}(\hat{\beta} - \beta)$. We need to (i) establish the asymptotic behavior of the matrix $\frac{1}{n}\hat{\boldsymbol{X}}_2'\hat{\boldsymbol{\eta}}\hat{\boldsymbol{X}}_2$, and (ii) establish the asymptotic normality of the term $\frac{1}{\sqrt{n}}\hat{\boldsymbol{X}}_2'\hat{\boldsymbol{\eta}}(\hat{\boldsymbol{Y}} - \hat{\boldsymbol{X}}_2\beta)$. Uniform orders of NW estimators derived in Theorem 2 will help take care of (i). However, to establish asymptotic normality for the second term, we need to investigate the behavior of U-Statistics up to degree 3. Yao and Martins-Filho (2015) provide a direct and convenient method to characterize the asymptotic magnitude of each component in the H-decomposition (see Hoeffding, 1948) of a U-Statistic, and many places in our proofs are based on their results. The next theorem establishes the asymptotic distribution of $\hat{\beta}$ after suitable centering and under $\sqrt{n}$-normalization.

Theorem 3. *Under A1–A5, assuming that matrix Φ_0 exists and is nonsingular, we have*

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} \mathscr{N}\left(0, \Phi_0^{-1}(\Phi_1 + \Phi_2)\Phi_0^{-1}\right), \tag{15}$$

where the matrices Φ_0, Φ_1, and Φ_2 have typical elements given by

$$\begin{aligned}
\Phi_{0_{(j,k)}} &= E\left[\eta_t\left(X_{2t,j} - m_{2j}(W_t) - g_{2j}(U_t) + \mu_{2j}\right)\left(X_{2t,k} - m_{2k}(W_t) - g_{2k}(U_t) + \mu_{2k}\right)\right];\\
\Phi_{1_{(j,k)}} &= E\left[\eta_t^2\left(X_{2t,j} - m_{2j}(W_t) - g_{2j}(U_t) + \mu_{2j}\right)\left(X_{2t,k} - m_{2k}(W_t) - g_{2k}(U_t) + \mu_{2k}\right)\right]\sigma_v^2;\\
\Phi_{2_{(j,k)}} &= E\left[\sum_{d=1}^{D_2}\sum_{\delta=1}^{D_2} E\Big(\left(\Pi_{2j}(Z_i) - \Pi_{2j}(Z_t)\right)\mathrm{D}_d g(U_t)\eta_t\Big|Z_i\Big)\right.\\
&\qquad \left.\times E\Big(\left(\Pi_{2k}(Z_i) - \Pi_{2k}(Z_t)\right)\mathrm{D}_\delta g(U_t)\eta_t\Big|Z_i\Big)U_{id}U_{i\delta}\right], \quad \text{for } j, k = 1, \ldots, D_{22}.
\end{aligned}$$

[8] As for our estimator, the exact rate of convergence of their spline estimators, as a function of n, depends on the degree of continuous differentiability of $m(\cdot)$, $g(\cdot)$, and $\Pi(\cdot)$, the dimension of Z, W, and the rate at which the number of knots (K and L in their notation) diverges to infinity. When the sequences of knots have the same order and for the DGPs used in our Section 4, we have calculated the exact rate of convergence for both our estimator ($n^{-1/3}$) and theirs ($n^{-5/14}$).

Remarks. 1. It follows directly from Theorem 3 that $\hat{\beta}$ is consistent and asymptotically unbiased. The explicit structure for the covariance of the limiting distribution allows for asymptotically valid inference and hypothesis testing when a consistent estimator for the covariance is available. Given the structure of its component covariance matrices, we provide consistent estimators for Φ_i, $i = 1, 2, 3$ as follows,

$$\hat{\Phi}_0 = \frac{1}{n}\hat{\boldsymbol{X}}_2'\hat{\boldsymbol{\eta}}\hat{\boldsymbol{X}}_2 \qquad \hat{\Phi}_1 = \frac{1}{n}\hat{\boldsymbol{X}}_2'\hat{\eta}\hat{\eta}\hat{\boldsymbol{X}}_2\hat{\sigma}_v^2, \qquad \hat{\Phi}_2 = \frac{1}{n}\boldsymbol{Q}'\boldsymbol{Q}, \tag{16}$$

where $\hat{\sigma}_v^2 \equiv \frac{1}{n}\hat{\boldsymbol{v}}'\hat{\boldsymbol{v}}$, $\hat{\boldsymbol{v}} \equiv \boldsymbol{Y} - \boldsymbol{X}_2\hat{\beta} - \hat{\beta}_0 - \hat{\boldsymbol{m}} - \hat{\boldsymbol{g}}$, $\hat{\boldsymbol{m}} \equiv (\hat{m}(W_1), \ldots, \hat{m}(W_n))'$, $\hat{\boldsymbol{g}} \equiv (\hat{g}(\hat{U}_1), \ldots, \hat{g}(\hat{U}_n))'$, $\boldsymbol{Q} \equiv (Q_1, \ldots, Q_n)'$, $Q_i \equiv \frac{1}{n}(\boldsymbol{1}_n\hat{\Pi}_2'(Z_i) - \hat{\boldsymbol{\Pi}}_2)'\hat{\boldsymbol{\eta}}\mathrm{D}\hat{\boldsymbol{g}}\hat{U}_i$, $\hat{\Pi}_2(Z_i) \equiv (\hat{\Pi}_{21}(Z_i), \ldots, \hat{\Pi}_{2D_{22}}(Z_i))'$, $\hat{\boldsymbol{\Pi}}_2 \equiv (\hat{\Pi}_2(Z_1), \ldots, \hat{\Pi}_2(Z_n))'$, $\boldsymbol{1}_n \equiv (1, \ldots, 1)'_{n\times 1}$, $\mathrm{D}\hat{\boldsymbol{g}} \equiv (\mathrm{D}_1\hat{\boldsymbol{g}}, \ldots, \mathrm{D}_{D_2}\hat{\boldsymbol{g}})$, $\mathrm{D}_d\hat{\boldsymbol{g}} \equiv (\mathrm{D}_d\hat{g}(\hat{U}_1), \ldots, \mathrm{D}_d\hat{g}(\hat{U}_n))'$, and $\mathrm{D}_d\hat{g}(\hat{U}_i)$ is the partial derivative of the estimator $\hat{g}(u)$ with respect to the dth element of u evaluated at $\hat{U}_i$. Given Eqs. (7) and (9), by taking partial derivatives, we have $\mathrm{D}_d\hat{g}(\hat{U}_i)$ given by

$$\mathrm{D}_d\hat{g}(u) = -\frac{1}{nh_2^{D_2+1}}\frac{1}{\hat{f}_{\hat{U}}(u)}\sum_{t=1}^{n}\mathrm{D}_dK_2\left(\frac{\hat{U}_t - u}{h_2}\right)\left[\hat{\eta}_t(Y_t - X_{2t}\hat{\beta}) - (\hat{g}_1(u) - \hat{g}_2'(u)\hat{\beta})\right].$$

2. The covariance $\Phi_0^{-1}(\Phi_1+\Phi_2)\Phi_0^{-1}$ differs from what one would obtain if U were observed. Hence, there is an asymptotic cost in using $\hat{U}$ in estimation. It manifests itself via the presence of Φ_2, which would be zero if U were observed. In this case, Theorem 3 gives $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}\left(0, \Phi_0^{-1}\Phi_1\Phi_0^{-1}\right)$ which is similar to the limiting distribution claimed by Manzan and Zerom (2005) for their estimator, under homoscedasticity of the regression error. The critical difference is that the expectations that characterize the elements of Φ_0 and Φ_1 for our estimator include η_t and η_t^2, respectively. This follows from the fact that the moment conditions we explore to construct our estimator must include η (see Eq. (6)). As we have pointed out in Section 2.1, when, as in Manzan and Zerom (2005), $\eta = 1$ the asymptotic normality and semiparametric efficiency of the proposed estimator remains unproven.

3. It is evident that the choice of $L = \eta$, which led to the moment conditions in Eq. (6) that we explored in our estimation procedure, has a direct impact on the variance of the limiting distribution of $\sqrt{n}(\hat{\beta} - \beta)$. In fact, other choices of L that satisfies $E(LX_2^*|W) = E(LX_2^*|U) = 0$ and can be estimated at a suitable rate, e.g., $O_p(L_n)$, would produce a class of estimators for β that are asymptotically normal with covariance depending on L. We have not investigated whether it is possible to obtain a minimum variance estimator in this class, leaving this important topic for future research. However, it seems apparent that estimators motivated by moment conditions that do not satisfy $E(LX_2^*|W) = E(LX_2^*|U) = 0$ are not in this class.

4. The covariance matrix of the limiting distribution does not meet the semiparametric efficiency bound of Chamberlain (1992), a characteristic that our estimator shares with that proposed in Li and Wooldridge (2002).[9] This is also the case for the class of estimators proposed in Mammen et al. (2016). In fact, since our estimation procedure explores the additive structure of the nonparametric component, we expect based on sufficiency, that the variance of its asymptotic distribution will be smaller than that derived by Mammen et al. (2016) for their estimator. However, due to the complexity of the expressions for these variances we have been unable to theoretically establish their relative magnitudes.

5. Given Theorems 2, 3 and Eq. (9), we have the uniform convergence rate of $\hat{g}(\cdot)$ at $O_p(L_n + L_{1n}/h_2)$, which is generally worse than that of the traditional NW estimator due to the presence of h_2 in the second term.

The following theorem gives asymptotic normality of $\hat{m}(\cdot)$ at the typical nonparametric rate, in our case, $\sqrt{nh_3^{D_3}}$.

Theorem 4. *Let $\mathrm{D}_j^q f(x) \equiv \frac{\partial^q}{\partial_j^q}f(x)$ and $\mathrm{D}_j^0 f(x) \equiv f(x)$, $\forall q \geq 1$, $1 \leq j \leq q$. Under A1–A5 and if we assume that $\mathrm{E}\left(v_m^2|W\right) = \sigma_{vm}^2 < \infty$, $\mathrm{E}\left(|v_m|^{2+\delta}|W\right) \leq C < \infty$ for some $\delta > 0$, $|k_3(x)||x|^{2s_3+1+a} \to 0$ as $|x| \to \infty$ for some $a > 0$, and $|k_3^{(1)}(x)||x|^{s_3} \leq C$ for all x and some $C < \infty$, we have*

$$\sqrt{nh_3^{D_3}}\left(\hat{m}(w) - m(w) - b_m(w)\right) \xrightarrow{d} \mathcal{N}\left(0, \Phi_3 + \Phi_4\right),$$

where

$$b_m(w) = h_3^{s_3}\frac{\mu_{k_3,s_3}}{f_W(w)}\sum_{k=0}^{s_3}\frac{1}{k!(s_3-k)!}\sum_{j=1}^{D_3}\mathrm{D}_j^k m(w)\mathrm{D}_j^{s_3-k}f_W(w) + o_p(h_3^{s_3}),$$

$$\Phi_3 = \frac{\sigma_{vm}^2}{f_W(w)}\int K_3^2(\gamma)\,\mathrm{d}\gamma, \quad \Phi_4 = \frac{m^2(w)}{f_W(w)}\int\left(\int K_3(\gamma_1)K_3(\gamma_1+\gamma_2)\,\mathrm{d}\gamma_1\right)^2\,\mathrm{d}\gamma_2.$$

[9] See Li (2000) for estimators that satisfy a semiparametric efficiency bound when all regressors are observed, i.e., in the absence of generated regressors.

Remarks. 1. Given the order and structure of the bias, it follows immediately from Theorem 4 that $\hat{m}(w) - m(w) = o_p(1)$.

2. The fact that η, β_0, and β have to be estimated is costly asymptotically. In particular, the variance of the limiting distribution contains the strictly positive term Φ_4 added to Φ_3. Φ_3 can be immediately recognized as the covariance of the limiting distribution of an "oracle" Nadaraya–Watson estimator constructed under the assumption that η, β_0, and β are known. Hence, $\hat{m}(\cdot)$ is not oracle efficient. It may be possible to eliminate Φ_4 by considering a new estimator that explores a one-step backfitting procedure using $\hat{g}(\cdot)$, as suggested by Kim et al. (1999) and especially Yu et al. (2011). We leave this potential improvement for future research.

4. Monte Carlo study

In this section, we provide some experimental evidence on the finite sample behavior of our estimators $(\hat{\beta}, \hat{m})$ and contrast it to that of some alternative estimation procedures. We consider the following data generating processes (DGPs):

$$\text{DGP}_1: \quad Y_i = \ln(|X_{1i} - 1| + 1)\,\text{sgn}(X_{1i} - 1) + X_{2i}'\beta + \beta_0 + \varepsilon_i,$$

$$\text{DGP}_2: \quad Y_i = \frac{\exp(X_{1i})}{1 + 3\exp(X_{1i})} + X_{2i}'\beta + \beta_0 + \varepsilon_i,$$

for $i = 1, \ldots, n$. The sample size n is set at 100, 200, and 400. In both DGPs, Z_{1i} and Z_{2i} are generated independently from $N(0, 1)$ and we construct $X_{1i} = Z_{1i} + Z_{2i} + U_{1i}$ and $X_{2i} = Z_{1i}^2 + Z_{2i}^2 + U_{2i}$. ε_i and $U_i = (U_{1i}, U_{2i})'$ are generated as

$$\begin{pmatrix} \varepsilon_i \\ U_i \end{pmatrix} \sim NID\left(0, \begin{pmatrix} 1 & \theta & \theta \\ \theta & 1 & \theta^2 \\ \theta & \theta^2 & 1 \end{pmatrix}\right),$$

where the values $\theta = 0.3,\ 0.6$, and 0.9 indicate weak, moderate, and strong endogeneity, respectively. It is easy to verify that $\mathrm{E}(\varepsilon_i|Z_i) = 0$, $\mathrm{E}(U_i|Z_i) = 0$, and thus $\mathrm{E}(\varepsilon_i|U_i, Z_i) = \mathrm{E}(\varepsilon_i|U_i) = \frac{\theta}{1+\theta^2}(U_{1i} + U_{2i})$. We set the parameters $\beta = 1$, $\beta_0 = 1$, and perform 1000 repetitions for each experiment design.

The implementation of our estimators requires a choice of kernel function $K_i(\cdot)$ for $i = 1, \ldots, 4$ and bandwidth sequences. For all kernels, we use products of a univariate Gaussian kernel of appropriate orders, as we discussed in Assumption A1. For both DPGs we have $D_1 = D_2 = 2$, $D_3 = 1$ and $D_4 = 3$. Setting $s_1 = 5$, $s_2 = 3$, $s_3 = 1$, $s_4 = 4$, we choose bandwidths orders in accordance to A5. For the estimation of densities we choose rule-of-thumb (ROT) bandwidths as proposed in Simonoff (1996, p. 105) giving $h_i = c_i\hat{\sigma}(M_i)n^{-1/(2s_i+D_i)}$ for $i = 2, 3, 4$ with $c_2 = 1$, $c_3 = 1.06$, $c_4 = 0.96$ and $\hat{\sigma}(M_i)$ the sample standard deviation of the variable M_i, with $M_2 = \hat{U}$, $M_3 = X_1$, and $M_4 = (X_1, \hat{U})$. Since $D_2 = 2$ in the first stage regression estimation, we choose two ROT bandwidths that minimize an asymptotic approximation for the weighted mean integrated squared error of the Nadaraya–Watson estimator (see Ruppert et al., 1995). Using knowledge of DGPs 1 and 2 and the Gaussian kernel we set $h_{1j} = n^{-\delta}(2\sqrt{\pi})^{-1/5}c_{jn}$ for $j = 1, 2$, and $\delta = 1/9$, where

$$c_{1n} = 0.76\left(\max_{1\le i\le n}\{Z_{1i}\} - \min_{1\le i\le n}\{Z_{1i}\}\right)^{1/5} \text{ and } c_{2n} = 0.49\left(\max_{1\le i\le n}\{Z_{1i}\} - \min_{1\le i\le n}\{Z_{1i}\}\right)^{1/5}.$$

We also implement the series estimators proposed by Newey et al. (1999), which we denote by $(\hat{\beta}_{SP}, \hat{m}_{SP})$. It should be noted that their estimator was developed for a model where $\beta_0 = 0$, and the use of a trimming function $w(\tau)$ (in their notation), prevents the use of our assumption $\mathrm{E}(\varepsilon) = 0$. Thus, we adapt their estimation procedure to the DGPs under consideration and use B-splines throughout the implementation. We use the same number of knots to estimate $\Pi(\cdot)$, $m(\cdot)$, and $g(\cdot)$, and follow their constraints on how fast the number of knots diverges to infinity to obtain the convergence results in their Theorem 5.1. Specifically, given D_i for $i = 1, \ldots, 4$ in the DGPs we must select B-splines of order 7 with $s_1 > 6$. Hence, the smallest degree of differentiability permitted for $\Pi(\cdot)$ is $s_1 = 7$, more than we need to assume to attain the uniform rates of convergence for our nonparametric estimator of $m(\cdot)$. The higher degree of smoothness they must assume provides some benefits, specifically, for the DGPs considered here, the rate of uniform convergence in probability of our estimator is $n^{-1/3}$ while theirs is $n^{-5/14}$. In Table 1 we provide results on bias (B), standard deviation (S), root mean squared error (R), and median of root squared error (D) for the estimation of β, and the mean of root mean squared error (M) for estimating $m(\cdot)$ obtained by averaging across the realized values of X_{1i}. We give results for $(\hat{\beta}, \hat{m})$ and for comparison, we also provide results for the oracle estimators of β by taking $m(\cdot)$ as given using two different methods. $\hat{\beta}_{2SLS}$ is derived using the traditional two-stage least square (2SLS) method for linear models, while $\hat{\beta}_{IV}$ is based on IV estimation using the nonparametric proxies $\hat{\Pi}_2$ as in Section 2.2. Lastly, we provide results for the estimators proposed by Robinson (1988), denoted here by $(\hat{\beta}_{Rob}, \hat{m}_{Rob})$, which ignore the endogeneity of X_i.[10] To avoid any extreme estimates or boundary bias in the nonparametric estimation, results on M for estimators of $m(\cdot)$ are only shown by the mean of $10 - 90\%$ quantile range of sample estimates.[11]

[10] We have also compared the performance of $\hat{\beta}$ when U is observed to that of the estimator proposed by Manzan and Zerom (2005). For both DGPs and all sample sizes, $\hat{\beta}$ outperforms the estimator in Manzan and Zerom (2005) as measured by root mean squared error (R) and median of root squared error (D). This superior performance is mainly driven by smaller biases. These simulation results are available from the authors upon request.

[11] Especially for the second DGP since it has a lower bound of zero for the range of the nonparametric component.

Table 1
Finite sample performance.

	$\theta = 0.3$					$\theta = 0.6$					$\theta = 0.9$				
	B	S	R	D	M	B	S	R	D	M	B	S	R	D	M
DGP$_1$	n = 100														
$(\hat{\beta}, \hat{m})$	0.068	0.058	0.090	0.070	0.290	0.104	0.056	0.118	0.102	0.272	0.120	0.053	0.131	0.122	0.266
$(\hat{\beta}_{SP}, \hat{m}_{SP})$	0.064	0.088	0.109	0.073	0.541	0.120	0.085	0.147	0.117	0.609	0.166	0.083	0.186	0.166	0.575
$(\hat{\beta}_{Rob}, \hat{m}_{Rob})$	0.075	0.054	0.092	0.071	0.541	0.137	0.054	0.147	0.132	0.567	0.178	0.056	0.186	0.178	0.603
$(\hat{\beta}_{2SLS}, m)$	0.015	0.763	0.763	0.156		0.028	0.496	0.496	0.160		0.056	0.488	0.491	0.174	
$(\hat{\beta}_{IV}, m)$	0.036	0.048	0.061	0.041		0.073	0.049	0.088	0.071		0.110	0.049	0.121	0.109	
	n = 200														
$(\hat{\beta}, \hat{m})$	0.054	0.040	0.067	0.055	0.270	0.083	0.038	0.091	0.083	0.251	0.096	0.035	0.103	0.097	0.243
$(\hat{\beta}_{SP}, \hat{m}_{SP})$	0.035	0.050	0.061	0.041	0.548	0.075	0.051	0.091	0.077	0.559	0.105	0.048	0.115	0.104	0.613
$(\hat{\beta}_{Rob}, \hat{m}_{Rob})$	0.071	0.037	0.080	0.071	0.526	0.132	0.039	0.138	0.133	0.563	0.175	0.038	0.179	0.174	0.600
$(\hat{\beta}_{2SLS}, m)$	0.041	0.546	0.547	0.175		0.039	0.526	0.527	0.164		0.072	0.495	0.500	0.176	
$(\hat{\beta}_{IV}, m)$	0.027	0.034	0.043	0.032		0.055	0.034	0.065	0.056		0.085	0.033	0.091	0.084	
	n = 400														
$(\hat{\beta}, \hat{m})$	0.047	0.028	0.054	0.048	0.261	0.066	0.027	0.071	0.065	0.235	0.076	0.024	0.080	0.076	0.230
$(\hat{\beta}_{SP}, \hat{m}_{SP})$	0.016	0.032	0.036	0.026	0.525	0.030	0.031	0.043	0.033	0.509	0.042	0.029	0.051	0.044	0.511
$(\hat{\beta}_{Rob}, \hat{m}_{Rob})$	0.073	0.026	0.077	0.073	0.527	0.132	0.027	0.135	0.132	0.564	0.174	0.027	0.176	0.173	0.604
$(\hat{\beta}_{2SLS}, m)$	0.029	0.627	0.627	0.159		0.053	0.603	0.605	0.167		0.074	0.944	0.946	0.177	
$(\hat{\beta}_{IV}, m)$	0.021	0.024	0.032	0.024		0.041	0.024	0.047	0.041		0.061	0.023	0.065	0.061	
DGP$_2$	n = 100														
$(\hat{\beta}, \hat{m})$	0.090	0.060	0.108	0.092	0.164	0.128	0.057	0.140	0.126	0.185	0.146	0.053	0.156	0.146	0.222
$(\hat{\beta}_{SP}, \hat{m}_{SP})$	0.062	0.090	0.109	0.075	0.339	0.122	0.086	0.149	0.120	0.318	0.167	0.084	0.187	0.167	0.332
$(\hat{\beta}_{Rob}, \hat{m}_{Rob})$	0.076	0.053	0.092	0.076	0.254	0.136	0.055	0.147	0.133	0.279	0.175	0.054	0.183	0.176	0.321
$(\hat{\beta}_{2SLS}, m)$	0.024	1.096	1.096	0.162		0.043	0.468	0.470	0.162		0.073	0.412	0.418	0.172	
$(\hat{\beta}_{IV}, m)$	0.037	0.048	0.061	0.043		0.073	0.051	0.089	0.070		0.109	0.048	0.119	0.109	
	n = 200														
$(\hat{\beta}, \hat{m})$	0.076	0.041	0.086	0.075	0.143	0.104	0.040	0.111	0.105	0.170	0.123	0.035	0.128	0.123	0.203
$(\hat{\beta}_{SP}, \hat{m}_{SP})$	0.040	0.052	0.066	0.045	0.342	0.075	0.052	0.092	0.077	0.322	0.104	0.047	0.115	0.103	0.283
$(\hat{\beta}_{Rob}, \hat{m}_{Rob})$	0.074	0.038	0.083	0.072	0.245	0.134	0.038	0.139	0.133	0.277	0.175	0.039	0.179	0.175	0.317
$(\hat{\beta}_{2SLS}, m)$	0.023	0.445	0.445	0.164		0.061	0.705	0.707	0.164		0.052	0.714	0.716	0.160	
$(\hat{\beta}_{IV}, m)$	0.029	0.034	0.045	0.030		0.057	0.035	0.066	0.057		0.085	0.033	0.091	0.084	
	n = 400														
$(\hat{\beta}, \hat{m})$	0.062	0.029	0.069	0.062	0.126	0.086	0.027	0.090	0.088	0.152	0.099	0.026	0.103	0.098	0.187
$(\hat{\beta}_{SP}, \hat{m}_{SP})$	0.017	0.034	0.038	0.027	0.246	0.031	0.031	0.044	0.034	0.297	0.043	0.029	0.052	0.044	0.230
$(\hat{\beta}_{Rob}, \hat{m}_{Rob})$	0.071	0.025	0.076	0.071	0.239	0.132	0.027	0.135	0.131	0.272	0.171	0.028	0.173	0.171	0.320
$(\hat{\beta}_{2SLS}, m)$	0.015	0.546	0.546	0.161		0.063	0.551	0.555	0.168		0.077	0.586	0.591	0.180	
$(\hat{\beta}_{IV}, m)$	0.021	0.024	0.032	0.024		0.041	0.024	0.048	0.041		0.061	0.023	0.065	0.061	

Note: The mean of root mean squared error (M) is intended to be left blank for $\left(\hat{\beta}_{2SLS}, m\right)$ and $\left(\hat{\beta}_{IV}, m\right)$ since m is treated as known and will not be estimated in these cases.

As shown in Table 1, the performances of $\left(\hat{\beta}, \hat{m}\right)$, $(\hat{\beta}_{SP}, \hat{m}_{SP})$ and $\hat{\beta}_{IV}$ improve with the sample size by all of the aforementioned measures (e.g., for DGP$_1$, when $\theta = 0.3$, root mean squared error of $\hat{\beta}$ drops nearly 40% from 0.090 to 0.054 when we increase the sample size from 100 to 400) for both DGPs. The performances of $(\hat{\beta}_{Rob}, \hat{m}_{Rob})$ and $\hat{\beta}_{2SLS}$ do not generally improve with sample size. For all DGPs, sample sizes and values of θ, our nonparametric estimators of m outperforms $\hat{m}_{SP}$ and, as expected, $\hat{m}_{Rob}$. The performance of $\hat{\beta}$ relative to that of $\hat{\beta}_{SP}$ is more nuanced. For DGP$_1$ and $n = 100$ it exhibits smaller B, S, R, and D than $\hat{\beta}_{SP}$ for all θ. For $n = 200$ these results are reversed for B when $\theta = 0.3$ and 0.6. For $n = 400$, $\hat{\beta}_{SP}$ outperforms $\hat{\beta}$ for all θ, except for S. For DGP$_2$, the pattern is more or less similar.

We note that $\hat{\beta}$ and $\hat{\beta}_{SP}$ seem to adequately account for the endogeneity problem since, given the same DGP and sample size, the performance of these estimators regarding bias (B) does not change significantly as the degree of endogeneity (θ) increases, contrasting with the estimator $\hat{\beta}_{Rob}$. In the latter case, as θ increases from 0.3 to 0.9, the bias more than doubles. The performance of $\hat{\beta}_{2SLS}$ is the worst among the five estimators, even though it is derived assuming $m(\cdot)$ is known. This result is not surprising since in 2SLS estimation we specify a linear structure when approximating the endogenous variables, which in our DGPs it is not. This illustrates the importance of nonparametric estimation when we are not able to specify the functional forms of interest. $\hat{\beta}_{IV}$ avoids that potential misspecification and gives the best performance among all estimators for β in every aspect, exactly as we expected.

To give a more visual description of the distribution of root squared error (RSE) for estimators of β across the simulated samples, we estimate and plot its density for each linear estimator with $n = 100$, 200, and 400 for DGP$_1$ in the top, middle,

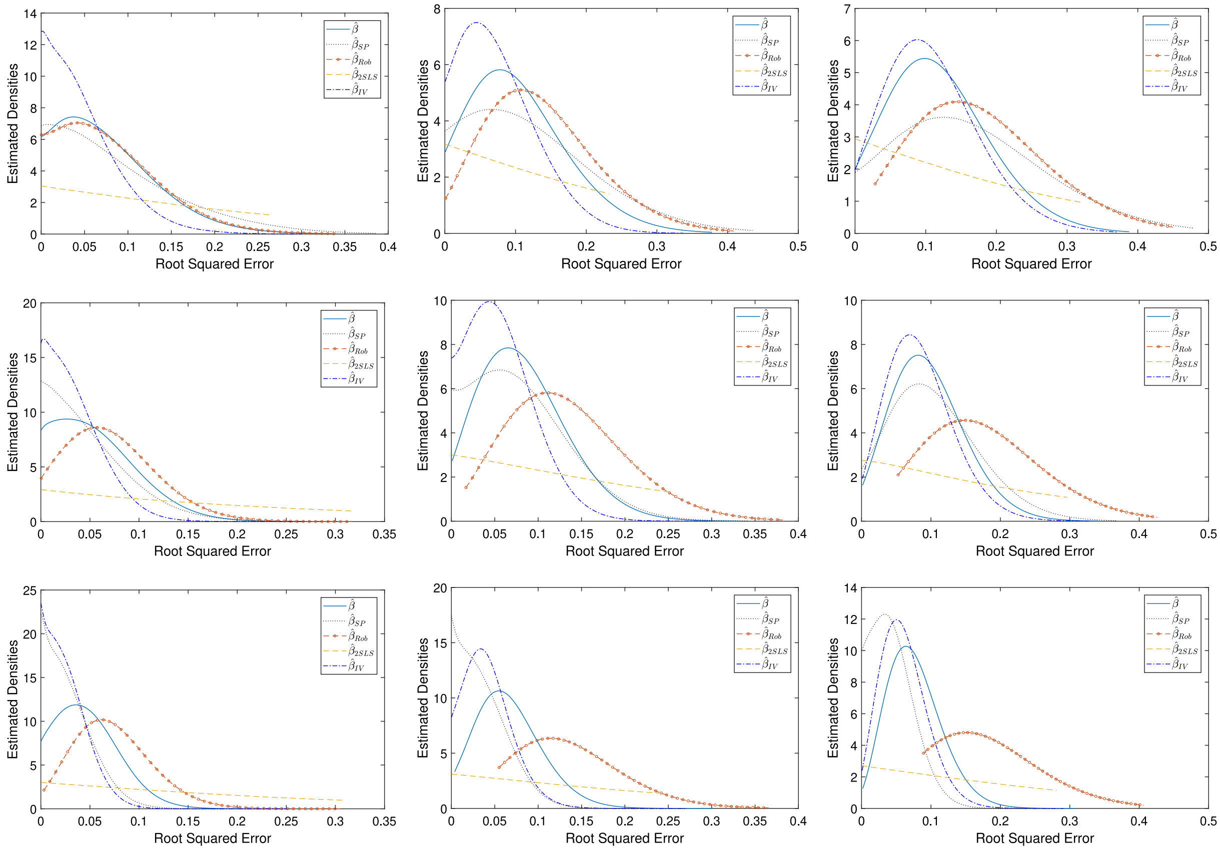

Fig. 1. Estimated densities for RSE of estimators of β for DGP$_1$ with $n = 100$ (top row), $n = 200$ (middle row), and $n = 400$ (bottom row); $\theta = 0.3$ (left column), $\theta = 0.6$ (middle column) and $\theta = 0.9$ (right column).

and bottom panels of Fig. 1. The left, middle, and right panels of Fig. 1 correspond to different degrees of endogeneity, $\theta = 0.3,\ 0.6$, and 0.9, respectively. Fig. 2 displays the same panels for DGP$_2$. The density estimation is performed using the gamma kernel density estimator proposed by Chen (2000) to avoid any boundary bias. It is apparent that the estimated densities for the RSE of estimators $\hat{\beta}_{IV}$ (dashed–dotted graph) are closest to the vertical axis, most concentrated around zero, and exhibit thinnest tails to the right across all the panels in both figures. In Fig. 1 the density associated with our estimator $\hat{\beta}$ (solid graph) is closer to the vertical axis and has thinner tails especially when $\theta = 0.6$ or 0.9. In Fig. 2, it is $\hat{\beta}_{SP}$ (dotted line) that is closer to the vertical axis with thinner tails. The densities associated with the other estimators exhibit particularly bad behavior, especially for large θ.

5. Summary and conclusion

In this paper we contribute to the literature on the estimation of partially linear regression models with generated covariates. We propose easily computable kernel-based estimators for the finite and infinite dimensional parameters of the model and establish their asymptotic distributions. Two critical steps are needed to establish these results: first, the choice of the normalizing function $L(\cdot)$ appearing in Section 2.1, and second the repeated use of the results on U-Statistics obtained in Yao and Martins-Filho (2015). Besides its role in assuring asymptotic normality of the proposed estimators, the choice of $L(\cdot)$ generates a class of estimators with different variances for their asymptotic distributions. Future research should be done on selecting optimal (minimal variance) estimators from this class. In fact, further investigation of the efficiency properties of these estimators may shed light on how to construct oracle efficient estimators for $m(\cdot)$ and semiparametric efficient estimators for β.

Appendix A

This appendix presents the proofs of the main theorems and the supporting lemmas. For the proofs for the lemmas and additional details on the proof of the theorems, e.g., analysis of the order of U-Statistics, we refer readers to the Online Appendix (OA). For a scalar variable x, $f'(x)$ denotes the derivative of $f(x)$ evaluated at x. For $D \times 1$ vectors γ, β, define

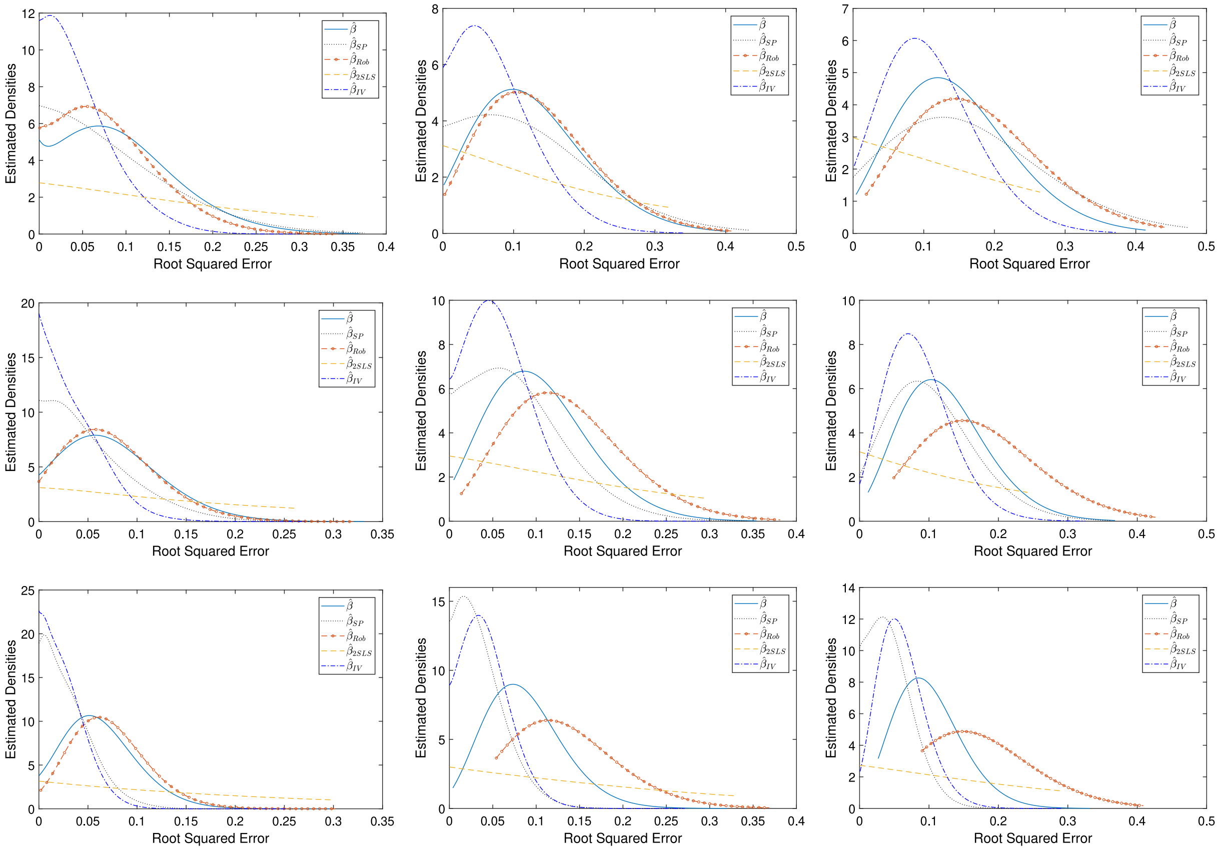

Fig. 2. Estimated densities for RSE of estimators of β for DGP$_2$ with $n = 100$ (top row), $n = 200$ (middle row), and $n = 400$ (bottom row); $\theta = 0.3$ (left column), $\theta = 0.6$ (middle column) and $\theta = 0.9$ (right column).

$\gamma^{\beta} = \prod_{d=1}^{D} \gamma_d^{\beta_d}$, $|\beta| = \sum_{d=1}^{D} \beta_d$, $\mathrm{D}_d f(\gamma) = \frac{\partial}{\partial_d} f(\gamma)$, $\mathrm{D}^2_{dk} f(\gamma) = \frac{\partial^2}{\partial_d \partial_k} f(\gamma)$, $\mathrm{D}^{\beta} f(\gamma) = \frac{\partial^{|\beta|}}{\partial_1^{\beta_1} \ldots \partial_D^{\beta_D}} f(\gamma)$. $\mathbf{J}f(\gamma)$ and $\mathbf{H}f(\gamma)$ denote the Jacobian and Hessian matrix of $f(\gamma)$, respectively. Note that for a scalar function $f(\gamma)$, $\mathbf{J}f(\gamma)$ is the transpose of the gradient vector of $f(\gamma)$. $A \times B$ denotes the Cartesian product of two sets A and B. χ_A denotes the indicator function for the set A. $P(A)$ denotes the probability of event A in the probability space $(\Omega, \mathscr{F}, P)$, $\mathrm{E}(\cdot)$ denotes expectation, and $\mathrm{V}(\cdot)$ denotes variance.

Proof of theorems

Proof of Theorem 1. We establish the uniform convergence order of $\hat{f}_{\hat{U}}(u)$ by taking a Taylor expansion of the kernel function up to order four. To obtain the desired order we explore the structure of the NW estimator that underlies the calculation of $\hat{U}_t - U_t$ and study the order of a U-statistic of degree two.

By the uniform convergence rate of the Rosenblatt density estimator (see, e.g., Theorem 1.4 of Li and Racine, 2007), we have $\sup_{u \in \mathscr{G}_U} \left|\hat{f}_U(u) - f_U(u)\right| = O_p(L_{2n})$, $\sup_{w \in \mathscr{G}_W} \left|\hat{f}_W(w) - f_W(w)\right| = O_p(L_{3n})$, and $\sup_{\{w,u\} \in \mathscr{G}_W \times \mathscr{G}_U} \left|\hat{f}_{WU}(w,u) - f_{WU}(w,u)\right| = O_p(L_{4n})$. Therefore, to prove the first equation in (12), we only need focus on $|\hat{f}_{\hat{U}}(u) - \hat{f}_U(u)|$.

Denote $K_{2ti} = K_2\left(\frac{U_t - U_i}{h_2}\right)$, $\hat{K}_{2ti} = K_2\left(\frac{\hat{U}_t - \hat{U}_i}{h_2}\right)$, $K_{2t} = K_2\left(\frac{U_t - u}{h_2}\right)$, $\hat{K}_{2t} = K_2\left(\frac{\hat{U}_t - u}{h_2}\right)$, and other kernels similarly. Since K_2 is 4-times partially continuously differentiable, by Taylor's Theorem,

$$\hat{f}_{\hat{U}}(u) - \hat{f}_U(u) = \frac{1}{nh_2^{D_2}} \sum_{t=1}^{n} \left(\hat{K}_{2t} - K_{2t}\right) = \frac{1}{nh_2^{D_2}} \sum_{t=1}^{n} \left(\sum_{|\beta|=1}^{3} \frac{H^{\beta}}{\beta!} \mathrm{D}^{\beta} K_{2t} + \sum_{|\beta|=4} \frac{H^{\beta}}{\beta!} \mathrm{D}^{\beta} K_2 \left(\frac{U_t - u}{h_2} + \lambda H\right)\right) \equiv \sum_{i=1}^{4} T_i,$$

where $H \equiv \frac{1}{h_2}(\hat{U}_t - U_t)$ and $\lambda \in (0, 1)$. Next, we examine the uniform order of T_i over $\mathscr{G}_U$ for $i = 1, \ldots, 4$ in four steps.

Step 1: We first rewrite T_1 as:

$$T_1=\sum_{d=1}^{D_2}\Big(\frac{1}{nh_2^{D_2+1}}\sum_{t=1}^{n}(\hat{U}_{td}-U_{td})\mathrm{D}_dK_{2t}\Big)\equiv\sum_{d=1}^{D_2}-T_{1d}.$$

Given $\hat{\Pi}_d(Z_t)=(nh_1^{D_1}\hat{f}_Z(Z_t))^{-1}\sum_{l=1}^{n}K_{1lt}X_{l,d}$, and $\hat{f}_Z(Z_t)=(nh_1^{D_1})^{-1}\sum_{l=1}^{n}K_{1lt}$, we have

$$-(\hat{U}_{td}-U_{td})=\hat{\Pi}_d(Z_t)-\Pi_d(Z_t)=\frac{1}{nh_1^{D_1}f_Z(Z_t)}\sum_{l=1}^{n}K_{1lt}\Big(U_{ld}+\Pi_d(Z_l)-\Pi_d(Z_t)\Big)+O_p(L_{1n}^2) \tag{A.1}$$

by the uniform order of $\hat{f}_Z(Z_t)-f_Z(Z_t)$ and $\hat{U}_{td}-U_{td}$. Thus, we have

$$\begin{aligned}T_{1d}&=\frac{1}{n^2}\sum_{t=1}^{n}\sum_{l=1}^{n}\frac{1}{h_1^{D_1}h_2^{D_2+1}f_Z(Z_t)}K_{1lt}\mathrm{D}_dK_{2t}U_{ld}+\frac{1}{n^2}\sum_{t=1}^{n}\sum_{l=1}^{n}\frac{1}{h_1^{D_1}h_2^{D_2+1}f_Z(Z_t)}K_{1lt}\mathrm{D}_dK_{2t}\Big(\Pi_d(Z_l)-\Pi_d(Z_t)\Big)\\&\quad+O_p(L_{1n}^2)\frac{1}{nh_2^{D_2+1}}\sum_{t=1}^{n}|\mathrm{D}_dK_{2t}|\equiv T_{1d1}+T_{1d2}+O_p(L_{1n}^2/h_2),\\T_{1d1}&=\frac{1}{n^2}\sum_{t=1}^{n}\frac{1}{h_1^{D_1}h_2^{D_2+1}f_Z(Z_t)}K_1(0)\mathrm{D}_dK_{2t}U_{td}+\frac{1}{n^2}\sum_{\substack{t=1\\t\neq l}}^{n}\sum_{l=1}^{n}\frac{1}{h_1^{D_1}h_2^{D_2+1}f_Z(Z_t)}K_{1lt}\mathrm{D}_dK_{2t}U_{ld}\equiv E_{1n}+E_{2n}.\end{aligned}$$

We can show that $E_{1n}=O_p\Big((nh_1^{D_1}h_2)^{-1}\Big)$ uniformly over $\mathscr{G}_U$ by Lemma 3, and $E_{2n}\leq C|U_n|=O_p\Big((\log n/n)^{1/2}+(n^2h_1^{D_1}h_2^{D_2+2})^{-1/2}\Big)$ (see OA 1.1). Together, $T_{1d1}=O_p(L_{1n})$ uniformly by A5.

The order of T_{1d2} can be analyzed in the same way, given that Π and f_Z are s_1 times partially continuously differentiable, and K_1 is a multivariate kernel of order s_1, we have $T_{1d2}=O_p\big(h_1^{s_1}+(\log n/n)^{1/2}+(n^2h_1^{D_1-2}h_2^{D_2+2})^{-1/2}\big)=O_p(L_{1n})$ uniformly by A5. In sum, $\sup_{u\in\mathscr{G}_U}T_1=O_p(L_{1n})$.

Step 2: $T_2=\sum_{|\beta|=2}\big(nh_2^{D_2}\big)^{-1}\sum_{t=1}^{n}H^\beta\mathrm{D}^\beta K_{2t}$, when 1 appears in the dth and the kth position of β, we have

$$\frac{1}{nh_2^{D_2}}\sum_{t=1}^{n}H^\beta\mathrm{D}^\beta K_{2t}=\frac{1}{2nh_2^{D_2+2}}\sum_{t=1}^{n}(\hat{U}_{td}-U_{td})(\hat{U}_{tk}-U_{tk})\mathrm{D}_{dk}^2K_{2t}.$$

Since $\sup_{Z\in\mathscr{G}_Z}\Big|\hat{U}_{tj}-U_{tj}\Big|=O_p(L_{1n})$, for $j=d,k$, we have $T_2=O_p\left(L_{1n}^2/h_2^2\right)\left(nh_2^{D_2}\right)^{-1}\sum_{t=1}^{n}\left|\mathrm{D}_{dk}^2K_{2t}\right|\equiv O_p\left(L_{1n}^2/h_2^2\right)C_2(u)$. $C_2(u)=O_p(1)$ as $\mathrm{E}(|C_2(u)|)=O(1)$ uniformly over $\mathscr{G}_U$. Thus, $\sup_{u\in\mathscr{G}_U}T_2=O_p\left(L_{1n}^2/h_2^2\right)$.

Step 3: Similarly, $\sup_{u\in\mathscr{G}_U}T_3=O_p\left(L_{1n}^3/h_2^3\right)$.

Step 4: T_4 is different from T_2 and T_3 in that $\sup_{u\in\mathscr{G}_U}C_4(u)=O_p(1/h_2^{D_2})$, where $C_4(u)\equiv\left(nh_2^{D_2}\right)^{-1}\sum_{t=1}^{n}\left|\mathrm{D}^\beta K_{2t}^*\right|$, for any $|\beta|=4$, and $\mathrm{D}^\beta K_{2t}^*\equiv\mathrm{D}^\beta K_2((U_t-u)/h_2+\lambda H)$. Thus, $\sup_{u\in\mathscr{G}_U}T_4=O_p\left(L_{1n}^4/h_2^{D_2+4}\right)$. By A5, it can be shown that $T_2,T_3,T_4=o_p(n^{-1/2})$, and $T_1=O_p(L_{1n})=O_p(L_{2n})$, which gives us

$$\sup_{u\in\mathscr{G}_U}|\hat{f}_{\hat{U}}(u)-f_U(u)|=O_p(L_{2n}).$$

The uniform order of $\left|\hat{f}_{W\hat{U}}(w,u)-f_{WU}(w,u)\right|$ can be derived in the similar way under A5, and consequently, here, we omit the details. □

Proof of Theorem 2. We follow the proof of Theorem 1 using a Taylor expansion of the kernel function up to order two to establish the uniform convergence rate of a NW estimator (e.g., $\hat{g}_2(u)$). In addition to taking care of the estimated covariate $\hat{U}_t$ appearing in the kernel, we need to deal with an estimated regressand such as $\hat{\eta}_tX_{2t}$ as well.

We start with the jth element of $\hat{g}_2(u)-g_2(u)$. Note that

$$\begin{aligned}\hat{g}_{2j}(u)-g_{2j}(u)&=\frac{1}{nh_2^{D_2}\hat{f}_{\hat{U}}(u)}\sum_{t=1}^{n}\hat{K}_{2t}\hat{\eta}_tX_{2t,j}-g_{2j}(u)\\&=\frac{1}{nh_2^{D_2}\hat{f}_{\hat{U}}(u)}\sum_{t=1}^{n}\hat{K}_{2t}\underbrace{\Big\{(\hat{\eta}_t-\eta_t)X_{2t,j}+v_{g2t,j}+\big(g_{2j}(U_t)-g_{2j}(u)\big)\Big\}}_{C_{g2t}}\end{aligned}$$

$$
\begin{aligned}
&= \left\{ \frac{1}{nh_2^{D_2} f_U(u)} \sum_{t=1}^{n} K_{2t} C_{g2t} + \frac{1}{nh_2^{D_2+1} f_U(u)} \sum_{t=1}^{n} \mathbf{J} K_{2t} \left(\hat{U}_t - U_t\right) C_{g2t} \right. \\
&\quad \left. + \frac{1}{nh_2^{D_2} f_U(u)} \sum_{t=1}^{n} R_t C_{g2t} \right\} \left(1 + O_p(L_{2n})\right) \qquad \text{(A.2)} \\
&\equiv \left(\sum_{k=1}^{3} T_k \right) \left(1 + O_p(L_{2n})\right),
\end{aligned}
$$

where R_t is the remainder term of a Taylor's expansion of $\hat{K}_{2t}$ at $(U_t - u)/h_2$, and $v_{g2t,j}$ is the jth element of v_{g2t}. We complete the proof by showing in three steps that $T_1 = O_p(L_n)$, $T_2 = O_p\,(L_{1n}/h_2)$, and $T_3 = o_p(n^{-1/2})$.

Step 1: Let $T_1 \equiv \sum_{k=1}^{3} T_{1k}$, corresponding to the three components in C_{g2t} separately. By Theorem 1 and A2, we have

$$
\sup_{\{\hat{U}_t, U_t, W_t\} \in \mathscr{G}_U \times \mathscr{G}_U \times \mathscr{G}_W} |\hat{\eta}_t - \eta_t| = \eta_t O_p(L_{2n} + L_{3n} + L_{4n}) \equiv \eta_t O_p(L_n).
$$

Thus $T_{11} = O_p(L_n)(nh_2^{D_2})^{-1} \sum_{t=1}^{n} |K_{2t}\eta_t X_{2t,j}| = O_p(L_n)$ uniformly, since by A3 and A4,

$$
\begin{aligned}
\mathrm{E}\left(\frac{1}{nh_2^{D_2}} \sum_{t=1}^{n} \left| K_{2t}\eta_t X_{2t,j} \right| \right) &= \frac{1}{h_2^{D_2}} \mathrm{E}\left(\left| K_{2t}(g_{2j}(U_t) + v_{g2t,j}) \right| \right) \\
&\leq \int |K_2(\gamma)|(|g_{2j}(u + h_2\gamma)| + C) f_U(u + h_2\gamma)\, \mathrm{d}\gamma \\
&\leq C\; |g_{2j}(u)| + C \int |K_2(\gamma)|(|g_{2j}(u + h_2\gamma)| - |g_{2j}(u)|)\, \mathrm{d}\gamma + C \\
&\leq C\; |g_{2j}(u)| + C\; h_2 \int |K_2(\gamma)| \sum_{d=1}^{D_2} |\gamma_d|\, \mathrm{d}\gamma + C \\
&\leq C\; |g_{2j}(u)| + C, \quad \text{which is bounded uniformly over } \mathscr{G}_U.
\end{aligned}
$$

By Lemma 3, we have $\sup_{u \in \mathscr{G}_U} |T_{12}| = O_p\left((\log\, n/nh_2^{D_2})^{1/2} \right) = O_p(L_{2n})$, given $\mathrm{E}(T_{12}) = 0$.

For T_{13}, note that by Taylor's Theorem, $\mathrm{E}(T_{13}) = h_2^{-D_2} f_U^{-1}(u) \mathrm{E}\left(K_{2t}\left(g_{2j}(U_t) - g_{2j}(u)\right) \right) = f_U^{-1}(u) \int K_2(\gamma)\left(g_{2j}(u + h_2\gamma) - g_{2j}(u)\right) f_U(u + h_2\gamma)\, \mathrm{d}\gamma = O(h_2^{s_2}) = O(L_{2n})$ uniformly over $\mathscr{G}_U$, given that K_2 is of order s_2, $g_{2j}, f_U \in C^{s_2}$ and all the partial derivatives of g_{2j} up to order s_2 are uniformly bounded by A4. By Lemma 3, we have $h_2^{-1} \sup_{u \in \mathscr{G}_U} |T_{13} - \mathrm{E}(T_{13})| = O_p\left((\log\, n/(nh_2^{D_2}))^{1/2} \right) = O_p(L_{2n})$. Thus, $\sup_{u \in \mathscr{G}_U} |T_{13}| = O_p(L_{2n})$, and we have $T_1 = O_p(L_n)$ uniformly.

Step 2: For T_2, similar to T_{11}, we have

$$
\begin{aligned}
T_2 &= \frac{1}{nh_2^{D_2+1} f_U(u)} \sum_{t=1}^{n} \left(\hat{U}_t - U_t\right) \mathbf{J} K_{2t} C_{g2t} \\
&= O_p\left(\frac{L_{1n}}{h_2} \right) \sum_{d=1}^{D_2} \frac{1}{nh_2^{D_2} f_U(u)} \sum_{t=1}^{n} \left| \mathrm{D}_d K_{2t} \left((\hat{\eta}_t - \eta_t) X_{2t,j} + v_{g2t,j} + \left(g_{2j}(U_t) - g_{2j}(u) \right) \right) \right| \\
&= O_p\left(\frac{L_{1n}}{h_2} \right).
\end{aligned}
$$

Step 3: R_t is the remainder term of a Taylor's expansion of $\hat{K}_{2t}$ at $(U_t - u)/h_2$, thus $R_t = \sum_{|\beta|=2}^{3} (\beta!)^{-1} \mathrm{D}^\beta K_{2t} H^\beta + \sum_{|\beta|=4} (\beta!)^{-1} \mathrm{D}^\beta K_{2t}^* H^\beta$, where $\mathrm{D}^\beta K_{2t}^* \equiv \mathrm{D}^\beta K_2((U_t - u)/h_2 + \lambda H)$, $\lambda \in (0, 1)$, and $H = \left(\hat{U}_t - U_t\right)/h_2$. Thus, let $T_3 \equiv \sum_{k=1}^{3} T_{3k}$, with

$$
\begin{aligned}
T_{31} &= \sum_{d=1}^{D_2} \sum_{l=1}^{D_2} \frac{1}{2nh_2^{D_2+2} f_U(u)} \sum_{t=1}^{n} D_{dl}^2 K_{2t} \left(\hat{U}_{td} - U_{td} \right) \left(\hat{U}_{tl} - U_{tl} \right) C_{g2t} \\
&= O_p\left(\frac{L_{1n}^2}{h_2^2} \right) \sum_{d=1}^{D_2} \sum_{l=1}^{D_2} \frac{1}{nh_2^{D_2}} \sum_{t=1}^{n} \left| D_{dl}^2 K_{2t} C_{g2t} \right| = O_p\left(\frac{L_{1n}^2}{h_2^2} \right),
\end{aligned}
$$

by A3. Similarly, $T_{32} = O_p\left(L_{1n}^3/h_2^3\right)$. By A1, $T_{33} = O_p\left(L_{1n}^4/h_2^{D_2+4}\right) \frac{1}{n} \sum_{t=1}^{n} \left| C_{g2t} \right| = O_p\left(L_{1n}^4/h_2^{D_2+4}\right)$. By A5, we can show that $T_3 = O_p\left(L_{1n}^2/h_2^2 + L_{1n}^3/h_2^3 + L_{1n}^4/h_2^{D_2+4}\right) = o_p(n^{-1/2})$ uniformly.

Combining Steps 1 to 3, we have $\sup_{u\in\mathscr{G}_U}|\hat{g}_2(u)-g_2(u)| = O_p\left(L_n+\frac{L_{1n}}{h_2}\right)$. For $\hat{m}_{2j}(w)-m_{2j}(w)$, note that

$$\begin{aligned}
\hat{m}_{2j}(w)-m_{2j}(w) &= \frac{1}{nh_3^{D_3}\hat{f}_W(w)}\sum_{t=1}^n K_{3t}\hat{\eta}_t X_{2t,j} - m_{2j}(w)\\
&= \left\{\frac{1}{nh_3^{D_3}f_W(w)}\sum_{t=1}^n K_{3t}\underbrace{\left\{(\hat{\eta}_t-\eta_t)X_{2t,j}+v_{m2t,j}+\left(m_{2j}(W_t)-m_{2j}(w)\right)\right\}}_{C_{m2t}}\right\}\left(1+O_p(L_{3n})\right)\\
&= O_p(L_n),
\end{aligned} \tag{A.3}$$

where the order can be found similarly to T_1 in part 1. The uniform orders of $\hat{g}_1(u)$, $\hat{m}_1(w)$, $\hat{g}_3(u)$, and $\hat{m}_3(w)$ can be found similarly by replacing $\hat{\eta}_t X_{2t,j}$ with $\hat{\eta}_t Y_t$ or $\hat{\eta}_t$, respectively. Thus, the details of these proofs are not provided here. □

Proof of Theorem 3. The proof for the asymptotic normality of $\hat{\beta}$ when suitably centered and standardized follows the typical steps once we account for the difficulties encountered in the proofs of Theorems 1 and 2. Using estimated residuals instead of the unknown errors leads to an additional term Φ_2 in the variance of the asymptotic distribution of $\hat{\beta}$ as will be shown in Steps 4 and 5.

Denote the vector matrix format in bold face, e.g., $\boldsymbol{m} \equiv \left(m(W_1),\ldots,m(W_n)\right)'$. Note that $\boldsymbol{m} = \boldsymbol{m}_1 - \boldsymbol{m}_2\beta - \beta_0$ and $\boldsymbol{g} = \boldsymbol{g}_1 - \boldsymbol{g}_2\beta - \beta_0$. Denote $\boldsymbol{V}_Y \equiv \sum_{k=\{m,g,\mu\}}\boldsymbol{V}_{k1}$ and $\boldsymbol{V}_X \equiv \sum_{k=\{m,g,\mu\}}\boldsymbol{V}_{k2}$, where $\boldsymbol{V}_{m1} \equiv \hat{\boldsymbol{m}}_1 - \boldsymbol{m}_1$, $\boldsymbol{V}_{g1} \equiv \hat{\boldsymbol{g}}_1 - \boldsymbol{g}_1$, $\boldsymbol{V}_{\mu 1} \equiv -(\hat{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1)$, $\boldsymbol{V}_{m2} \equiv \hat{\boldsymbol{m}}_2 - \boldsymbol{m}_2$, $\boldsymbol{V}_{g2} \equiv \hat{\boldsymbol{g}}_2 - \boldsymbol{g}_2$, and $\boldsymbol{V}_{\mu 2} \equiv -(\hat{\boldsymbol{\mu}}_2 - \boldsymbol{\mu}_2)$. Thus, since $\hat{\boldsymbol{Y}} = \boldsymbol{Y}^* - \boldsymbol{V}_Y$, $\hat{\boldsymbol{X}}_2 = \boldsymbol{X}_2^* - \boldsymbol{V}_X$, and $\hat{\boldsymbol{Y}} - \hat{\boldsymbol{X}}_2\beta = \boldsymbol{v} - \sum_{k=\{m,g,\mu\}}(\boldsymbol{V}_{k1} - \boldsymbol{V}_{k2}\beta)$, we have

$$\hat{\beta}-\beta = \left(\frac{1}{n}\hat{\boldsymbol{X}}_2'\hat{\boldsymbol{\eta}}\hat{\boldsymbol{X}}_2\right)^{-1}\frac{1}{n}\hat{\boldsymbol{X}}_2'\hat{\boldsymbol{\eta}}\left(\hat{\boldsymbol{Y}}-\hat{\boldsymbol{X}}_2\beta\right),$$

where
$$\begin{aligned}
\frac{1}{n}\hat{\boldsymbol{X}}_2'\hat{\boldsymbol{\eta}}\hat{\boldsymbol{X}}_2 &= \frac{1}{n}\boldsymbol{X}_2^{*\prime}\hat{\boldsymbol{\eta}}\boldsymbol{X}_2^* - \frac{1}{n}\boldsymbol{X}_2^{*\prime}\hat{\boldsymbol{\eta}}\boldsymbol{V}_X - \frac{1}{n}\boldsymbol{V}_X'\hat{\boldsymbol{\eta}}\boldsymbol{X}_2^* + \frac{1}{n}\boldsymbol{V}_X'\hat{\boldsymbol{\eta}}\boldsymbol{V}_X \equiv \sum_{k=1}^4 A_k,\\
\frac{1}{n}\hat{\boldsymbol{X}}_2'\hat{\boldsymbol{\eta}}(\hat{\boldsymbol{Y}}-\hat{\boldsymbol{X}}_2\beta) &= \frac{1}{n}\hat{\boldsymbol{X}}_2'\hat{\boldsymbol{\eta}}\boldsymbol{v} - \frac{1}{n}\hat{\boldsymbol{X}}_2'\hat{\boldsymbol{\eta}}(\boldsymbol{V}_{m1}-\boldsymbol{V}_{m2}\beta) - \frac{1}{n}\hat{\boldsymbol{X}}_2'\hat{\boldsymbol{\eta}}(\boldsymbol{V}_{g1}-\boldsymbol{V}_{g2}\beta) - \frac{1}{n}\hat{\boldsymbol{X}}_2'\hat{\boldsymbol{\eta}}(\boldsymbol{V}_{\mu 1}-\boldsymbol{V}_{\mu 2}\beta) \equiv \sum_{k=1}^4 B_k.
\end{aligned}$$

The proof has five steps:

(1) We show that $A_1 \xrightarrow{p} \Phi_0$ and $A_2, A_3, A_4 = o_p(1)$.
(2) We show that $\sqrt{n}B_1 \xrightarrow{d} \mathscr{N}(0,\Phi_1)$.
(3) We show that $B_2, B_4 = o_p(n^{-1/2})$.
(4) We show that $B_3 = \frac{1}{n}\sum_{i=1}^n a_{ni} + o_p(n^{-1/2})$, where $a_{ni} \equiv -\sum_{d=1}^{D_2}(h_1^{D_1}h_2^{D_2})^{-1}U_{id}\mathrm{E}\left(\frac{\eta_l X_{2l}^* D_d K_{2tl} K_{1il}}{f_U(U_l)f_Z(Z_l)}\mathbf{J}g(U_l)\left(\frac{U_t-U_l}{h_2}\right)\Big|Z_i\right)$.
(5) Combining (1)–(4), we show that $\sqrt{n}(\hat{\beta}-\beta) \xrightarrow{d} \mathscr{N}\left(0, \Phi_0^{-1}(\Phi_1+\Phi_2)\Phi_0^{-1}\right)$.

Step 1: By uniform order of $|\hat{\eta}_i-\eta_i|$, Kolmogorov's LLN and A3, we have

$$A_1 = \frac{1}{n}\sum_{i=1}^n \hat{\eta}_i X_{2i}^* X_{2i}^{*\prime} = \frac{1}{n}\sum_{i=1}^n \eta_i X_{2i}^* X_{2i}^{*\prime} + O_p(L_n)\frac{1}{n}\sum_{i=1}^n |X_{2i}^* X_{2i}^{*\prime}| \xrightarrow{p} \Phi_0,$$

where $\Phi_{0_{(j,k)}} \equiv \mathrm{E}(\eta_t X_{2t,j}^* X_{2t,k}^*) = \mathrm{E}\Big\{\eta_t(X_{2t,j}-m_{2j}(W_t)-g_{2j}(U_t)+\mu_{2j})(X_{2t,k}-m_{2k}(W_t)-g_{2k}(U_t)+\mu_{2k})\Big\} < \infty$, since $\{\eta_i X_{2i}^* X_{2i}^{*\prime}\}_{i=1}^n$ is an IID sequence, and $\mathrm{E}|\eta_i X_{2i,k}^* X_{2i,j}^*| < \infty$ due to (i) η_i is uniformly bounded; (ii) $\mathrm{E}|X_{2i,j}X_{2i,k}| \le \left(\mathrm{E}(X_{2i,j}^2)\mathrm{E}(X_{2i,k}^2)\right)^{1/2} < \infty$ by Cauchy–Schwarz Inequality; (iii) $\mathrm{E}|X_{2i,j}m_{2k}(W_i)| \le \left(\mathrm{E}(X_{2i,j}^2)\mathrm{E}(m_{2k}^2(W_i))\right)^{1/2}$; (iv) $\mathrm{E}(m_{2k}^2(W_i)) = \mathrm{E}\left(\mathrm{E}(\eta_i X_{2i,k}|W_i)^2\right) \le \mathrm{E}\left(\mathrm{E}(\eta_i^2 X_{2i,k}^2|W_i)\right) = \mathrm{E}(\eta_i^2 X_{2i,k}^2) < \infty$. By the non-singularity of Φ_0 we have $A_1^{-1} \xrightarrow{p} \Phi_0^{-1}$. And for $-A_2 = \frac{1}{n}\sum_{i=1}^n \hat{\eta}_i X_{2i}^* V_{Xi}'$, the $(k,j)^{th}$ element is $-A_{2_{(k,j)}} = \frac{1}{n}\sum_{i=1}^n \hat{\eta}_i X_{2i,k}^* V_{Xi,j} \le O_p(L_n + L_{1n}/h_2)\frac{1}{n}\sum_{i=1}^n |X_{2i,k}^*| = o_p(1)$ by Theorem 2. Similarly we have $A_3, A_4 = o_p(1)$. Thus, $\left(\frac{1}{n}\hat{\boldsymbol{X}}_2'\hat{\boldsymbol{\eta}}\hat{\boldsymbol{X}}_2\right)^{-1} \xrightarrow{p} \Phi_0^{-1}$.

Step 2: We rewrite B_1 into four elements:

$$B_1 = \frac{1}{n}\sum_{i=1}^n \hat{X}_{2i}\hat{\eta}_i v_i = \frac{1}{n}\sum_{i=1}^n X_{2i}^*\eta_i v_i - \frac{1}{n}\sum_{i=1}^n V_{Xi}(\hat{\eta}_i-\eta_i)v_i + \frac{1}{n}\sum_{i=1}^n X_{2i}^*(\hat{\eta}_i-\eta_i)v_i - \frac{1}{n}\sum_{i=1}^n V_{Xi}\eta_i v_i \equiv \sum_{k=1}^4 B_{1k},$$

and show that $\sqrt{n}B_1 \xrightarrow{d} \mathscr{N}(0,\Phi_1)$ by establishing that $\sqrt{n}B_{11} \xrightarrow{d} \mathscr{N}(0,\Phi_1)$, and $B_{12}, B_{13}, B_{14} = o_p(n^{-1/2})$.

First, by Levy's Central Limit Theorem and the Cramer–Wold device, we have $\sqrt{n}B_{11} \xrightarrow{d} \mathcal{N}(0, \Phi_1)$, since (i) $\{X_{2i}^*\eta_i v_i\}_{i=1}^n$ is IID; (ii) $\mathrm{E}(X_{2i}^*\eta_i v_i) = 0$; (iii) $\mathrm{E}(v_i^2|Z_i, U_i) = \sigma_v^2$; (iv) $\mathrm{V}(X_{2i}^*\eta_i v_i) = \mathrm{E}(X_{2i}^*\eta_i^2 v_i^2 X_{2i}^{*\prime}) = \sigma_v^2 \mathrm{E}(\eta_i^2 X_{2i}^* X_{2i}^{*\prime}) \equiv \Phi_1 < \infty$, where $\Phi_{1_{(j,k)}} = \sigma_v^2\ \mathrm{E}(\eta_t^2 X_{2t,j}^* X_{2t,k}^*) = \sigma_v^2\ \mathrm{E}\Big\{\eta_t^2(X_{2t,j} - m_{2j}(W_t) - g_{2j}(U_t) + \mu_{2j})(X_{2t,k} - m_{2k}(W_t) - g_{2k}(U_t) + \mu_{2k})\Big\} < \infty$.

Second, given that $|V_{Xi}|, |\hat{\eta}_i - \eta_i| = O_p(L_n + L_{1n}/h_2)$, we have $B_{12} = O_p(L_n^2 + L_{1n}^2/h_2^2)\frac{1}{n}\sum_{i=1}^n |v_i| = o_p(n^{-1/2})$.

Third, the jth element of B_{13} is $\frac{1}{n}\sum_{i=1}^n G(M_i)(\hat{\eta}_i(W_i, \hat{U}_i) - \eta_i(W_i, U_i))$, where $G(M_i) \equiv X_{2i,j}^* v_i$ and $M_i \equiv (X_i, Z_i, U_i, \varepsilon_i)$. Note that since $\mathrm{E}(v_i|X_i, Z_i, U_i) = 0$, $\mathrm{E}(G(M_i)|X_i, Z_i, U_i) = 0$. In addition, $\mathrm{E}(G^2(M_i)) = \mathrm{E}(X_{2i,j}^{*2} v_i^2) < \infty$ by A3. By A4, $G(M_i)$ is continuous, hence using Lemma 4, $B_{13} = o_p(n^{-1/2})$.

Fourth, for B_{14}, the jth element can be written as

$$-B_{14,j} = \frac{1}{n}\sum_{i=1}^n V_{Xi,j}\eta_i v_i = \frac{1}{n}\sum_{i=1}^n V_{m2i,j}\eta_i v_i + \frac{1}{n}\sum_{i=1}^n V_{g2i,j}\eta_i v_i + \frac{1}{n}\sum_{i=1}^n V_{\mu 2i,j}\eta_i v_i \equiv \sum_{k=1}^3 B_{14k}.$$

We show that $B_{14k} = o_p(n^{-1/2})$ for $k = 1, 2, 3$.

Note that $B_{143} = -\frac{1}{n}\sum_{i=1}^n(\hat{\mu}_{2j} - \mu_{2j})\eta_i v_i = -(\hat{\mu}_{2j} - \mu_{2j})\frac{1}{n}\sum_{i=1}^n \eta_i v_i = O_p(L_n)O_p(n^{-1/2}) = o_p(n^{-1/2})$.

For B_{141}, given that $\big(nh_3^{D_3} f_W(W_i)\big)^{-1}\sum_{t=1}^n K_{3ti}C_{m2ti} = O_p(L_n)$, and by the decomposition of $\hat{m}_{2j}(W_i) - m_{2j}(W_i)$ in (A.3) from the proof of Theorem 2, we have

$$B_{141} = \frac{1}{n^2}\sum_{i=1}^n\sum_{t=1}^n \frac{\eta_i v_i K_{3ti}}{h_3^{D_3} f_W(W_i)} C_{m2ti,j} + \frac{1}{n}\sum_{i=1}^n \big|\eta_i v_i\big| O_p(L_n)O_p(L_{3n}) \equiv \sum_{k=1}^3 B_{141k} + o_p(n^{-1/2}),$$

where
$$B_{1411} = \frac{1}{n^2}\sum_{i=1}^n\sum_{t=1}^n \frac{\eta_i v_i K_{3ti}}{h_3^{D_3} f_W(W_i)}(\hat{\eta}_t - \eta_t)X_{2t,j}, \qquad B_{1412} = \frac{1}{n^2}\sum_{i=1}^n\sum_{t=1}^n \frac{\eta_i v_i K_{3ti}}{h_3^{D_3} f_W(W_i)} v_{m2t,j},$$
$$B_{1413} = \frac{1}{n^2}\sum_{i=1}^n\sum_{t=1}^n \frac{\eta_i v_i K_{3ti}}{h_3^{D_3} f_W(W_i)}\big(m_{2j}(W_t) - m_{2j}(W_i)\big).$$

We have $B_{141} = o_p(n^{-1/2})$ by showing $B_{141k} = o_p(n^{-1/2})$ for $k = 1, 2, 3$.

1.1. Let $Q_t \equiv \frac{1}{n}\sum_{i=1}^n\big(h_3^{D_3} f_W(W_i)\big)^{-1}\eta_i v_i K_{3ti}$. So $B_{1411} = \frac{1}{n}\sum_{t=1}^n(\hat{\eta}_t - \eta_t)X_{2t,j}Q_t$. By Lemma 3, we can show that $Q_t = O_p(L_{3n})$ uniformly over $\mathcal{G}_W$, given A3 and $\mathrm{E}(Q_t) = 0$. Given $\hat{\eta}_t - \eta_t = \eta_t O_p(L_n)$ uniformly, we have $B_{1411} = O_p(L_n)O_p(L_{3n})\frac{1}{n}\sum_{t=1}^n |\eta_t X_{2t,j}| = o_p(n^{-1/2})$ by A5.

1.2. Let $B_{1412} \equiv \frac{1}{n^2}\sum_{i=1}^n\sum_{t=1}^n \psi_{nit} \equiv E_{1n} + E_{2n}$, where $\psi_{nit} \equiv \big(h_3^{D_3} f_W(W_i)\big)^{-1}\eta_i v_i K_{3ti} v_{m2t,j}$. Thus, $E_{1n} = \frac{1}{n^2}\sum_{i=1}^n \psi_{nii} = o_p(n^{-1/2})$ by Chebyshev's Inequality, since $\mathrm{E}(E_{1n}) = 0$, $\mathrm{V}(E_{1n}) = \frac{1}{n^3}\mathrm{E}(\psi_{nii}^2) = O\big(n^{-3}h_3^{-D_3}\big) = o(n^{-1})$. And $|E_{2n}| \le C|U_n|$, where U_n is a U-statistic of degree 2 such that $U_n = \binom{n}{2}^{-1}\sum_{\substack{i=1\\ i<t}}^n\sum_{t=1}^n \phi_{nit}$ with $\phi_{nit} \equiv \psi_{nit} + \psi_{nti}$. U_n is $o_p(n^{-1/2})$ as shown in OA 3.1. Thus, $B_{1412} = o_p(n^{-1/2})$.

1.3. Given $B_{1413} = \frac{1}{n^2}\sum_{\substack{i=1\\ i\ne t}}^n\sum_{t=1}^n \psi_{nit}$, where $\psi_{nit} = h_3^{-D_3} f_W^{-1}(W_i)\eta_i v_i K_{3ti}\big(m_{2j}(W_t) - m_{2j}(W_i)\big)$, we have $|B_{1413}| \le C|U_n| = o_p(n^{-1/2})$ as shown in OA 3.2. Thus, we have $B_{1413} = o_p(n^{-1/2})$.

For B_{142}, as in the proof of Theorem 2, $V_{g2i,j} = \hat{g}_{2j}(\hat{U}_i) - g_{2j}(U_i) \equiv \big(\sum_{k=1}^3 T_k\big)(1 + O_p(L_{2n}))$, where $T_1 = O_p(L_n)$, $T_2 = O_p\,(L_{1n}/h_2)$, and $T_3 = o_p(n^{-1/2})$. Thus, by the decomposition of $V_{g2i,j}$ in (A.2), we have

$$\begin{aligned} B_{142} &= \frac{1}{n}\sum_{i=1}^n V_{g2i,j}\eta_i v_i = \sum_{k=1}^3 B_{142k} + \frac{1}{n}\sum_{i=1}^n |\eta_i v_i| \Big(\big(o_p(n^{-\frac{1}{2}}) + O_p(L_n) + O_p(L_{1n}/h_2)\big)O_p(L_{2n})\Big) \\ &\equiv \sum_{k=1}^3 B_{142k} + o_p(n^{-1/2}) \quad \text{by A5,} \end{aligned}$$

where
$$B_{1421} = \frac{1}{n^2}\sum_{i=1}^n\sum_{t=1}^n \frac{\eta_i v_i}{h_2^{D_2} f_U(U_i)} K_{2ti}C_{g2ti}, \qquad B_{1422} = -\frac{1}{n^2}\sum_{i=1}^n\sum_{t=1}^n \frac{\eta_i v_i}{h_2^{D_2+1} f_U(U_i)}\mathbf{J}K_{2ti}(\hat{U}_i - U_i)C_{g2ti},$$
$$B_{1423} = \frac{1}{n^2}\sum_{i=1}^n\sum_{t=1}^n \frac{\eta_i v_i}{h_2^{D_2+1} f_U(U_i)}\mathbf{J}K_{2ti}(\hat{U}_t - U_t)C_{g2ti}, \quad C_{g2ti} = (\hat{\eta}_t - \eta_t)X_{2t,j} + v_{g2t,j} + \big(g_{2j}(U_t) - g_{2j}(U_i)\big).$$

Similar to B_{141} we just analyzed, we have $B_{1421} = o_p(n^{-1/2})$, with U_i replacing W_i. B_{1422} and B_{1423} are similar in structure, so here we only show that $B_{1422} = o_p(n^{-1/2})$. Given the three components in C_{g2ti}, let $B_{1422} = \sum_{k=1}^3 B_{1422k}$.

We show $B_{1422k} = o_p(n^{-1/2})$ for $k = 1, 2, 3$.

2.1.

$$
\begin{aligned}
B_{14221} &= -\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i v_i}{h_2^{D_2+1}f_U(U_i)}\mathbf{J}K_{2ti}(\hat{U}_i-U_i)(\hat{\eta}_t-\eta_t)X_{2t,j}\\
&\le O_p(L_n)O_p\left(\frac{L_{1n}}{h_2}\right)\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{d=1}^{D_2}\frac{\left|\eta_i v_i\eta_t X_{2t,j}\mathrm{D}_d K_{2ti}\right|}{h_2^{D_2}f_U(U_i)}\\
&= O_p(L_n)O_p\left(\frac{L_{1n}}{h_2}\right)=o_p(n^{-1/2}),\quad \text{by A5.}
\end{aligned}
$$

2.2. By (A.1) in the proof of Theorem 1, we have

$$
\begin{aligned}
B_{14222} &= -\sum_{d=1}^{D_2}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i v_i v_{g2t,j}\mathrm{D}_d K_{2ti}}{h_2^{D_2+1}f_U(U_i)}(\hat{U}_{id}-U_{id})\\
&= \sum_{d=1}^{D_2}\Bigg\{\frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\frac{\eta_i v_i v_{g2t,j}\mathrm{D}_d K_{2ti}K_{1li}}{h_1^{D_1}h_2^{D_2+1}f_U(U_i)f_Z(Z_i)}\Big(U_{ld}+\big(\Pi_d(Z_l)-\Pi_d(Z_i)\big)\Big)\\
&\quad +O_p(L_{1n}^2)\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\left|\frac{\eta_i v_i v_{g2t,j}\mathrm{D}_d K_{2ti}}{h_2^{D_2+1}f_U(U_i)}\right|\Bigg\}\\
&\equiv \sum_{d=1}^{D_2}(T_{1d}+T_{2d})+o_p(n^{-1/2}).
\end{aligned}
$$

where the last equality follows by Markov's Inequality and $O_p(L_{1n}^2/h_2)=o_p(n^{-1/2})$ by A5. We have $B_{14222}=o_p(n^{-1/2})$ by showing that T_{1d} and T_{2d} are $o_p(n^{-1/2})$ in OA 3.3.

2.3. Similar to part 2.2, we have

$$
\begin{aligned}
B_{14223} &= -\sum_{d=1}^{D_2}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i v_i \mathrm{D}_d K_{2ti}\big(g_{2j}(U_t)-g_{2j}(U_i)\big)}{h_2^{D_2+1}f_U(U_i)}(\hat{U}_{id}-U_{id})\\
&= \sum_{d=1}^{D_2}\Bigg\{\frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\frac{\eta_i v_i \mathrm{D}_d K_{2ti}K_{1li}\big(g_{2j}(U_t)-g_{2j}(U_i)\big)}{h_1^{D_1}h_2^{D_2+1}f_U(U_i)f_Z(Z_i)}\Big(U_{ld}+\big(\Pi_d(Z_l)-\Pi_d(Z_i)\big)\Big)\\
&\quad +O_p(L_{1n}^2)\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\left|\frac{\eta_i v_i \mathrm{D}_d K_{2ti}\big(g_{2j}(U_t)-g_{2j}(U_i)\big)}{h_2^{D_2+1}f_U(U_i)}\right|\Bigg\}\equiv\sum_{d=1}^{D_2}(T_{1d}+T_{2d})+o_p(n^{-1/2}).
\end{aligned}
$$

We have $B_{14223}=o_p(n^{-1/2})$ by showing that T_{1d} and T_{2d} are $o_p(n^{-1/2})$ in OA 3.4. In sum, we have $B_{142}=o_p(n^{-1/2})$.

Combining all the terms in Step 2, we have $B_1=B_{11}+o_p(n^{-1/2})$, where $\sqrt{n}B_{11}\xrightarrow{d}\mathcal{N}(0,\Phi_1)$. Thus, $\sqrt{n}B_1\xrightarrow{d}\mathcal{N}(0,\Phi_1)$.

Step 3: We first show that $B_4=o_p(n^{-1/2})$. Note that

$$
-B_4=\frac{1}{n}\hat{\boldsymbol{X}}_2'\hat{\boldsymbol{\eta}}(\boldsymbol{V}_{\mu1}-\boldsymbol{V}_{\mu2}\beta)=\frac{1}{n}\hat{\boldsymbol{X}}_2'\boldsymbol{\eta}\boldsymbol{V}_{\mu1}-\frac{1}{n}\hat{\boldsymbol{X}}_2'\boldsymbol{\eta}\boldsymbol{V}_{\mu2}\beta+\frac{1}{n}\hat{\boldsymbol{X}}_2'(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta})(\boldsymbol{V}_{\mu1}-\boldsymbol{V}_{\mu2}\beta)\equiv\sum_{k=1}^{3}B_{4k}.
$$

By Theorems 1 and 2, we have $|\hat{\eta}_i-\eta_i|$, $V_{\mu2i}$, $V_{\mu1i}=O_p(L_n)$. Given that $V_{\mu1i}$ is the same across i, we have $B_{41}=V_{\mu1i}\big(\frac{1}{n}\sum_{i=1}^n X_{2i}^*\eta_i-\frac{1}{n}\sum_{i=1}^n V_{Xi}\eta_i\big)=O_p(L_n)\big(O_p(n^{-1/2})+O_p(L_n)\big)=o_p(n^{-1/2})$ by A5. $B_{42}=o_p(n^{-1/2})$ follows similarly, and $B_{43}=O_p(L_n^2)=o_p(n^{-1/2})$ by A5.

Then, we show $B_2=o_p(n^{-1/2})$. Note

$$
-B_2=\frac{1}{n}\hat{\boldsymbol{X}}_2'\boldsymbol{\eta}\boldsymbol{V}_{m1}-\frac{1}{n}\hat{\boldsymbol{X}}_2'\boldsymbol{\eta}\boldsymbol{V}_{m2}\beta+\frac{1}{n}\hat{\boldsymbol{X}}_2'(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta})(\boldsymbol{V}_{m1}-\boldsymbol{V}_{m2}\beta)\equiv\sum_{k=1}^{3}B_{2k}.
$$

$B_{23}=O_p(L_n^2)=o_p(n^{-1/2})$ by A5. B_{22} is of the same structure as B_{21}, thus we only show that $B_{21}=o_p(n^{-1/2})$.

Note that $B_{21}=\frac{1}{n}\sum_{i=1}^n X_{2i}^*\eta_i V_{m1i}-\frac{1}{n}\sum_{i=1}^n V_{Xi}\eta_i V_{m1i}\equiv B_{21}'+o_p(n^{-1/2})$ by Theorem 2. By the decomposition of V_{m1i}, similar to V_{m2i} given in (A.3) from the proof of Theorem 2, we have the jth element of B_{21}' as

$$
B_{21}'=\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i X_{2i,j}^* K_{3ti}}{h_3^{D_3}f_W(W_i)}C_{m1ti}+O_p(L_{3n})O_p(L_n)\frac{1}{n}\sum_{i=1}^{n}\left|\eta_i X_{2i,j}^*\right|\equiv\sum_{k=1}^{3}B_{21k}+o_p(n^{-\frac{1}{2}}),
$$

where
$$B_{211} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i X^*_{2i,j}K_{3ti}}{h_3^{D_3}f_W(W_i)}(\hat{\eta}_t-\eta_t)Y_t, \qquad B_{212} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i X^*_{2i,j}K_{3ti}}{h_3^{D_3}f_W(W_i)}v_{m1t},$$
$$B_{213} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i X^*_{2i,j}K_{3ti}}{h_3^{D_3}f_W(W_i)}\big(m_1(W_t)-m_1(W_i)\big).$$

We show that $B_{21k}=o_p(n^{-1/2})$ for $k=1,2,3$.

3.1. Let $Q_t \equiv \frac{1}{n}\sum_{i=1}^{n}\big(h_3^{D_3}f_W(W_i)\big)^{-1}\eta_i X^*_{2i,j}K_{3ti}$. So $B_{211} = \frac{1}{n}\sum_{t=1}^{n}(\hat{\eta}_t-\eta_t)Y_tQ_t$. By Lemma 3, we can show that $Q_t = O_p(L_{3n})$ uniformly over $\mathscr{G}_W$, given A3 and $\mathrm{E}(Q_t)=0$. Given $\hat{\eta}_t-\eta_t=\eta_tO_p(L_n)$ uniformly, we have $B_{211}=O_p(L_n)O_p(L_{3n})\frac{1}{n}\sum_{t=1}^{n}|\eta_tY_t| = o_p(n^{-1/2})$ by A5.

3.2. $B_{212} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}h_3^{-D_3}f_W^{-1}(W_i)\eta_iX^*_{2i,j}K_{3ti}v_{m1t} \equiv \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\psi_{nit} \equiv E_{1n}+E_{2n}$, where $E_{1n} = \frac{1}{n^2}\sum_{i=1}^{n}\psi_{nii} = \frac{1}{n^2}\sum_{i=1}^{n}h_3^{-D_3}f_W^{-1}(W_i)\eta_iX^*_{2i,j}K_3(0)v_{m1i} = O_p\big((nh_3^{D_3})^{-1}\big)=o_p(n^{-1/2})$, and $|E_{2n}|\le C|U_n| = o_p(n^{-1/2})$ as shown in OA 3.5 with $U_n=\binom{n}{2}^{-1}\sum_{i=1, i\neq t}^{n}\sum_{t=1}^{n}\psi_{nit}$. We have $B_{212}=o_p(n^{-1/2})$.

3.3. $|B_{213}| \le C|U_n| = o_p(n^{-1/2})$ as shown in OA 3.6, where $U_n = \binom{n}{2}^{-1}\sum_{i=1, i\neq t}^{n}\sum_{t=1}^{n}\psi_{nit}$ with $\psi_{nit}\equiv h_3^{-D_3}f_W^{-1}(W_i)\eta_iX^*_{2i,j}K_{3ti}\big(m_1(W_t)-m_1(W_i)\big)$.

By 3.1–3.3, we have $B_{21}=o_p(n^{-1/2})$.

Step 4: For B_3, we have $-B_3 = \frac{1}{n}\hat{\boldsymbol{X}}_2'\boldsymbol{\eta}(\boldsymbol{V}_{g1}-\boldsymbol{V}_{g2}\beta)+o_p(n^{-1/2}) \equiv B_{31}+B_{32}+o_p(n^{-1/2})$. We will focus on B_{31} here, since B_{32} has a similar structure to B_{31} and could be analyzed accordingly. By Theorem 2, we have $B_{31} = \frac{1}{n}\sum_{i=1}^{n}X^*_{2i}\eta_iV_{g1i} - \frac{1}{n}\sum_{i=1}^{n}V_{Xi}\eta_iV_{g1i} \equiv B'_{31}+o_p(n^{-1/2})$. Similar to (A.2) given in the proof of Theorem 2, by Taylor's Theorem, we have

$$V_{g1i} = \hat{g}_1(\hat{U}_i)-g_1(U_i) = \left\{\frac{1}{nh_2^{D_2}f_U(U_i)}\sum_{t=1}^{n}K_{2ti}C_{g1ti} + \frac{1}{nh_2^{D_2+1}f_U(U_i)}\sum_{t=1}^{n}\mathbf{J}K_{2ti}\Big(\hat{U}_t-U_t-\big(\hat{U}_i-U_i\big)\Big)C_{g1ti}\right.$$
$$\left. + \frac{1}{nh_2^{D_2}f_U(U_i)}\sum_{t=1}^{n}R_{ti}C_{g1ti}\right\}\Big(1+O_p(L_{2n})\Big),$$

where $C_{g1ti} \equiv (\hat{\eta}_t-\eta_t)Y_t+v_{g1t}+\big(g_1(U_t)-g_1(U_i)\big)$, and R_{ti} is the remainder term of a Taylor's expansion of $\hat{K}_{2ti}$ at $(U_t-U_i)/h_2$.

Similar to the T_3 term in the proof of Theorem 2, we have $(nh_2^{D_2}f_U(U_i))^{-1}\sum_{t=1}^{n}R_{ti}C_{g1ti}=o_p(n^{-1/2})$ uniformly. Thus, we have the jth element of B'_{31} as

$$B'_{31,j} = \frac{1}{n}\sum_{i=1}^{n}X^*_{2i,j}\eta_iV_{g1i} = \sum_{k=1}^{3}B_{31k}+\frac{1}{n}\sum_{i=1}^{n}|\eta_iX^*_{2i,j}|\left(\big(o_p(n^{-\frac{1}{2}})+O_p(L_n)+O_p(L_{1n}/h_2)\big)O_p(L_{2n})\right)$$
$$\equiv \sum_{k=1}^{3}B_{31k}+o_p(n^{-1/2}) \quad \text{by A5},$$

where
$$B_{311} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_iX^*_{2i,j}}{h_2^{D_2}f_U(U_i)}K_{2ti}C_{g1ti}, \qquad B_{312} = -\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_iX^*_{2i,j}}{h_2^{D_2+1}f_U(U_i)}\mathbf{J}K_{2ti}\big(\hat{U}_i-U_i\big)C_{g1ti},$$
$$B_{313} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_iX^*_{2i,j}}{h_2^{D_2+1}f_U(U_i)}\mathbf{J}K_{2ti}\big(\hat{U}_t-U_t\big)C_{g1ti}.$$

We show that $B_{311}, B_{313}=o_p(n^{-1/2})$ and $B_{312}=\frac{1}{n}\sum_{i=1}^{n}a_{1ni,j}+o_p(n^{-1/2})$, where

$$a_{1ni,j} = \sum_{d=1}^{D_2}\frac{U_{id}}{h_1^{D_1}h_2^{D_2}}\mathrm{E}\left(\frac{\eta_lX^*_{2l,j}\mathrm{D}_dK_{2tl}K_{1il}}{f_U(U_l)f_Z(Z_l)}\mathbf{J}g_1(U_l)\Big(\frac{U_t-U_l}{h_2}\Big)\middle| Z_i\right).$$

B_{311} is of similar structure as B_{141} with U_i replacing W_i, $\eta_iX^*_{2i,j}$ replacing η_iv_i, C_{g1ti} replacing $C_{m2ti,j}$, and $\mathrm{E}(\eta_iX^*_{2i,j}|U_i)=0$ replacing $E(\eta_iv_i|W_i)=0$. By the same arguments in $1.1-1.3$, we have $B_{311}=o_p(n^{-1/2})$. Given the three components in C_{g1ti}, let $-B_{312}\equiv\sum_{k=1}^{3}B_{312k}$, with

$$B_{3121} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_iX^*_{2i,j}}{h_2^{D_2+1}f_U(U_i)}\mathbf{J}K_{2ti}\big(\hat{U}_i-U_i\big)(\hat{\eta}_t-\eta_t)Y_t, \qquad B_{3122} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_iX^*_{2i,j}}{h_2^{D_2+1}f_U(U_i)}\mathbf{J}K_{2ti}\big(\hat{U}_i-U_i\big)v_{g1t},$$

$$B_{3123} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i X^*_{2i,j}}{h_2^{D_2+1}f_U(U_i)}\mathbf{J}K_{2ti}\big(\hat{U}_i - U_i\big)\big(g_1(U_t) - g_1(U_i)\big).$$

We show that $B_{3121}, B_{3122} = o_p(n^{-1/2})$, and $B_{3123} = \frac{1}{n}\sum_{i=1}^{n} a_{1ni,j} + o_p(n^{-1/2})$.

4.1. Given $\hat{\eta}_t - \eta_t = O_p(L_n)$ and $\hat{U}_i - U_i = O_p(L_{1n})$ uniformly, by Markov's Inequality and A5, we have

$$B_{3121} = O_p(L_n)O_p\left(\frac{L_{1n}}{h_2}\right)\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{d=1}^{D_2}\frac{\left|\eta_i X^*_{2i,j}\eta_t Y_t \mathrm{D}_d K_{2ti}\right|}{h_2^{D_2}f_U(U_i)} = o_p(n^{-1/2}).$$

4.2. By (A.1) in the proof of Theorem 2, we have

$$\begin{aligned}
B_{3122} &= \sum_{d=1}^{D_2}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i X^*_{2i,j} v_{g1t}\mathrm{D}_d K_{2ti}}{h_2^{D_2+1}f_U(U_i)}(\hat{U}_{id} - U_{id})\\
&= -\sum_{d=1}^{D_2}\Bigg\{\frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\frac{\eta_i X^*_{2i,j} v_{g1t}\mathrm{D}_d K_{2ti}K_{1li}}{h_1^{D_1}h_2^{D_2+1}f_U(U_i)f_Z(Z_i)}\Big(U_{ld} + \big(\Pi_d(Z_l) - \Pi_d(Z_i)\big)\Big)\\
&\quad + O_p(L_{1n}^2)\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\left|\frac{\eta_i X^*_{2i,j} v_{g1t}\mathrm{D}_d K_{2ti}}{h_2^{D_2+1}f_U(U_i)}\right|\Bigg\} \equiv -\sum_{d=1}^{D_2}(T_{1d} + T_{2d}) + o_p(n^{-1/2}).
\end{aligned}$$

We have $B_{3122} = o_p(n^{-1/2})$ by showing that T_{1d} and T_{2d} are $o_p(n^{-1/2})$ in OA 3.7.

4.3.

$$\begin{aligned}
B_{3123} &= \sum_{d=1}^{D_2}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i X^*_{2i,j}\big(g_1(U_t) - g_1(U_i)\big)\mathrm{D}_d K_{2ti}}{h_2^{D_2+1}f_U(U_i)}(\hat{U}_{id} - U_{id})\\
&= -\sum_{d=1}^{D_2}\Bigg\{\frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\frac{\eta_i X^*_{2i,j}\big(g_1(U_t) - g_1(U_i)\big)\mathrm{D}_d K_{2ti}K_{1li}}{h_1^{D_1}h_2^{D_2+1}f_U(U_i)f_Z(Z_i)}\Big(U_{ld} + \big(\Pi_d(Z_l) - \Pi_d(Z_i)\big)\Big)\\
&\quad + O_p(L_{1n}^2)\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\left|\frac{\eta_i X^*_{2i,j}\big(g_1(U_t) - g_1(U_i)\big)\mathrm{D}_d K_{2ti}}{h_2^{D_2+1}f_U(U_i)}\right|\Bigg\} \equiv -\sum_{d=1}^{D_2}(W_{1d} + W_{2d}) + o_p(n^{-1/2}).
\end{aligned}$$

We show that $\sum_{d=1}^{D_2} W_{1d} = \frac{1}{n}\sum_{i=1}^{n} a_{1ni,j} + o_p(n^{-1/2})$ and $W_{2d} = o_p(n^{-1/2})$ in OA 3.8 and OA 3.9 respectively, where

$$\begin{aligned}
W_{1d} &= \frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\frac{\eta_i X^*_{2i,j}\big(g_1(U_t) - g_1(U_i)\big)\mathrm{D}_d K_{2ti}K_{1li}}{h_1^{D_1}h_2^{D_2+1}f_U(U_i)f_Z(Z_i)}U_{ld},\\
W_{2d} &= \frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\frac{\eta_i X^*_{2i,j}\mathrm{D}_d K_{2ti}K_{1li}}{h_1^{D_1}h_2^{D_2+1}f_U(U_i)f_Z(Z_i)}\big(g_1(U_t) - g_1(U_i)\big)\big(\Pi_d(Z_l) - \Pi_d(Z_i)\big).
\end{aligned}$$

By 4.1–4.3, we have $B_{312} = \frac{1}{n}\sum_{i=1}^{n} a_{1ni,j} + o_p(n^{-1/2})$. For B_{313}, the analysis is exactly similar to B_{312}, but note that for the term having order $O_p(n^{-1/2})$ in B_{3123}, the corresponding term in B_{3133}, denoted as W'_{1d}, is

$$W'_{1d} = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\frac{\eta_i X^*_{2i,j}\big(g_1(U_t) - g_1(U_i)\big)\mathrm{D}_d K_{2ti}K_{1lt}}{h_1^{D_1}h_2^{D_2+1}f_U(U_i)f_Z(Z_t)}U_{ld}.$$

The difference here is we have Z_t instead of Z_i, such that $\mathrm{E}(\psi_{nitl}|P_l) = 0$ in that $\mathrm{E}(\eta_i X^*_{2i,j}|U_i) = 0$. Thus, by the same arguments for the rest of terms, we have $B_{313} = o_p(n^{-1/2})$.

As to B_{32}, the analysis is similar to B_{31} given above. For the component with order $O_p(n^{-1/2})$, we can actually combine that in B_{31} and the one in B_{32} together to have a more intuitive result. Note that

$$\begin{aligned}
V_{g1i} - V_{g2i}\beta &= \Bigg\{\frac{1}{nh_2^{D_2}f_U(U_i)}\sum_{t=1}^{n}\hat{K}_{2ti}\Big[(\hat{\eta} - \eta_t)(Y_t - X_{2t}\beta) + (v_{g1t} - v_{g2t}\beta)\\
&\quad + \Big(\big(g_1(U_t) - g_1(U_i)\big) - \big(g_2(U_t) - g_2(U_i)\big)\beta\Big)\Big]\Bigg\}\Big(1 + O_p(L_{2n})\Big),
\end{aligned}$$

and the component of order $O_p(n^{-1/2})$ involves the third term in brackets, which is $\big(g_1(U_t) - g_2(U_t)\beta - \beta_0\big) - \big(g_1(U_i) - g_2(U_i)\beta - \beta_0\big) = g(U_t) - g(U_i)$. Thus using $\big(g(U_t) - g(U_i)\big)$ instead of $\big(g_1(U_t) - g_1(U_i)\big)$ in W_{1d}, we have

$$B_3 = -\frac{1}{n}\sum_{i=1}^{n} a_{ni} + o_p(n^{-1/2}),$$

$$\text{where}\quad a_{ni} = \sum_{d=1}^{D_2} \frac{U_{id}}{h_1^{D_1}h_2^{D_2}} \mathrm{E}\left(\frac{\eta_l X_{2l}^{*}\mathrm{D}_d K_{2tl}K_{1il}}{f_U(U_l)f_Z(Z_l)}\mathbf{J}g(U_l)\Big(\frac{U_t-U_l}{h_2}\Big)\middle| Z_i\right).$$

Step 5: Combining orders of B_1, B_2, B_3, B_4, we have $\frac{1}{n}\hat{\boldsymbol{X}}_2'\hat{\boldsymbol{\eta}}(\hat{\boldsymbol{Y}} - \hat{\boldsymbol{X}}_2\beta) = B_{11} - \frac{1}{n}\sum_{i=1}^{n} a_{ni} + o_p(n^{-1/2})$. Next we investigate $\sqrt{n}(B_{11} - \frac{1}{n}\sum_{i=1}^{n} a_{ni})$.

Let $\lambda \in \mathbb{R}^{D_2}$ be a non-stochastic vector such that $\lambda'\lambda = 1$. Denote $B_{11} + \frac{1}{n}\sum_{i=1}^{n} a_{ni} = \frac{1}{n}\sum_{i=1}^{n}(X_{2i}^{*}\eta_i v_i + a_{ni}) \equiv \frac{1}{n}\sum_{i=1}^{n} b_{ni}$, and we have $\mathrm{E}(\lambda' b_{ni}) = 0$ as $\mathrm{E}(X_{2i}^{*}\eta_i v_i)$, $\mathrm{E}(a_{ni}) = 0$, and $\mathrm{E}(\lambda' b_{ni} b_{ni}'\lambda) = \lambda'\mathrm{E}(X_{2i}^{*}\eta_i^2 v_i^2 X_{2i}^{*\prime})\lambda + \lambda'\mathrm{E}(a_{ni}a_{ni}')\lambda = \lambda'\Phi_1\lambda + \lambda'\mathrm{E}(a_{ni}a_{ni}')\lambda$. Denote $X_{2i,j} = \Pi_{2j}(Z_i) + U_{2i,j}$, the jth element of a_{ni} can be written as

$$\begin{aligned}
a_{ni,j} &= \sum_{d=1}^{D_2} \frac{U_{id}}{h_1^{D_1}h_2^{D_2}} \mathrm{E}\left(\frac{\eta_l X_{2l,j}^{*}\mathrm{D}_d K_{2tl}K_{1il}}{f_U(U_l)f_Z(Z_l)}\mathbf{J}g(U_l)\Big(\frac{U_t-U_l}{h_2}\Big)\middle| Z_i\right)\\
&= \int \frac{1}{h_1^{D_1}h_2^{D_2}}\Big(\Pi_{2j}(Z_l) + U_{2l,j} - m_{2j}(W_l) - g_{2j}(U_l) + \mu_{2j}\Big)\sum_{d=1}^{D_2} U_{id}\mathrm{D}_d K_{2tl}K_{1il}\mathbf{J}g(U_l)\left(\frac{U_t-U_l}{h_2}\right)\\
&\quad \times \frac{\eta(W_l,U_l)}{f_U(U_l)f_Z(Z_l)} f_U(U_t) f_{ZUW}(Z_l,U_l,W_l)\,\mathrm{d}U_t\,\mathrm{d}Z_l\,\mathrm{d}U_l\,\mathrm{d}W_l\\
&= \int \Big(\Pi_{2j}(Z_i - h_1\gamma) + U_{2t,j} - h_2\psi_{2j} - m_{2j}(W_l) - g_{2j}(U_t - h_2\psi) + \mu_{2j}\Big)\sum_{d=1}^{D_2} U_{id}\mathrm{D}_d K_2(\psi)K_1(\gamma)\\
&\quad \times \mathbf{J}g(U_t - h_2\psi)\psi\,\frac{\eta(W_l, U_t - h_2\psi)}{f_U(U_t - h_2\psi)f_Z(Z_i - h_1\gamma)} f_U(U_t) f_{ZUW}(Z_i - h_1\gamma, U_t - h_2\psi, W_l)\,\mathrm{d}\gamma\,\mathrm{d}\psi\,\mathrm{d}U_t\,\mathrm{d}W_l\\
&\to \int \Big(\Pi_{2j}(Z_i) + U_{2t,j} - m_{2j}(W_l) - g_{2j}(U_t) + \mu_{2j}\Big)\sum_{d=1}^{D_2} U_{id}\,(-\mathrm{D}_d g(U_t))\,\eta(W_l,U_t) f_{UW|Z}(U_t, W_l|Z_i)\,\mathrm{d}U_i\,\mathrm{d}W_i\\
&= -\sum_{d=1}^{D_2}\mathrm{E}\left(\Big(\Pi_{2j}(Z_i) + U_{2t,j} - m_{2j}(W_t) - g_{2j}(U_t) + \mu_{2j}\Big)\mathrm{D}_d g(U_t)\eta_t\middle| Z_i\right)U_{id}\\
&= -\sum_{d=1}^{D_2}\mathrm{E}\left(\Big(\Pi_{2j}(Z_i) - \Pi_{2j}(Z_t)\Big)\mathrm{D}_d g(U_t)\eta_t\middle| Z_i\right)U_{id}.
\end{aligned}$$

The convergence follows by A3, and that $\int \mathrm{D}_d K_2(\psi)\psi d\psi = (0,\ldots,-1,\ldots,0)'$, where -1 appears on the dth position of the vector. The last equation follows by $\mathrm{E}(\eta_t X_{2t}^{*}|U_t) = 0$. Hence, the (j,k)th element of $\mathrm{E}(a_{ni}a_{ni}')$ converges to

$$\Phi_{2(j,k)} \equiv \mathrm{E}\left[\sum_{d=1}^{D_2}\sum_{\delta=1}^{D_2} \mathrm{E}\left(\Big(\Pi_{2j}(Z_i) - \Pi_{2j}(Z_t)\Big)\mathrm{D}_d g(U_t)\eta_t\middle| Z_i\right)\mathrm{E}\left(\Big(\Pi_{2k}(Z_i) - \Pi_{2k}(Z_t)\Big)\mathrm{D}_\delta g(U_t)\eta_t\middle| Z_i\right)U_{id}U_{i\delta}\right].$$

By Lyapunov's Central Limit Theorem, we have $\sqrt{n}\big(B_{11} - \frac{1}{n}\sum_{i=1}^{n} a_{ni}\big) \xrightarrow{d} \mathcal{N}(0, \Phi_1 + \Phi_2)$, provided $\lim_{n\to\infty}\sum_{i=1}^{n}\mathrm{E}\big|n^{-1/2}\lambda' a_{ni}\big|^{2+\delta} = 0$ for some $\delta > 0$. Note that by C_r Inequality,

$$\sum_{i=1}^{n}\mathrm{E}\big|n^{-1/2}\lambda' a_{ni}\big|^{2+\delta} = n^{-\delta/2}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left|\sum_{j=1}^{D_{22}}\lambda_j a_{ni,j}\right|^{2+\delta} \le n^{-\delta/2} D_{22}^{1+\delta}\sum_{j=1}^{D_{22}}\lambda_j^{2+\delta}\mathrm{E}|a_{ni,j}|^{2+\delta},$$

$$\begin{aligned}
\text{where}\quad \mathrm{E}|a_{ni,j}|^{2+\delta} &\to \int\left|\sum_{d=1}^{D_2}\mathrm{E}\Big(\big(\Pi_{2j}(Z_i) + U_{2t,j} - m_{2j}(W_t) - g_{2j}(U_t) + \mu_{2j}\big)\mathrm{D}_d g(U_t)\eta_t\Big| Z_i\Big)\right|^{2+\delta}\\
&\quad \times |U_{id}|^{2+\delta} f_{ZU}(Z_i,U_i)\,\mathrm{d}Z_i\,\mathrm{d}U_i
\end{aligned}$$

$$
\begin{aligned}
&\le C\sum_{d=1}^{D_2}\int \left|\mathrm{E}\Big(\big(\Pi_{2j}(Z_i)+U_{2t,j}-m_{2j}(W_t)-g_{2j}(U_t)+\mu_{2j}\big)\Big|Z_i\Big)\right|^{2+\delta}|U_{id}|^{2+\delta}f_{ZU}(Z_i,U_i)\,\mathrm{d}Z_i\,\mathrm{d}U_i\\
&<\infty \quad \text{since} \quad \mathrm{E}\big(|U_{id}|^{2+\delta}|Z_i\big)<C<\infty \quad \text{and} \quad \mathrm{E}|X_{2i,j}|^{2+\delta}<\infty.
\end{aligned}
$$

Thus $\lim_{n\to\infty}\sum_{i=1}^{n}\mathrm{E}\big|n^{-1/2}\lambda' a_{ni}\big|^{2+\delta}=0$ for some $\delta>0$, and we have $\frac{1}{n}\hat{\boldsymbol{X}}_2'\hat{\boldsymbol{\eta}}(\hat{\boldsymbol{Y}}-\hat{\boldsymbol{X}}_2\beta)\xrightarrow{d}\mathcal{N}(0,\Phi_1+\Phi_2)$. From Step 1, we have $\left(\frac{1}{n}\hat{\boldsymbol{X}}_2'\hat{\boldsymbol{\eta}}\hat{\boldsymbol{X}}_2\right)^{-1}\xrightarrow{p}\Phi_0^{-1}$. All together, we have $\sqrt{n}(\hat\beta-\beta)\xrightarrow{d}\mathcal{N}\big(0,\Phi_0^{-1}(\Phi_1+\Phi_2)\Phi_0^{-1}\big)$. $\square$

Proof of Theorem 4. $\hat m(w)$ can be written as $\hat m(w)=\hat m_1(w)-\hat m_2'(w)\hat\beta-\hat m_3(w)\hat\beta_0$, where

$$
\hat m_3(w)\equiv(nh_3^{D_3}\hat f_W(w))^{-1}\sum_{t=1}^{n}K_3\left(\frac{W_t-w}{h_3}\right)\hat\eta_t.
$$

Then by Eq. (4), we have

$$
\begin{aligned}
\hat m(w)-m(w)=&\big(\hat m_1(w)-m_1(w)\big)-\big(\hat m_2(w)-m_2(w)\big)'\beta-\big(\hat m_3(w)-1\big)\beta_0\\
&-\big(\hat m_2(w)-m_2(w)\big)'\big(\hat\beta-\beta\big)-\big(\hat m_3(w)-1\big)\big(\hat\beta_0-\beta_0\big)-m_2(w)'\big(\hat\beta-\beta\big)-\big(\hat\beta_0-\beta_0\big).
\end{aligned}
$$

Since, by Theorems 2 and 3, $\hat\beta_0-\beta_0=O_p(n^{-1/2})$, $\hat\beta-\beta=O_p(n^{-1/2})$, and $\hat m_2(w)-m_2(w)=o_p(1)$, the last four terms in $\hat m(w)-m(w)$ when multiplied by $(nh_3^{D_3})^{1/2}$ are $o_p(1)$. Thus,

$$
\sqrt{nh_3^{D_3}}\Big(\hat m(w)-m(w)\Big)=\sqrt{nh_3^{D_3}}\Big((\hat m_1(w)-m_1(w))-\big(\hat m_2(w)-m_2(w)\big)'\beta-\big(\hat m_3(w)-1\big)\beta_0\Big)+o_p(1).
$$

We first investigate $\sqrt{nh_3^{D_3}}\big(\hat m_1(w)-m_1(w)\big)$, and then the asymptotic distribution of $\hat m(w)$ follows immediately due to the similar structure of $\hat m(w)$ and $\hat m_1(w)$. Given the expressions for $\hat m_1(w)$ and $\hat f_W(w)$, and the uniform order of $\hat f_W(w)$, letting $K_{3t}\equiv K_3\left(\frac{W_t-w}{h_3}\right)$, we have

$$
\begin{aligned}
\hat m_1(w)-m_1(w)&=\left\{\frac{1}{nh_3^{D_3}f_W(w)}\sum_{t=1}^{n}K_{3t}\Big(\big(m_1(W_t)-m_1(w)\big)+\big(\eta_tY_t-m_1(W_t)\big)+(\hat\eta_t-\eta_t)Y_t\Big)\right\}\Big(1+O_p(L_{3n})\Big)\\
&\equiv\left\{\sum_{k=1}^{3}T_k\right\}\Big(1+O_p(L_{3n})\Big).
\end{aligned}
$$

The proof has four steps:

(1) We show that $T_1=b_{m1,1}(w)$, where $b_{m1,1}(w)\equiv h_3^{s_3}\frac{\mu_{k_3,s_3}}{f_W(w)}\sum_{k=1}^{s_3}\frac{1}{k!(s_3-k)!}\sum_{j=1}^{D_3}\mathrm{D}_j^k m_1(w)\mathrm{D}_j^{s_3-k}f_W(w)+o_p(h_3^{s_3})$.

(2) We show that $\sqrt{nh_3^{D_3}}T_2\xrightarrow{d}\mathcal{N}(0,\Phi_{m1,1})$, where $\Phi_{m1,1}\equiv\frac{\sigma_{vm1}^2}{f_W(w)}\int K_3^2(\gamma)\,\mathrm{d}\gamma$.

(3) We show that $\sqrt{nh_3^{D_3}}\big(T_3-b_{m1,2}(w)\big)\xrightarrow{d}\mathcal{N}(0,\Phi_{m1,2})$, where $b_{m1,2}(w)\equiv h_3^{s_3}\frac{\mu_{k_3,s_3}}{f_W(w)}\frac{1}{s_3!}\sum_{j=1}^{D_3}m_1(w)\mathrm{D}_j^{s_3}f_W(w)+o_p(h_3^{s_3})$, $\Phi_{m1,2}\equiv\frac{m_1^2(w)}{f_W(w)}\int\left(\int K_3(\gamma_1)K_3(\gamma_1+\gamma_2)\,\mathrm{d}\gamma_1\right)^2\mathrm{d}\gamma_2$.

(4) Combining (1)-(3), we show that $\sqrt{nh_3^{D_3}}\big(\hat m_1(w)-m_1(w)-b_{m1}(w)\big)\xrightarrow{d}\mathcal{N}(0,\Phi_{m1,1}+\Phi_{m1,2})$, where $b_{m1}(w)=b_{m1,1}(w)+b_{m1,2}(w)$.

Step 1: By Taylor's Theorem, we have

$$
\begin{aligned}
T_1&=\frac{1}{nh_3^{D_3}f_W(w)}\sum_{t=1}^{n}K_{3t}\big(m_1(W_t)-m_1(w)\big)\\
&=\frac{1}{nh_3^{D_3}f_W(w)}\sum_{t=1}^{n}K_{3t}\left(\sum_{|\beta|=1}^{s_3}\frac{1}{\beta!}\mathrm{D}^\beta m_1(w)(W_t-w)^\beta+\sum_{|\beta|=s_3+1}\frac{1}{\beta!}\mathrm{D}^\beta m_1(\tilde w)(W_t-w)^\beta\right)\equiv\sum_{|\beta|=1}^{s_3+1}T_{1,|\beta|},
\end{aligned}
$$

where $\tilde w\equiv w+\lambda(W_t-w)$, for some $\lambda\in(0,1)$. For each $|\beta|=1,\ldots,s_3$, we rewrite $T_{1,|\beta|}$ as

$$
T_{1,|\beta|}=\frac{h_3^{|\beta|}}{|\beta|!f_W(w)}\sum_{|\beta|}\mathrm{D}^\beta m_1(w)t_{|\beta|},\quad\text{where}\quad t_{|\beta|}\equiv\frac{1}{nh_3^{D_3}}\sum_{t=1}^{n}K_{3t}\left(\frac{W_t-w}{h_3}\right)^\beta.
$$

Using Lemma 3 and the assumptions on the statement of the theorem, $\sup_{w\in\mathcal{G}_W}|t_{|\beta|}-\mathrm{E}(t_{|\beta|})|=O_p\left(\left(\log n/nh_3^{D_3}\right)^{1/2}\right)$. If $|\beta|=1$, by Taylor's Theorem and given that k_3 is of order s_3, we have

$$
\begin{aligned}
\mathrm{E}(t_1) &= \sum_{|\beta|=1}\int K_3(\gamma)\gamma^\beta f_W(w+h_3\gamma)\,\mathrm{d}\gamma \\
&= \sum_{|\beta|=1}\int K_3(\gamma)\gamma^\beta\left(f_W(w)+\sum_{|\alpha|=1}^{s_3-1}\frac{1}{\alpha!}\mathrm{D}^\alpha f_W(w)(h_3\gamma)^\alpha+\sum_{|\alpha|=s_3}\frac{1}{\alpha!}\mathrm{D}^\alpha f_W(\tilde{w})(h_3\gamma)^\alpha\right)\,\mathrm{d}\gamma \\
&= h_3^{s_3-1}\frac{\mu_{k_3,s_3}}{(s_3-1)!}\sum_{j=1}^{D_3}\mathrm{D}_j^{s_3-1}f_W(w)+o\left(h_3^{s_3-1}\right).
\end{aligned}
$$

Thus, given that $h_3=n^{-1/(2s_3+D_3)}$, we have $h_3(\log n/nh_3^{D_3})^{1/2}=o(h_3^{s_3})$, and

$$
T_{11}=h_3^{s_3}\frac{\mu_{k_3,s_3}}{f_W(w)}\frac{1}{(s_3-1)!}\sum_{j=1}^{D_3}\mathrm{D}_j m_1(w)\mathrm{D}_j^{s_3-1}f_W(w)+o_p(h_3^{D_3}).
$$

Similarly, if $|\beta|=2$ and 2 is in the jth position of the vector β, 0 elsewhere, we have $\mathrm{E}(t_{|\beta|})=h_3^{s_3-2}\frac{\mu_{k_3,s_3}}{(s_3-2)!}\mathrm{D}_j^{s_3-2}f_W(w)+o\left(h_3^{s_3-2}\right)$. And for any remaining β such that $|\beta|=2$, $\mathrm{E}(t_{|\beta|})=o\left(h_3^{s_3-2}\right)$. Thus,

$$
T_{12}=h_3^{s_3}\frac{\mu_{k_3,s_3}}{f_W(w)}\frac{1}{2!(s_3-2)!}\sum_{j=1}^{D_3}\mathrm{D}_j^2 m_1(w)\mathrm{D}_j^{s_3-2}f_W(w)+o_p(h_3^{D_3}).
$$

In a similar manner, we have,

$$
T_{1,|\beta|}=h_3^{s_3}\frac{\mu_{k_3,s_3}}{f_W(w)}\frac{1}{|\beta|!(s_3-|\beta|)!}\sum_{j=1}^{D_3}\mathrm{D}_j^{|\beta|}m_1(w)\mathrm{D}_j^{s_3-|\beta|}f_W(w)+o_p(h_3^{D_3}),\quad \text{for any } |\beta|=1,\dots,s_3.
$$

For $|\beta|=s_3+1$, we have $T_{1,(s_3+1)}=\frac{h_3^{s_3+1}}{nh_3^{D_3}f_W(w)}\sum_{t=1}^n K_{3t}\left(\sum_{|\beta|=s_3+1}\frac{1}{(s_3+1)!}\mathrm{D}^\beta m_1(\tilde{w})\left(\frac{W_t-w}{h_3}\right)^\beta\right)=o_p(h_3^{s_3})$, by Markov's Inequality and $\mathrm{E}|T_{1,(s_3+1)}|=O(h_3^{s_3+1})=o(h_3^{s_3})$ since $m_1(w)\in C^{s_3+1}$. Combining all the $T_{1,|\beta|}$ terms, we have

$$
T_1=b_{m1,1}(w),\quad \text{where}\ \ b_{m1,1}(w)\equiv h_3^{s_3}\frac{\mu_{k_3,s_3}}{f_W(w)}\sum_{k=1}^{s_3}\frac{1}{k!(s_3-k)!}\sum_{j=1}^{D_3}\mathrm{D}_j^k m_1(w)\mathrm{D}_j^{s_3-k}f_W(w)+o_p(h_3^{s_3}).
$$

Step 2: Given $\eta_t Y_t=m_1(W_t)+v_{m1t}$, we have $T_2=\sum_{t=1}^n a_{1tn}$, where $a_{1tn}\equiv(nh_3^{D_3}f_W(w))^{-1}K_{3t}v_{m1t}$. Since $\mathrm{E}(v_{m1t}|W_t)=0$ and $\mathrm{E}(v_{m1t}^2|W_t)=\sigma_{vm1}^2<\infty$, we have $\mathrm{E}(a_{1tn})=0$, and $\mathrm{V}(a_{1tn})=n^{-2}h_3^{-D_3}f_W^{-2}(w)\sigma_{vm1}^2\int K_3^2(\gamma)f_W(w+h_3\gamma)\,\mathrm{d}\gamma$. Let $S_{1n}^2\equiv\sum_{t=1}^n\mathrm{V}(a_{1tn})=(nh_3^{D_3})^{-1}f_W^{-2}(w)\sigma_{vm1}^2\int K_3^2(\gamma)f_W(w+h_3\gamma)\,\mathrm{d}\gamma$. Then, by Lyapunov's CLT, if $\sum_{t=1}^n\mathrm{E}|a_{1tn}/S_{1n}|^{2+\delta}\to 0$ for some $\delta>0$ as $n\to\infty$, we have $\sum_{t=1}^n a_{1tn}/S_{1n}\xrightarrow{d}\mathcal{N}(0,1)$, i.e., given $\sqrt{nh_3^{D_3}}S_{1n}\to\Phi_{m1,1}^{1/2}$,

$$
\sqrt{nh_3^{D_3}}T_2\xrightarrow{d}\mathcal{N}(0,\Phi_{m1,1}),\quad \text{where}\ \ \Phi_{m1,1}\equiv\frac{\sigma_{vm1}^2}{f_W(w)}\int K_3^2(\gamma)\,\mathrm{d}\gamma.
$$

Given that $nh_3^{D_3}S_{1n}^2\to\Phi_{m1,1}>0$ and $\mathrm{E}(|v_{m1t}|^{2+\delta}|W_t)<C$, Lyapunov's condition is satisfied since

$$
\sum_{t=1}^n\mathrm{E}\left|\frac{a_{1tn}}{S_{1n}}\right|^{2+\delta}=\frac{\left(nh_3^{D_3}\right)^{\delta/2+1}}{\left(nh_3^{D_3}S_{1n}^2\right)^{\delta/2+1}}\sum_{t=1}^n\mathrm{E}\left(\left|\frac{K_{3t}v_{m1t}}{nh_3^{D_3}f_W(w)}\right|^{2+\delta}\right)\le C\left(nh_3^{D_3}\right)^{-\delta/2}\int|K_3(\gamma)|^{2+\delta}\,\mathrm{d}\gamma\to 0,\ \text{as } n\to\infty.
$$

Step 3: Denote $\hat{f}_{\hat{U}}(\hat{U}_t)=\hat{f}_{\hat{U}_t}$, $\hat{f}_W(W_t)=\hat{f}_{W_t}$, $\hat{f}_{W\hat{U}}(W_t,\hat{U}_t)=\hat{f}_{W_t\hat{U}_t}$, $f_U(U_t)=f_{U_t}$, $f_W(W_t)=f_{W_t}$, $f_{WU}(W_t,U_t)=f_{W_tU_t}$. According to the uniform order of these density estimators from Theorem 1 and L_n^2, $(L_{1n}/h_2)^2=o(n^{-1/2})$ by A5, we have

$$
\hat{\eta}_t-\eta_t=\frac{1}{f_{W_tU_t}^2}\Big(f_{W_tU_t}f_{W_t}(\hat{f}_{\hat{U}_t}-f_{U_t})-f_{U_t}f_{W_t}(\hat{f}_{W_t\hat{U}_t}-f_{W_tU_t})+f_{W_tU_t}f_{U_t}(\hat{f}_{W_t}-f_{W_t})\Big)+o_p(n^{-1/2}).
$$

Since $T_3=(nh_3^{D_3}f_W(w))^{-1}\sum_{t=1}^n K_{3t}\big((\hat{\eta}_t-\eta_t)Y_t\big)$, and $(nh_3^{D_3}f_W(w))^{-1}\sum_{t=1}^n\left|K_{3t}Y_t\right|=O_p(1)$, we have

$$
T_3=\sum_{k=1}^3 T_{3k}+o_p\big(n^{-1/2}\big),
$$

where $$T_{31}=\frac{1}{nh_3^{D_3}f_W(w)}\sum_{t=1}^{n}\frac{1}{f_{U_t}}(\hat f_{\hat U_t}-f_{U_t})K_{3t}\eta_tY_t,\quad T_{32}=-\frac{1}{nh_3^{D_3}f_W(w)}\sum_{t=1}^{n}\frac{1}{f_{W_tU_t}}(\hat f_{W_t\hat U_t}-f_{W_tU_t})K_{3t}\eta_tY_t,$$
$$T_{33}=\frac{1}{nh_3^{D_3}f_W(w)}\sum_{t=1}^{n}\frac{1}{f_{W_t}}(\hat f_{W_t}-f_{W_t})K_{3t}\eta_tY_t.$$

From Theorem 1, we have $|\hat f_{\hat U_t}-f_{U_t}|=O_p(L_{2n})$ and $|\hat f_{W_t\hat U_t}-f_{W_tU_t}|=O_p(L_{4n})$ uniformly. Thus, $\sqrt{nh_3^{D_3}}T_{31}=O_p\left(\sqrt{nh_3^{D_3}}L_{2n}\right)$ $=o_p(1)$ by Assumption A5(iii). Similarly, $\sqrt{nh_3^{D_3}}T_{32}=O_p\left(\sqrt{nh_3^{D_3}}L_{4n}\right)=o_p(1)$. Let $T_{33}=T_{331}+T_{332}$, where

$$T_{331}=\frac{1}{nh_3^{D_3}f_W(w)}\sum_{t=1}^{n}\frac{1}{f_{W_t}}\big(\mathrm{E}(\hat f_{W_t})-f_{W_t}\big)K_{3t}\eta_tY_t,\qquad T_{332}=\frac{1}{nh_3^{D_3}f_W(w)}\sum_{t=1}^{n}\frac{1}{f_{W_t}}\big(\hat f_{W_t}-\mathrm{E}(\hat f_{W_t})\big)K_{3t}\eta_tY_t.$$

We show that T_{331} contributes to a bias and T_{332} to a normal distribution.

For T_{331}, given that $\mathrm{E}(\hat f_{W_t})-f_{W_t}=h_3^{s_3}\frac{\mu_{k_3,s_3}}{s_3!}\sum_{j=1}^{D_3}\mathrm{D}_j^{s_3}f_W(W_t)+o(h_3^{s_3})$ by Taylor's Theorem and the high order of kernel k_3, we have

$$T_{331}=h_3^{s_3}\frac{\mu_{k_3,s_3}}{s_3!}\sum_{j=1}^{D_3}t_j+o(h_3^{s_3}),\qquad\text{where}\quad t_j=\frac{1}{nh_3^{D_3}f_W(w)}\sum_{t=1}^{n}\frac{1}{f_{W_t}}\mathrm{D}_j^{s_3}f_W(W_t)K_{3t}\eta_tY_t.$$

Since $\eta_tY_t=v_{m1t}+(m_1(W_t)-m_1(w))+m_1(w)$, let $t_j=\sum_{k=1}^{3}t_{jk}$, where

$$t_{j1}=\frac{1}{nh_3^{D_3}f_W(w)}\sum_{t=1}^{n}\frac{1}{f_{W_t}}\mathrm{D}_j^{s_3}f_W(W_t)K_{3t}v_{m1t},\qquad t_{j2}=\frac{1}{nh_3^{D_3}f_W(w)}\sum_{t=1}^{n}\frac{1}{f_{W_t}}\mathrm{D}_j^{s_3}f_W(W_t)K_{3t}\big(m_1(W_t)-m_1(w)\big),$$
$$t_{j3}=\frac{m_1(w)}{nh_3^{D_3}f_W(w)}\sum_{t=1}^{n}\frac{1}{f_{W_t}}\mathrm{D}_j^{s_3}f_W(W_t)K_{3t}.$$

By Markov's Inequality and $\mathrm{E}(t_{j1})=0$, $\mathrm{E}(t_{j1}^2)=O\big((nh_3^{D_3})^{-1}\big)$ due to $\mathrm{E}(v_{m1t}|W_t)=0$ and $\mathrm{E}(v_{m1t}^2|W_t)\le C$, we have $t_{j1}=O_p\big((nh_3^{D_3})^{-1/2}\big)=o_p(1)$. And $t_{j2}=O_p(h_3)=o_p(1)$ since $\mathrm{E}|t_{j2}|\le Ch_3^{-D_3}\mathrm{E}\big|K_{3t}\big(m_1(W_t)-m_1(w)\big)\big|=O(h_3)$. For t_{j3}, since $\mathrm{E}(t_{j3})=m_1(w)f_W(w)^{-1}\int\mathrm{D}_j^{s_3}f_W(w+h_3\phi)K_3(\phi)\,\mathrm{d}\phi\to m_1(w)f_W^{-1}(w)\mathrm{D}_j^{s_3}f_W(w)$, and $\mathrm{E}(t_{j3}^2)=O\big((nh_3^{D_3})^{-1}\big)=o(1)$, we have $t_{j3}=m_1(w)f_W^{-1}(w)\mathrm{D}_j^{s_3}f_W(w)+o_p(1)$. In sum, $T_{331}=h_3^{s_3}\frac{\mu_{k_3,s_3}}{f_W(w)}\frac{1}{s_3!}\sum_{j=1}^{D_3}m_1(w)\mathrm{D}_j^{s_3}f_W(w)+o_p(h_3^{s_3})\equiv b_{m1,2}(w)$.

For T_{332}, we show that $(nh_3^{D_3})^{1/2}T_{332}=(nh_3^{D_3})^{1/2}\sum_{t=1}^{n}a_{2tn}+o_p(1)\xrightarrow{d}\mathcal{N}(0,\Phi_{m1,2})$, where

$$a_{2tn}=(nh_3^{2D_3})^{-1}m_1(w)\mathrm{E}\Big(f_{W_i}^{-1}\big(K_{3it}-\mathrm{E}_t(K_{3it})\big)K_{3i}\Big|W_t\Big),\quad \Phi_{m1,2}\equiv\frac{m_1^2(w)}{f_W(w)}\int\Big(\int K_3(\gamma_1)K_3(\gamma_1+\gamma_2)\,\mathrm{d}\gamma_1\Big)^2\mathrm{d}\gamma_2.$$

Since $\hat f_{W_t}-\mathrm{E}(\hat f_{W_t})=(nh_3^{D_3})^{-1}\sum_{i=1}^{n}\big(K_{3ti}-\mathrm{E}_i(K_{3ti})\big)$, we have $T_{332}=T_{3321}+T_{3322}$, where

$$T_{3321}=\frac{1}{n^2h_3^{2D_3}f_W(w)}\sum_{t=1}^{n}\frac{1}{f_{W_t}}\big(K_3(0)-\mathrm{E}_i(K_{3ti})\big)K_{3t}\eta_tY_t,\quad T_{3322}=\frac{1}{n^2h_3^{2D_3}f_W(w)}\sum_{\substack{t=1\\t\ne i}}^{n}\sum_{i=1}^{n}\frac{1}{f_{W_t}}\big(K_{3ti}-\mathrm{E}_i(K_{3ti})\big)K_{3t}\eta_tY_t.$$

Since $\mathrm{E}_i(K_{3ti})=O(h_3^{D_3})$, we have $T_{3321}\le C(nh_3^{D_3})^{-2}\sum_{t=1}^{n}|K_{3t}\eta_tY_t|=O_p\big((nh_3^{D_3})^{-1}\big)$, thus $(nh_3^{D_3})^{1/2}T_{3321}=o_p(1)$.

For T_{3322}, we have $T_{3322}=\frac{1}{n^2}\binom{n}{2}U_n=\frac{n-1}{n}\frac{1}{2}U_n$, where $U_n\equiv\binom{n}{2}^{-1}\sum_{\substack{t=1\\t<i}}^{n}\sum_{i=1}^{n}\phi_{nti}=\theta_n+2H_n^{(1)}+H_n^{(2)}$, $\phi_{nti}=\psi_{nti}+\psi_{nit}$, and $\psi_{nti}=(h_3^{2D_3}f_{W_t}f_W(w))^{-1}\big(K_{3ti}-\mathrm{E}_i(K_{3ti})\big)K_{3t}\eta_tY_t$. Then $\theta_n=\mathrm{E}(\phi_{nti})=0$, $\sigma_{2n}^2=\mathrm{V}(\phi_{nti})\le C\mathrm{E}(\psi_{nti}^2)=O(h_3^{-2D_3})$, $H_n^{(2)}=O_p\big((\sigma_{2n}^2/n^2)^{1/2}\big)=O_p\big((nh_3^{D_3})^{-1}\big)$, and we have $(nh_3^{D_3})^{1/2}H_n^{(2)}=o_p(1)$. For $H_n^{(1)}=n^{-1}\sum_{t=1}^{n}\mathrm{E}(\psi_{nit}|W_t)$, given that $\mathrm{E}(\eta_iY_i|W_i)=m_1(W_i)$, we have $H_n^{(1)}=Q_1+Q_2$, where

$$Q_1\equiv\frac{1}{nh_3^{2D_3}}\sum_{t=1}^{n}\mathrm{E}\Big(\frac{1}{f_{W_i}f_W(w)}\big(K_{3it}-\mathrm{E}_t(K_{3it})\big)K_{3i}\big(m_1(W_i)-m_1(w)\big)\Big|W_t\Big),$$
$$Q_2\equiv\frac{m_1(w)}{nh_3^{2D_3}}\sum_{t=1}^{n}\mathrm{E}\Big(\frac{1}{f_{W_i}f_W(w)}\big(K_{3it}-\mathrm{E}_t(K_{3it})\big)K_{3i}\Big|W_t\Big).$$

Since $\mathrm{E}(Q_1)=0$, $\mathrm{E}(Q_1^2)=O\big((nh_3^{D_3})^{-1}h_3\big)$, we have $(nh_3^{D_3})^{1/2}Q_1=O_p(h_3^{1/2})=o_p(1)$. Since $Q_2=\sum_{t=1}^{n}a_{2tn}$, let $Z_{tn}=(nh_3^{2D_3})^{-1}m_1(w)\mathrm{E}_i\big((f_{W_i}f_W(w))^{-1}K_{3it}K_{3i}\big)$, and $\mu_n=(nh_3^{2D_3})^{-1}m_1(w)\mathrm{E}\big((f_{W_i}f_W(w))^{-1}K_{3it}K_{3i}\big)$, so that $a_{2tn}=Z_{tn}-\mu_n$

and $\mathrm{E}(Z_{tn}) = \mu_n$. Then, we have $Z_{tn} = (nh_3^{D_3})^{-1}m_1(w)f_W^{-1}(w)\int K_3(\gamma_1)K_3\left(\frac{W_t-w}{h_3}+\gamma_1\right)\mathrm{d}\gamma_1$, $\mu_n = n^{-1}m_1(w)f_W^{-1}(w)\int K_3(\gamma_1)K_3(\gamma_1+\gamma_2)\,\mathrm{d}\gamma_1\,\mathrm{d}\gamma_2 = O(n^{-1})$, and $\mathrm{V}(a_{2tn}) = \mathrm{E}(Z_{tn}^2)-\mu_n^2 = n^{-2}h_3^{-D_3}m_1^2(w)f_W^{-2}(w)\int\left(\int K_3(\gamma_1)K_3(\gamma_1+\gamma_2)\,\mathrm{d}\gamma_1\right)^2 f_W(w+h_3\gamma_2)\,\mathrm{d}\gamma_2-\mu_n^2$. Letting $S_{2n}^2 \equiv \sum_{t=1}^n \mathrm{V}(a_{2tn})$, we have $nh_3^{D_3}S_{2n}^2 = m_1^2(w)f_W^{-2}(w)\int\left(\int K_3(\gamma_1)K_3(\gamma_1+\gamma_2)\,\mathrm{d}\gamma_1\right)^2 f_W(w+h_3\gamma_2)\,\mathrm{d}\gamma_2 - n^2h_3^{D_3}\mu_n^2 \to \Phi_{m1,2}$. Thus, by Lyapunov's CLT, if $\sum_{t=1}^n \mathrm{E}|a_{2tn}/S_{2n}|^{2+\delta} \to 0$ for some $\delta > 0$ as $n\to\infty$, we have $\sum_{t=1}^n a_{2tn}/S_{2n} \xrightarrow{d} \mathcal{N}(0,1)$, i.e., combining previous results on other terms in T_3,

$$\sqrt{nh_3^{D_3}}\left(T_3 - b_{m1,2}(w)\right) \xrightarrow{d} \mathcal{N}(0,\Phi_{m1,2}), \quad \text{where } \Phi_{m1,2} \equiv \frac{m_1^2(w)}{f_W(w)}\int\left(\int K_3(\gamma_1)K_3(\gamma_1+\gamma_2)\,\mathrm{d}\gamma_1\right)^2\mathrm{d}\gamma_2.$$

Given that $nh_3^{D_3}S_{2n}^2 \to \Phi_{m1,2} > 0$, Lyapunov's condition is satisfied since

$$\sum_{t=1}^n \mathrm{E}\left|\frac{a_{2tn}}{S_{2n}}\right|^{2+\delta} \le \frac{C\left(nh_3^{D_3}\right)^{\delta/2+1}}{\left(nh_3^{D_3}S_{2n}^2\right)^{\delta/2+1}}\sum_{t=1}^n \mathrm{E}(|Z_{tn}|^{2+\delta}) \le C\left(nh_3^{D_3}\right)^{-\delta/2} \to 0, \text{ as } n\to\infty.$$

Step 4: Combining results from (1) to (3), we have $\sqrt{nh_3^{D_3}}\left(\hat{m}_1(w) - m_1(w) - b_{m1}(w)\right) = \sqrt{nh_3^{D_3}}\sum_{t=1}^n(a_{1tn}+a_{2tn})$, where $b_{m1}(w) = b_{m1,1}(w)+b_{m1,2}(w) = h_3^{s_3}\frac{\mu_{k_3,s_3}}{f_W(w)}\sum_{k=0}^{s_3}\frac{1}{k!(s_3-k)!}\sum_{j=1}^{D_3}\mathrm{D}_j^k m_1(w)\mathrm{D}_j^{s_3-k}f_W(w)+o_p(h_3^{s_3})$. Reapplying Lyapunov's CLT, given that $S_n^2 \equiv \mathrm{V}\left(\sum_{t=1}^n(a_{1tn}+a_{2tn})\right) = S_{1n}^2+S_{2n}^2+2\sum_{t=1}^n \mathrm{Cov}(a_{1tn},a_{2tn}) = S_{1n}^2+S_{2n}^2$ as $\mathrm{E}(a_{1tn}a_{2tn}) = 0$, and $nh_3^{D_3}S_n^2 \to \Phi_{m1,1}+\Phi_{m1,2}$, we have $\sqrt{nh_3^{D_3}}\left(\hat{m}_1(w)-m_1(w)-b_{m1}(w)\right) \xrightarrow{d} \mathcal{N}(0,\Phi_{m1,1}+\Phi_{m1,2})$. Lyapunov's condition can be easily verified using C_r Inequality.

Next, we extend this result for $\hat{m}_1(w)$ to $\hat{m}(w)$. Recall that,

$$\hat{m}_1(w) = \frac{1}{nh_3^{D_3}f_W(w)}\sum_{t=1}^n K_{3t}\hat{\eta}_t Y_t, \qquad \hat{m}(w) = \frac{1}{nh_3^{D_3}f_W(w)}\sum_{t=1}^n K_{3t}\hat{\eta}_t\left(Y_t - X_{2t}'\beta-\beta_0\right).$$

We see that $\hat{m}(w)$ shares a similar structure as $\hat{m}_1(w)$ except using $\hat{\eta}_t\left(Y_t - X_{2t}'\beta-\beta_0\right)$ instead of $\hat{\eta}_tY_t$ as the regressand. Given that $\eta_t\left(Y_t - X_{2t}'\beta-\beta_0\right) = m(W_t)+v_{mt}$, $\mathrm{E}(v_{mt}^2|W_t) = \sigma_{vm}^2 \le C$, and $\mathrm{E}(|v_{mt}|^{2+\delta}|W_t) \le C$, by repeating Step 1–4, we have $\sqrt{nh_3^{D_3}}\left(\hat{m}(w)-m(w)-b_m(w)\right) \xrightarrow{d} \mathcal{N}(0,\Phi_3+\Phi_4)$, where $b_m(w) = h_3^{s_3}\frac{\mu_{k_3,s_3}}{f_W(w)}\sum_{k=0}^{s_3}\frac{1}{k!(s_3-k)!}\sum_{j=1}^{D_3}\mathrm{D}_j^k m(w) \times \mathrm{D}_j^{s_3-k}f_W(w)+o_p(h_3^{s_3})$, $\Phi_3 = \frac{\sigma_{vm}^2}{f_W(w)}\int K_3^2(\gamma)\,\mathrm{d}\gamma$, $\Phi_4 = \frac{m^2(w)}{f_W(w)}\int\left(\int K_3(\gamma_1)K_3(\gamma_1+\gamma_2)\,\mathrm{d}\gamma_1\right)^2\mathrm{d}\gamma_2$. □

Lemmas

We start by noting that for any kernel K that satisfies Assumption A1, and for any function $f(x):\mathbb{R}^D\to\mathbb{R}$ such that $\int|f(\gamma)|\,\mathrm{d}\gamma<\infty$, we have that if x is a point of continuity of $f(x)$,

$$\int K(\gamma)f(x+h_n\gamma)\,\mathrm{d}\gamma \to f(x)\int K(\gamma)\,\mathrm{d}\gamma \quad \text{as} \quad n\to\infty.$$

This result follows directly from Theorem 1A in Parzen (1962).

Lemma 1. *Assume that $K(x):\mathbb{R}^D\to\mathbb{R}$ is a product kernel $K(x)=\prod_{j=1}^D k(x_j)$ with $k(x):\mathbb{R}\to\mathbb{R}$ such that: (a) $k(x)$ is continuously differentiable everywhere; (b) $|k(x)||x|^3\le C$, for any $x\in\mathbb{R}$ and some $C>0$; (c) $|k^{(1)}(x)||x|^3\le C$, for any $x\in\mathbb{R}$ and some $C>0$. Thus, for any $|\beta|=0,\ldots,3$, $K(x)x^\beta$ satisfies a local Lipschitz condition, i.e., for any $x\ne y\in A$, where $A\subset\mathbb{R}^D$ is a bounded convex set, we have $|K(x)x^\beta-K(y)y^\beta|\le C\|x-y\|_E$, for some $C>0$.*

Lemma 2. *Let $\{X_i\}_{i=1}^n$ be a sequence of independent and identically distributed (IID) random variables, $G_n(X_i,x):\mathbb{R}\times\mathbb{R}^K\to\mathbb{R}$ such that: (a) $|G_n(X_i,x)-G_n(X_i,x')|\le B_n(X_i)\|x-x'\|$ for all x,x' and $B_n(X_i)>0$ with $E(B_n(X_i))<C<\infty$; (b) $E(G_n(X_i,x))<\infty$ and $E\left(|G_n(X_i,x)-E(G_n(X_i,x))|^p\right)\le C^{p-2}p!E\left((G_n(X_i,x)-E(G_n(X_i,x)))^2\right)<\infty$ for some $C>0$ for all $i=1,2,\ldots$ and $p=3,4,\ldots$. Then, if $S_n(x)=\frac{1}{n}\sum_{i=1}^n G_n(X_i,x)$, for $x\in G_x$, an arbitrary convex compact subset of $\mathbb{R}^K$,*

$$\sup_{x\in G_x}|S_n(x)-E(S_n(x))| = O_p\left(\left(\frac{\log n}{n}\right)^{1/2}\right).$$

Lemma 3. *Assume that $K(x):\mathbb{R}^D\to\mathbb{R}$ is a product kernel $K(x)=\prod_{j=1}^D k(x_j)$ with $k(x):\mathbb{R}\to\mathbb{R}$ such that: (a) $k(x)$ is continuously differentiable everywhere; (b) $|k(x)||x|^{7+a}\to 0$ as $|x|\to\infty$ for some $a>0$; (c) $|k^{(1)}(x)||x|^3\le C$ for all x and some $C<\infty$. In addition, assume that (1) $\{(X_t,\varepsilon_t)'\}_{t=1,2,\ldots}$ is an independent and identically distributed sequence of random vectors; (2) The joint density of X_t and ε_t is given by $f_{X\varepsilon}(x,\varepsilon)=f_X(x)f_{\varepsilon|X}(\varepsilon|x)$; (3) $f_X(x)$ is everywhere continuous and uniformly* *bounded. Let* $w(X_t - x; x) : \mathbb{R}^D \to \mathbb{R}$ *and* $g(\varepsilon) : \mathbb{R} \to \mathbb{R}$ *be measurable functions. Define*

$$s(x) = \frac{1}{nh_n^D}\sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right)\left(\frac{X_t - x}{h_n}\right)^{\beta} w(X_t - x; x) g(\varepsilon_t),$$

where $|\beta| = 0, 1, 2, 3$. *If*

(i) $E\left(|g(\varepsilon_t)|^a \middle| X_t\right) \le C < \infty$ *for some* $a \ge 2$*;*

(ii) $w(X_t - x; x)$ *satisfies a Lipschitz condition of order 1 in x, i.e.,* $|w(X_t - x; x) - w(X_t - x^k; x^k)| \le C\|x - x^k\|_E$ *for some* $C > 0$*, and* $|w(X_t - x, x)| < C$ *for all* $x \in \mathbb{R}^D$.

Then, for an arbitrary compact set $\mathscr{G} \subseteq \mathbb{R}^D$*, we have* $\sup_{x\in\mathscr{G}} |s(x) - E(s(x))| = O_p\left(\left(\frac{\log n}{nh_n^D}\right)^{1/2}\right)$*, provided that* $h_n \to 0$*,* $nh_n^{D+2} \to \infty$ *and* $\frac{nh_n^D}{\log n} \to \infty$ *as* $n \to \infty$.

Lemma 4. *Let* $\{M_i\}_{i=1}^n$ *be a sequence of independent and identically distributed random vectors with the same distribution as* $M = (X \quad Z \quad U \quad \varepsilon\ )$ *and* $G(M)$ *a continuous function of* M *with* $\mathrm{E}(G^2(W)|Z) \le C < \infty$*. Then, if the joint density* f_M *of* M *is continuous,*

$$S_n = \frac{1}{n}\sum_{i=1}^{n} G(M_i)\left(\hat{\eta}(W_i, \hat{U}_i) - \eta(W_i, U_i)\right) = \begin{cases} o_p(n^{-1/2}), & \text{if} \quad \mathrm{E}(G(M_i)|X_i, Z_i, U_i) = 0 \\ O_p\left(n^{-1/2} + \sum_{i=1}^{4} h_i^{s_i}\right), & \text{if} \quad \mathrm{E}(G(M_i)|X_i, Z_i, U_i) \ne 0 \end{cases}.$$

Appendix B. Supplementary material: proofs of lemmas and technical details

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jspi.2020.02.002.

References

Chamberlain, G., 1992. Efficiency bounds for semiparametric regression. Econometrica 60, 567–596.
Chen, S.X., 2000. Probability density function estimation using gamma kernels. Ann. Inst. Statist. Math. 52 (3), 471–480.
Cheng, Y., Zerom, D., 2015. A quantile regression model for time-series data in the presence of additive components. Comm. Statist. Theory Methods 44, 4354–4379.
Cuzick, J., 1992. Efficient estimates in semiparametric additive regression models with unknown error distribution. Ann. Statist. 20, 1129–1136.
Fan, J., 1992. Design-adaptive nonparametric regression. J. Amer. Statist. Assoc. 87, 998–1004.
Fan, J., Härdle, W., Mammen, E., 1998. Direct estimation of low dimensional components in additive models. Ann. Statist. 26, 943–971.
Fan, Y., Li, Q., 1999. Root-n-consistent estimation of partially linear time series models. J. Nonparametr. Stat. 11, 251–269.
Fan, Y., Li, Q., 2003. A kernel-based method for estimating additive partially linear models. Statist. Sinica 13, 739–762.
Heckman, N.E., 1986. Spline smoothing in a partly linear model. J. R. Stat. Soc. Series B, 244–248.
Henderson, D., Parmeter, C.F., 2015. Applied nonparametric econometrics. Cambridge University Press, New York, NY.
Hoeffding, W., 1948. A class of statistics with asymptotically normal distribution. Ann. Math. Stat. 19, 293–325.
Juhl, T., Xiao, Z., 2005. Partially linear models with unit roots. Econometric Theory 21, 877–906.
Kim, W., Linton, O.B., Hengartner, N., 1999. A computationally efficient oracle estimator for additive nonparametric regression with bootstrap confidence intervals. J. Comput. Graph. Statist. 8, 278–297.
Li, Q., 2000. Efficient estimation of additive partially linear models. Internat. Econom. Rev. 41, 1073–1092.
Li, Q., Racine, J., 2007. Nonparametric econometrics: theory and practice. Princeton University Press, Princeton, NJ.
Li, Q., Wooldridge, J., 2000. Estimating semiparametric econometrics models by local linear method: with an application to cross-country growth. Ann. Econ. Finance 1, 337–357.
Li, Q., Wooldridge, J., 2002. Semiparametric estimation of partially linear models for dependent data with generated regressors. Econometric Theory 18, 625–645.
Linton, O.B., 1995. Second order approximation in the partially linear regression model. Econometrica 63, 1079–1112.
Linton, O.B., Härdle, W., 1996. Estimation of additive regression models with known links. Biometrika 83, 529–540.
Mammen, E., Rothe, C., Schienle, M., 2016. Semiparametric estimation with generated covariates. Econometric Theory 32, 1140–1177.
Manzan, S., Zerom, D., 2005. Kernel estimation of a partially linear additive model. Statist. Probab. Lett. 72, 313–322.
Manzan, S., Zerom, D., 2010. A semiparametric analysis of gasoline demand in the United States: reexamining the impact of price. Econometric Rev. 29, 439–468.
Martins-Filho, C., Yang, K., 2007. Finite sample performance of kernel-based regression methods for non-parametric additive models under common bandwidth selection criterion. J. Nonparametr. Stat. 19, 23–62.
Martins-Filho, C., Yao, F., 2012. Kernel-based estimation of semiparametric regression in triangular systems. Econom. Lett. 115, 24–27.
Newey, W.K., Powell, J.L., Vella, F., 1999. Nonparametric estimation of triangular simultaneous equation models. Econometrica 67, 565–603.
Ozabaci, D., Henderson, D., 2015. Additive kernel estimates of returns to schooling. Empir. Econ. 48, 227–251.
Ozabaci, D., Henderson, D., Su, L., 2014. Additive nonparametric regression in the presence of endogenous regressors. J. Bus. Econom. Statist. 32, 555–575.
Parzen, E., 1962. On estimation of a probability density and mode. Ann. Math. Stat. 33, 1065–1076.
Pinkse, J., 2000. Nonparametric two-step regression estimation when regressors and error are dependent. Canad. J. Statist. 28, 289–300.
Robinson, P.M., 1988. Root-N-consistent semiparametric regression. Econometrica 56, 931–954.
Ruppert, D., Sheather, S., Wand, M., 1995. An effective bandwidth selector for local least squares regression. J. Amer. Statist. Assoc. 90, 1257–1270.
Schick, A., 1986. On asymptotically efficient estimation in semiparametric models. Ann. Statist. 14, 1139–1151.
Simonoff, J., 1996. Smoothing methods in statistics. Springer Series in Statistics, Springer Verlag.
Su, L., Ullah, A., 2008. Local polynomial estimation of nonparametric simultaneous equations models. J. Econometrics 144, 193–218.
Yao, F., Martins-Filho, C., 2015. An asymptotic characterization of finite degree U-statistics with sample size dependent kernels: applications to nonparametric estimators and test statistics. Commun. Stat. - Theory Methods 44, 3251–3265.
Yu, K., Mammen, E., Park, B.U., 2011. Semi-parametric regression: efficiency gains from modeling the nonparametric part. Bernoulli 17, 736–748.