On Endogeneity and Shape Invariance in Extended Partially Linear Single Index Models

Jiti Gao[1]
Monash University, Australia

Namhyun Kim[2]
University of Exeter Business School, United Kingdom

Patrick W. Saart
Cardiff Business School, United Kingdom

Abstract

In this paper, the important (but so far unrevealed) usefulness of the extended generalized partially linear single-index (EGPLSI) model introduced by Xia et al. (1999) in its ability to model a flexible shape-invariant specification is elaborated. More importantly, a control function approach is proposed to address the potential endogeneity problems in the EGPLSI model in order to enhance its applicability to empirical studies. In this process, it is shown that the attractive asymptotic features of the single-index type of semiparametric model are still valid in our proposed estimation procedure, given the intrinsic generated covariates. Our newly developed method is then applied to address the endogeneity of expenditure in the semiparametric analysis of a system of empirical Engel curves using the British data, which highlights the convenient applicability of our proposed method.

JEL Classification: C14, C18, C51.

Keyword: Extended generalized partially linear single-index, control function approach, endogeneity, semiparametric regression models.

[1]This author acknowledges the Australian Research Council Discovery Grants Program for its support under grant numbers: DP150101012 & DP170104421.

[2]Namhyun, Kim, University of Exeter Business School, UK, email: n.kim@exeter.ac.uk

The authors thank the editor and two anonymous referees for their kind and thoughtful suggestions and comments on the paper. All remaining errors are our own.

1. Introduction

Xia et al. (1999) introduced the extended generalized partially linear single-index (EGPLSI) model of the form

$$Y_i = X_i'\beta_0 + g(X_i'\alpha_0) + \epsilon_i, \tag{1.1}$$

where (i) (X, Y) is a set of $\mathbb{R}^q \times \mathbb{R}$-valued observable random vectors; (ii) β_0 and α_0 are vectors of unknown parameters such that $\beta_0 \perp \alpha_0$, showing the orthogonality of β_0 and α_0 for identifiability, with $||\alpha_0|| = 1$ and (iii) $g(\cdot)$ is an unknown structural link function such that $g(\cdot) : \mathbb{R} \to \mathbb{R}$ and $g''(\cdot) \neq 0$. In addition, it is assumed that $E(\epsilon|X) = 0$, a usual exogeneity assumption suggesting that $E(\epsilon|V_0) = 0$ where $V_0 = X'\alpha_0$. In fact, the EGPLSI model is an extended version of the generalized partially linear single-index (GPLSI) model of Carroll et al. (1997) and Xia and Härdle (2006) and hence a number of non- and semiparametric models are special cases of the EGPLSI model.

The current paper refines the usefulness of the EGPLSI model in modelling the kind of flexible shape-invariant specification often considered in pooling nonparametric regression curves (see Härdle and Marron (1990), and Robinson and Pinkse (1995) for examples). Furthermore, the paper also aims to address a breakdown of the exogeneity assumption in the EGPLSI model, particularly the endogeneity problems that cause unidentification of the structural link function, in order to enhance the applicability of the EGPLSI model to empirical studies.

Recently, a number of methods have been discussed in the literature on how the endogeneity problems can be best addressed in non- and semiparametric models. Among these, two of the most popular alternatives are nonparametric instrumental variable estimation (NPIV) and the control function (CF) approach (see Blundell and Powell (2003) for an excellent review of endogeneity in non- and semiparametric models). The NPIV estimation relies on different stochastic assumptions from the CF approach and is performed without estimating a first-stage reduced-form equation. Nonetheless, there are a few well-known difficulties that are intrinsic to the NPIV estimation, particularly the so-called ill-posed inverse problem (see Ai and Chen (2003), and Blundell et al. (2007) for details). On the other hand, the CF approach allows the specification of endogeneity, which is based on the intuitive triangular structure of a model (see Blundell et al. (1998), and Blundell and Powell (2003) for details). This triangular structure of the CF approach also provides an accessible way of addressing the weak instruments problem in a nonparametric regression model by translating the weak instruments problem into a simpler one, namely the multicollinearity problem (see Han (2012) for details). Hence, the development of the CF approach in the EGPLSI model also provides a foundation for addressing the presence of weak instruments in semiparametric regression models.

This paper also aims to develop the CF approach, mainly for the reasons given above. Although the generated covariates issue is intrinsic in the development of the CF approach, similar to the study of Mammen et al. (2016), the proposed method maintains the attractive features of the single-index (SI) model with the relatively mild conditions seen in the literature and shows an accessible extension to strictly stationary and strongly mixing (α-mixing) process. In a SI model, Härdle et al. (1993) showed that the optimal bandwidth for estimating a structural link function (in the sense of the integrated mean square error (IMSE)) can be used for the $\sqrt{n}$-consistent estimation of the index coefficients. Xia et al. (1999) then extended the optimization technique and asymptotic results of Härdle et al. (1993) to estimation of the EGPLSI model for a strictly stationary and strongly mixing process. The current paper proceeds one step further by showing that under-smoothing for estimating a first-stage reduced-form equation is not required in the newly proposed CF's two-stage nonparametric/EGPLSI estimation in order to achieve $\sqrt{n}$-consistent estimation of α_0. These results are developed in details with the simplest data structure, namely an independently and identically distributed (i.i.d.) random sample, then extended to a strictly stationary and strongly mixing case. Furthermore, the convenient applicability of our newly developed CF approach to an empirical study is explored by analyzing empirical Engel curves based on British data.

The structure of the rest of the paper is as follows. In Section 2, the usefulness of the EGPLSI model for modelling a flexible shape-invariant specification is elaborated. In addition, the development of the CF approach in the EGPLSI model and a Monte Carlo exercise assessing the finite-sample performance of the proposed estimation procedure are also presented. In Section 3, the implementation of the empirical study of the cross-sectional relationships between specific goods and the level of total expenditure is investigated. Finally, Section 4 concludes the paper with a summary of the main findings and further issues to be investigated. All mathematical proofs of the main theoretical results of the paper are presented in the Appendix.

2. EGPLSI Model, Shape-Invariant Specification and Endogeneity

In this section, the usefulness of the EGPLSI model introduced by Xia et al. (1999) is firstly elaborated for specifying a flexible shape-invariant specification. This section then introduces endogeneity into the EGPLSI model, establishes the CF approach to address the endogeneity problems and presents the asymptotic properties and finite sample performances of a Monte Carlo simulation exercise for the proposed estimators.

2.1. Shape-Invariant Specification within EGPLSI Model Framework

A shape-invariant specification in modelling an aggregate structural relationship incorporating individual heterogeneity is easily found in various areas of economics. For instance, Blundell and Stoker (2007) suggested modelling consumption patterns with the demographic differences of individual households, and Nagin and Odgers (2010) and LaFree et al. (2009) proposed specifying group-based trajectories in, respectively, clinical research with heterogeneous subject groups over time and in cross-national politically motivated violence over time. The EGPLSI model allows this type of shape-invariant specification with functional flexibility because both the scale and shift parameters can be incorporated into the model. Below, we discuss how to model a flexible shape-invariant specification within the EGPLSI model framework in detail.

Let us consider a flexible shape-invariant specification within the EGPLSI model framework by considering the two sets of observations below. The first set of observations $(X_1, Y_1), \ldots, (X_n, Y_n)$, for example, is assumed to follow the data-generating process (d.g.p.) below

$$Y_i = m_1(X_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{2.1}$$

where ε is assumed to be independent with a mean of 0 and the common variance σ^2. Suppose the second set of observations $(X_1', Y_1'), \ldots, (X_n', Y_n')$ is from the following nonparametric regression model

$$Y_i' = m_2(X_i') + \varepsilon_i', \tag{2.2}$$

where ε' is independent from ε but otherwise has the same stochastic structure as ε and has the common variance σ'^2. The main interest here is to model the curves whose parametric nature is modelled by

$$m_2(X') = S_{\theta_0}^{-1}(m_1(T_{\theta_0}^{-1}(X'))), \tag{2.3}$$

where T_θ and S_θ are invertible transformations, particularly scalings and shifts of the axes indexed by the parameters $\theta \in \Theta \subseteq R^d$, and θ_0 is the vector of the true values of the parameters. A good estimate of θ_0 is provided by θ for which the curve $m_1(X)$ is closely approximated by

$$m(X, \theta) = S_\theta(m_2(T_\theta(X))). \tag{2.4}$$

For the sake of illustration, the simple models are considered as follows

$$m_1(X) = (X - 0.4)^2 \text{ and } m_2(X') = (X' - 0.5)^2 - 0.2, \tag{2.5}$$

which fit the framework described by (2.1) to (2.4) by defining the following

$$\begin{aligned} T_\theta(X) &= \theta^{(1)}X + \theta^{(2)} \\ m_2(T_\theta(X)) &= \left(\theta^{(1)}X + \theta^{(2)} - 0.5\right)^2 - 0.2 \\ S_\theta(m_2(T_\theta(X))) &= \left(\theta^{(1)}X + \theta^{(2)} - 0.5\right)^2 - 0.2 + \theta^{(3)}X + \theta^{(4)}, \end{aligned}$$

where $\theta_0 = \left(\theta_0^{(1)}, \theta_0^{(2)}, \theta_0^{(3)}, \theta_0^{(4)}\right) = (1, 0.1, 0, 0.2)$.

When a curve comparison problem with a similar parametric nature to (2.3) needs to be considered, Härdle and Marron (1990) suggested an estimation procedure by which separated kernel smoothers are used in order to compute the estimates of $m_1(\cdot)$ and $m_2(\cdot)$. The estimator of θ_0 is then found by minimizing a L^2-norm objective function of the kernel estimates of $m_1(\cdot)$ and $m_2(\cdot)$, and the approximation in (2.4). Alternatively, pooling the two sets of observations is more desirable. Modelling the data within the EGPLSI model framework enables this type of pooling nonparametric regression. The shift and scaling of the axes illustrated in the example above fit in the EGPLSI framework, shown as below

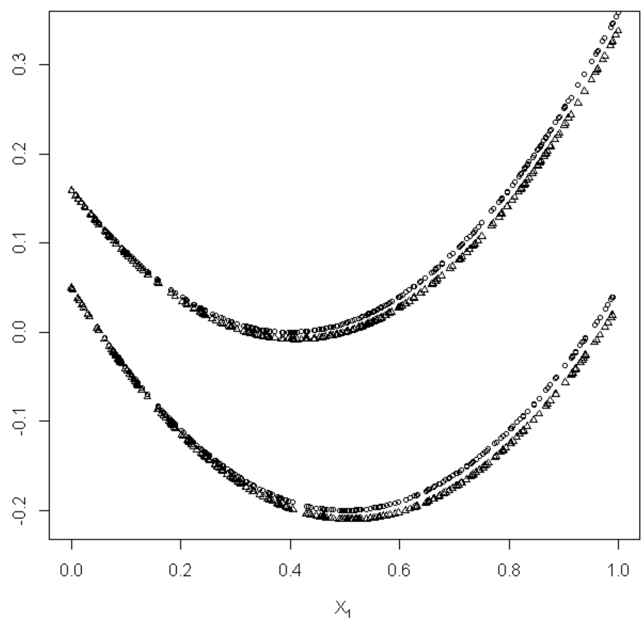
$$m_3(X_1, X_2) = [\beta_{01}X_1 + \beta_{02}X_2] + \left\{([\alpha_{01}X_1 + \alpha_{02}X_2] - 0.5)^2 - 0.2\right\}, \tag{2.6}$$

where $X_1 = \begin{cases} X \\ X' \end{cases}$ and $X_2 = \begin{cases} 1 & \text{if } X_1 = X \\ 0 & \text{if } X_1 = X' \end{cases}$. The model examples in (2.5) can be obtained by defining

$$(\beta_{01}, \beta_{02}, \alpha_{01}, \alpha_{02}) = (0, 0.2, 1, 0.1). \tag{2.7}$$

Five hundred simulated observations of the model are represented by circles in Figure 2.1, where X_{1i} on the x-axis is a uniform random variable on $[0, 1]$ for $i = 1, \ldots, 500$.

Figure 2.1. *Five hundred simulated observations based on* $m_3(\cdot, \cdot)$.



The two sets of observations are determined by X_2, which is a Bernoulli random variable with the parameter $p = 0.5$. It should be noted, however, that the set of values of the parameters in (2.7) does not satisfy the identification conditions which require that $\beta_0 \perp \alpha_0$ with $||\alpha_0|| = 1$. An approximate model that satisfies these identification conditions is obtained by first setting $\beta_{02} = 0.2$ and $\alpha_{02} = 0.1$, so that $\beta_{01} = -0.02$ and $\alpha_{01} = 0.99$ can be derived. five hundred simulated observations of this type of model are represented by triangles in Figure 2.1. In practice, when there is enough reason to believe (perhaps based on economic theory) that $\beta_{01} = 0$ and $\alpha_{01} = 1$, then such a model can be obtained by scaling and shifting, respectively, as follows

$$X_2 = v_{01} - \beta_{01} X_1 \text{ and } X_1 + \frac{\alpha_{02}}{\alpha_{01}} X_2 = \frac{v_{02}}{\alpha_{01}},$$

where $\beta_{01}X_1 + \beta_{02}X_2 = v_{01}$ and $\alpha_{01}X_1 + \alpha_{02}X_2 = v_{02}$. This method is illustrated in the empirical analysis in Section 3.

2.2. Endogeneity and Newly Proposed Estimation Procedure

Despite its ability to model a flexible shape-invariant specification, the applicability of the EGPLSI model in (1.1) to an empirical study is limited because of its shortfalls in addressing endogeneity problems. There are two potential sources of endogeneity in the model, namely endogeneity in the parametric and in the nonparametric components. Hereafter, let us refer to these as "parametric-endogeneity" and "nonparametric-endogeneity", respectively. The simultaneous occurrence of these two types of endogeneity is also possible. If it is present, parametric-endogeneity can be dealt via parametric IV estimation in place of the usual least-squares (LS) estimation method.[3] Because of the partialling-out process, as in the estimation procedure of the partially linear (PL) type of semiparametric model of Robinson (1988) and Speckman (1988), the $\sqrt{n}$-consistent LS estimator of β_0 is still obtainable even in the presence of the nonparametric-endogeneity unless the parametric covariates are endogenous. Hence, to simplify the argument, the parametric covariates are assumed to belong to a subset $X_1 \subseteq \mathbb{R}^{q_1}$ for $q_1 < q$ of X such that $E(\epsilon|X_1) = 0$, namely the parametric covariates are exogenous, without loss of generality.

In this case, nonparametric-endogeneity exists when $E(\epsilon|X) \neq 0$, which implies that $E(\epsilon|V_0) \neq 0$. An unexpected property from the SI type of semiparametric models is that estimators of the index coefficients are still $\sqrt{n}$-consistent even with the presence of nonparametric-endogeneity. The literature, particularly Ichimura (1993), Härdle et al. (1993), and Xia and Härdle (2006), suggested estimating the index coefficients by minimizing a L^2-norm objective function measuring the distance between a structural link function and its approximation by the conditional expectation relationship given a set of the initial values of the index coefficients. The disturbance term in the minimizing objective function is then endogeneity-free because of the partialling-out process of the estimation procedure of the SI type of semiparametrics. Nonetheless, the structural link function in the EGPLSI model is

[3]A comprehensive discussion of parametric IV estimation of the PL type of semiparametric models can be found in Li and Racine (2007).

unidentifiable by using the conditional expectation relationship in the presence of nonparametric-endogeneity. As a result of this, the optimization procedure in Xia et al. (1999) is no longer applicable.

In the following, let us present the development of the CF approach in the EGPLSI model. For the sake of notational simplicity, the simplest case is considered, namely the presence of an endogenous nonparametric covariate denoted by X_2.[4] Hereafter, let Z denote a vector of valid instruments for X_2 as follows

$$X_{2i} = g_x(Z_i) + \eta_i, \tag{2.8}$$

where $E(\eta|Z) = 0$, and

$$E(\epsilon|X_2) = E(\epsilon|Z, \eta) = E(\epsilon|\eta) \equiv \iota(\eta), \tag{2.9}$$

where (X_2, Z) is a set of $\mathbb{R} \times \mathbb{R}^{q_z}$-valued observable random vectors, and $g_x(Z)$ and $\iota(\eta)$ are unknown real functions such that $g_x(\cdot) : \mathbb{R}^{q_z} \to \mathbb{R}$ and $\iota(\cdot) : \mathbb{R} \to \mathbb{R}$, respectively. The stochastic assumption of (2.9) is standard in the CF literature, suggesting the exogeneity condition of Z, particularly $E(\epsilon|Z, \eta) = E(\epsilon|\eta)$ (see Newey et al. (1999), Blundell and Powell (2004), and Su and Ullah (2008) for examples). Furthermore, the necessary identification condition for the structural link function ($g(\cdot)$ function) as discussed in Newey et al. (1999) is the non-existence of a linear functional relationship between X_2 and η.

By imposing the structure of (2.8) and (2.9), the EGPLSI model in (1.1) with the presence of nonparametric-endogeneity is rewritten as

$$Y_i = X_i'\beta_0 + m(V_{0i}, \eta_i) + e_i, \tag{2.10}$$

where $m(v_0, \eta) \equiv g(v_0) + \iota(\eta)$ with $\iota(\eta) \neq 0$ being the endogeneity control function, and $E(e|X) = 0$. The conditional expectation relationship, based on (2.10), is obtained as follows

$$m_y(v_0, \eta) \equiv m(v_0, \eta) + m_x(v_0, \eta)'\beta_0, \tag{2.11}$$

where $m_y(v_0, \eta) \equiv E(y|V_0, \eta)$ and $m_x(v_0, \eta) \equiv E(x|V_0, \eta)$.

[4]The generalized version (namely when there are more than one endogenous nonparametric covariates) is available on request from the author.

In the following, the performance of the CF approach in the EGPLSI model based on (2.8) to (2.11) is discussed. The identification issue is first presented as follows. Given α and β, let

$$J(\alpha,\beta) = E\left[Y - E(Y|V,\eta) - \{X - E(X|V,\eta)\}'\beta\right]^2$$
$$\mathcal{V} = E(\{X - E(X|V,\eta)\}\{X - E(X|V,\eta)\}')$$
$$\mathcal{W} = E(\{X - E(X|V,\eta)\}\{Y - E(Y|V,\eta)\}),$$

where $V = X'\alpha$. Suppose that $g(\cdot)$ is twice differentiable and that X has a positive density function on a union of a finite number of open convex subsets in $\mathbb{R}^q$. The minimum point of $J(\alpha,\beta)$ with $\alpha \perp \beta$ is then unique at α_0 and $\beta_{\alpha_0} = \{\mathcal{V}(\alpha_0)\}^+\mathcal{W}(\alpha_0)$, where $\{\mathcal{V}(\alpha_0)\}^+$ is the Moore-Penrose inverse.

Before we discuss the optimization procedure, the necessary notation is defined for the sake of convenience. We assume that the random sample $\{(X_i', Z_i', Y_i); i = 1,\ldots,n\}$ is i.i.d. Let $f_x(x)$ and $f_z(z)$ denote the joint density functions of X' and Z', respectively. Let us also denote $f_\alpha(v)$ as the density function of $V = X'\alpha$. We assume that $\mathcal{A}_j \subset \mathbb{R}^k$ is the union of a finite number of open sets such that $f_j(s) > C$ on $\mathcal{A}_j$, where $k = q$ or q_z and $j = x$ or z for some constant $C > 0$. Hereafter, this region is considered to avoid the boundary points. Because the region is not known in practice, Xia and Härdle (2006) suggested using the weight function such that $I_n(s) = 1$ if $\frac{1}{n}\sum_{i=1}^n K_{j,i}(s) > C$ and 0 otherwise, where K_j is a corresponding kernel function. In this paper, $I_n(s)$ is omitted for notational simplicity. In addition, C, C' and C'' denote generic constants varying from one place to another. The conditional expectations, namely $E(Y|V,\eta)$ and $E(X|V,\eta)$, are then estimated with the leave-one-out nonparametric estimation as follows

$$\hat{E}_i(Y_i|V_i,\eta_i) = \frac{\sum_{j\neq i} L_{h_v h_\eta}(V_j - V_i, \eta_j - \eta_i)Y_j}{\sum_{j\neq i} L_{h_v h_\eta}(V_j - V_i, \eta_j - \eta_i)} \tag{2.12}$$

$$\hat{E}_i(X_i|V_i,\eta_i) = \frac{\sum_{j\neq i} L_{h_v h_\eta}(V_j - V_i, \eta_j - \eta_i)X_j}{\sum_{j\neq i} L_{h_v h_\eta}(V_j - V_i, \eta_j - \eta_i)}, \tag{2.13}$$

where $L_{h_v h_\eta}$ is a product kernel function constructed from the product of the univariate kernel functions of $k_{h_v}(\cdot) \times k_{h_\eta}(\cdot)$ with the relevant bandwidth parameters, h_v and h_η. Furthermore, the first-stage leave-one-out nonparametric estimation of the reduced equation in (2.8) used to estimate η_i is as follows

$$\hat{\eta}_i = X_i - \hat{g}_{x,i}(Z_i), \tag{2.14}$$

where $\hat{g}_{x,i}(Z_i) = \frac{\sum_{j\neq i} K_{h_z}(Z_j - Z_i)X_j}{\sum_{j\neq i} K_{h_z}(Z_j - Z_i)}$ with $K_{h_z}(\cdot)$ being the product kernel function constructed from $k_{h_{z_1}}(\cdot) \times \cdots \times k_{h_{z q_z}}(\cdot)$, and h_{z_j}, for $j = 1, \ldots, q_z$, is the relevant bandwidth parameter.

The LS estimates of the unknown parametric coefficients are then computed, given the initial values of the index coefficients denoted by α, as follows

$$\beta = \left(S_{\hat{U}_2}\right)^{-} S_{\hat{U}_2 \hat{W}_2}, \tag{2.15}$$

where $S_{AB} = \frac{1}{n}\sum_{i=1}^{n} A_i B_i'$, $S_A = S_{AA}$, $(S_A)^{-}$ is a generalized inverse of (S_A), $\hat{W}_{2i} \equiv Y_i - \hat{E}_i(Y_i|V_i, \hat{\eta}_i)$ and $\hat{U}_{2i} \equiv X_i - \hat{E}_i(X_i|V_i, \hat{\eta}_i)$. These are estimated by replacing η_i with $\hat{\eta}_i$ in (2.12) and (2.13), respectively. Next, based on $\beta \in B_n$, $\hat{\alpha}$, $\hat{h}_v$ and $\hat{h}_{\hat{\eta}}$ are computed by minimizing the objective function as follows

$$\min_{\alpha \in A_n, h_v, h_{\hat{\eta}} \in \mathcal{H}_n} \hat{J}(\alpha, h_v, h_{\hat{\eta}}) \equiv \min_{\alpha \in A_n, h_v, h_{\hat{\eta}} \in \mathcal{H}_n} \frac{1}{n} \sum_{i=1}^{n} (\hat{W}_{2i} - \hat{U}_{2i}'\beta)^2, \tag{2.16}$$

where

$$A_n = \{\alpha : ||\alpha - \alpha_0|| \leq Cn^{-1/2}\},\ B_n = \{\beta : ||\beta - \beta_0|| \leq Cn^{-1/2}\}$$
$$\text{and } \mathcal{H}_n = \left\{h_z, h_v, h_\eta : Cn^{-1/5} \leq h_z, h_v, h_\eta \leq C'n^{-1/5}\right\} \tag{2.17}$$

for $0 < C < C' < \infty$. Finally, we re-estimate β_0 by using $\hat{\alpha}$, $\hat{h}_{\hat{v}}$ and $\hat{h}_{\hat{\eta}}$ as follows

$$\hat{\beta} = \left(S_{\hat{U}_3}\right)^{-} S_{\hat{U}_3 \hat{W}_3}, \tag{2.18}$$

where $\hat{W}_{3i} \equiv Y_i - \hat{E}_i(Y_i|\hat{V}_i, \hat{\eta}_i)$ and $\hat{U}_{3i} \equiv X_i - \hat{E}_i(X_i|\hat{V}_i, \hat{\eta}_i)$ with $\hat{V}_i = X_i'\hat{\alpha}$, and $\hat{U}_{3i}$ and $\hat{W}_{3i}$ are estimated by replacing V_i and η_i with $\hat{V}_i$ and $\hat{\eta}_i$ in (2.12) and (2.13), respectively.

Remark 2.1. *The conditions for the finite-dimensional parameters in (2.17) seem to be restrictive at first glance. However, they are not restrictive, given that $\hat{\alpha}$ and $\hat{\beta}$ are $\sqrt{n}$-consistent. Furthermore, as shown in the mathematical proof, $\sqrt{n}$-consistency is achieved without under-smoothing in the first-stage of the proposed estimation procedure (i.e. estimation of the reduced-form equation in (2.8)). In general, under-smoothing in the first-stage of the estimation procedure is not required when $q_z < 3$ and $q - q_1 < 3/2$.*

The remaining task is then to identify the unknown structural link function. It is plausible to apply the marginal integration technique of Linton and Nielsen (1995), and Tjøstheim and Auestad (1994) to identify each of the functions because of the additive specification of the conditional expectation relation (see (2.10) below). As extensively discussed in the literature, a standard identification condition is to assume that $E(g(V_0)) = E(\iota(\eta)) = 0$ (see Hastie and Tibshirani (1990), Gao et al. (2006) and Gao (2007) for details). Hence, the marginal integration technique identifies $g(\cdot)$ and $\iota(\cdot)$ functions up to some constant value as follows

$$m_1(V_0) \equiv \int m(V_0, \eta)\, dQ(\eta) = g(V_0) + C \text{ and } m_2(\eta) \equiv \int m(V_0, \eta)\, dQ(V_0) = \iota(\eta) + C',$$

where $C \equiv \int \iota(\eta) dQ(\eta)$, $C' \equiv \int g(V_0) dQ(V_0)$ and Q is a probability measure in $\mathbb{R}$ with $\int dQ(\eta) = \int dQ(V_0) = 1$. The estimate of the structural link function can therefore be obtained by

$$\hat{m}_1(\hat{V}) = \frac{1}{n} \sum_{i=1}^{n} \hat{m}(\hat{V}, \hat{\eta}_i) \text{ and } \hat{g}(\hat{V}) = \hat{m}_1(\hat{V}) - \hat{C}, \tag{2.19}$$

where $\hat{m}(\hat{V}, \hat{\eta}_i) = \hat{E}(Y|\hat{V}, \hat{\eta}_i) - \hat{E}(X|\hat{V}, \hat{\eta}_i)'\hat{\beta}$, $\hat{C} = \frac{1}{n}\sum_{i=1}^{n} \hat{m}_1(\hat{V}_i)$, and $\hat{m}_1(\hat{V})$ is estimated by keeping $\hat{V}_i$ at $\hat{V}$ when taking an average over $\hat{\eta}_i$.

Before discussing the main theoretical results of the estimators proposed above, the estimation procedure is briefly summarized as follows.

Step 2.1: Estimate the endogeneity control covariate, $\hat{\eta}$, as in (2.14).

Step 2.2: Estimate β as in (2.15) with $\hat{\eta}_i$ from Step 2.1 and α.

Step 2.3: Estimate $\hat{\alpha}$ and $\hat{\beta}$ as in (2.16) and (2.18), respectively.

Step 2.4: Estimate $\hat{m}(\hat{V}_i, \hat{\eta}_i)$ by using (2.11) with $\hat{\alpha}$ and $\hat{\beta}$ from Step 2.3, then perform the marginal integration technique to estimate $\hat{g}(\hat{V})$ as in (2.19).

2.3. Asymptotic Properties of Proposed Estimators

In this subsection, the asymptotic properties of the estimators proposed above are discussed as follows. The required necessary conditions are presented first. Given ρ, let $\mathcal{A}_{j'}^{\rho}$ denote the set of all points in $\mathbb{R}^{k'}$, where $k' = q$ or 1, at a distance no greater than ρ from $\mathcal{A}_{j'}$ for $j' = x, \eta$. Let $\mathcal{U} = \{(V_0, \eta) : X \in \mathcal{A}_x^{\rho} \text{ and } \eta \in \mathcal{A}_{\eta}^{\rho}\}$ and $f(V_0, \eta)$ denote the joint density function of (V_0, η) with random arguments of X' and η. The necessary regularity conditions are then as follows.

Assumption 2.1. *Suppose that there is a vector of instrumental variables* $\{Z_i : i \geq 1\}$ *such that Equations (2.8) and (2.9) hold.*

Assumption 2.2. *(i) The joint density function of* $f_z(Z)$ *is bounded and is bounded away from zero with bounded and continuous second derivatives on* $\mathcal{A}_z$*. (ii) The joint density function of* $f(V, \eta)$ *is bounded and is bounded away from zero with bounded and continuous second derivatives on* $\mathcal{U}$ *for all* $\alpha \in A_n$*.*

Assumption 2.3. *(i) Assume that* $g_x(Z)$ *has bounded and continuous second derivatives on* $\mathcal{A}_z$*. (ii) Let* $m(V, \eta)$*,* $m_y(V, \eta)$ *and* $m_x(V, \eta)$ *have bounded and continuous second derivatives on* $\mathcal{U}$ *for all values of* $\alpha \in A_n$*.*

Assumption 2.4. *Suppose that a univariate kernel function* $k(\cdot)$ *and its first derivative* $k^{(1)}(\cdot)$ *are supported on the interval* $(-1, 1)$ *and* $k(\cdot)$ *is a symmetric density function. Furthermore, both* $k(\cdot)$ *and* $k^{(1)}(\cdot)$ *satisfy the Lipschitz conditions.*

Assumption 2.5. *Let* $E(\eta|Z) = 0$ *and* $E(\eta^2|Z) = \sigma_1^2(Z)$*,* $E(e|X, \eta) = 0$ *and* $E(e^2|X, \eta) = \sigma^2(X, \eta)$*,* $E(u|X, \eta) = 0$ *and* $E(u^2|X, \eta) = \sigma_2^2(X, \eta)$ *almost surely, and the functions* σ^2*,* σ_1^2 *and* σ_2^2 *are bounded and continuous. In addition,* $\sup_i E||X_i||^l < \infty$*,* $\sup_i E|Y_i|^l < \infty$ *and* $\sup_i E||Z_i||^l < \infty$ *for some large enough* $l \geq 2$*.*

Assumption 2.2 permits us to estimate the functions in the regions of $\mathcal{A}_z$ and $\mathcal{U}$, and to avoid the random denominator problem. In practice, the weight function of Xia and Härdle (2006) discussed above can be used. Assumptions 2.2 and 2.3 ensure that the kernel function in Assumption 2.4 leads to second-order bias in kernel smoothing. Higher-order bias can be achieved by imposing more restrictive conditions on the smoothness of the functions (see Robinson (1988) for details). The condition on the first derivative of the kernel function in Assumption 2.4 permits the use of the Taylor expansion argument to address the generated covariate, $\hat{\eta}_i$ (a similar condition on the derivatives of the kernel function can be found in Hansen (2008)). The Lipschitz conditions for both the kernel function and its derivative are convenient for the proof of the uniform convergence. Finally, Assumption 2.5 allows us the use of the Chebyshev inequality.

Now let us introduce some necessary notations used in the main theoretical results below. Let $\mathcal{K}_{z,2} = \int z^2 K_{h_z}(z)dz$, $\mathcal{K}_{\eta,2} = \int \eta^2 k_{h_\eta}(\eta)d\eta$ and $\mathcal{K}_{v,2} = \int v_0^2 k_{h_v}(v_0)dv_0$.

Furthermore, let $\mathcal{K}_z = \int k_{h_{z,j}}(z)^2 dz$ and $\mathcal{K} = \mathcal{K}_v \mathcal{K}_\eta$, where $\mathcal{K}_v = \int k_{h_v}(v_0)^2 dv_0$ and $\mathcal{K}_\eta = \int k_{h_\eta}(\eta)^2 d\eta$. Let $f_{z,j}^{(r)}$ be the r^{th} derivatives of $f_z(z)$ with respect to Z_j, for $j = 1, \cdots, q_z$, and let $f_{v_0}^{(r)}(v_0, \eta)$ and $f_\eta^{(r)}(v_0, \eta)$ be the r^{th} partial derivatives of $f(v_0, \eta)$ with respect to V_0 and η, respectively. Moreover, let $g_{x,j}^{(r)}(z)$ be the r^{th} partial derivatives of $g_x(z)$ with respect to Z_j, and let $m_{v_0}^{(r)}(V_0, \eta)$ and $m_\eta^{(r)}(v_0, \eta)$ be that of $m(v_0, \eta)$ with respect to V_0 and η, respectively. Then, let

$$B_z(z) \equiv \frac{\mathcal{K}_{z,2}}{2f(z)} \left\{ 2f_{z,j}^{(1)}(z) g_{x,j}^{(1)}(z) + f_z(z) g_{x,j}^{(2)}(z) \right\}$$

$$B_v(v_0, \eta) \equiv \frac{\mathcal{K}_{v,2}}{2f(v_0, \eta)} \left\{ 2f_{v_0}^{(1)}(v_0, \eta) m_{v_0}^{(1)}(v_0, \eta) + f(v_0, \eta) m_{v_0}^{(2)}(v_0, \eta) \right\}$$

$$B_\eta(v_0, \eta) \equiv \frac{\mathcal{K}_{\eta,2}}{2f(v_0, \eta)} \left\{ 2f_\eta^{(1)}(v_0, \eta) m_\eta^{(1)}(v_0, \eta) + f(v_0, \eta) m_\eta^{(2)}(v_0, \eta) \right\}.$$

In addition, let

$$IMSE_1(h_z) \asymp \int \left\{ \left[\sum_{j=1}^{q_z} B_{z,j}(z) h_{z,j}^2 \right]^2 + \frac{\mathcal{K}_z^{q_z}}{n h_{z,1} \ldots h_{z,q_2}} \frac{\sigma_1^2(z)}{f_z(z)} \right\} f(z) dz$$

$$IMSE_2(h_v, h_\eta) \asymp \int \left\{ \left[B_v(v_0, \eta) h_v^2 + B_\eta(v_0, \eta) h_\eta^2 \right]^2 + \frac{\mathcal{K}}{n h_v h_\eta} \frac{\sigma^2(V_0, \eta)}{f(v_0, \eta)} \right\} f(x, \eta) dx d\eta,$$

where $\asymp$ means that the quotient of the two sides tends to 1 as $n \to \infty$.

Theorem 2.1. *Under Assumptions 2.1 to 2.5, the minimizing objective function in (2.16) is rewritten as follows*

$$\hat{J}(\alpha, h_v, h_{\hat{\eta}}) = \tilde{J}(\alpha) + T_1(h_z) + T_2(h_v, h_\eta) + R_1(\alpha, h_v, h_\eta) + R_2(\alpha, h_v, h_\eta, h_z), \quad (2.20)$$

where

$$T_1(h_z) \equiv \frac{1}{n} \sum_{i=1}^{n} \left\{ \hat{g}_{x,i}(Z_i) - g_x(Z_i) \right\}^2 = IMSE_1(h_z) + R_3(h_z)$$

$$T_2(h_v, h_\eta) \equiv \frac{1}{n} \sum_{i=1}^{n} \left\{ \hat{m}_i(V_{0i}, \eta_i) - m(V_{0i}, \eta_i) \right\}^2 = IMSE_2(h_v, h_\eta) + R_4(h_v, h_\eta)$$

$$\sup_{\alpha \in A_n, h_v, h_\eta \in \mathcal{H}_n} |R_1(\alpha, h_v, h_\eta)| = o_p(n^{-1/2}), \quad \sup_{\alpha \in A_n, h_v, h_\eta, h_z \in \mathcal{H}_n} |R_2(\alpha, h_v, h_\eta, h_z)| = o_p(n^{-1/2})$$

with $\hat{m}_i(\cdot)$ *and* $\hat{g}_{x,i}(\cdot)$ *being the leave-one-out local constant estimators of* $m(\cdot)$ *and* $g_x(\cdot)$, *respectively. More importantly*

$$\tilde{J}(\alpha) = \frac{1}{n}\sum_{i=1}^{n}\{W_i - U_i'\beta\}^2$$

where $W_i \equiv Y_i - E(Y_i|V_i,\eta_i)$ *and* $U_i \equiv X_i - E(X_i|V_i,\eta_i)$. *Furthermore,* $\sup_{h_z\in\mathcal{H}_n}|R_3(h_z)| = o_p(n^{1/5})$ *and* $\sup_{h_v,h_\eta\in\mathcal{H}_n}|R_4(h_v,h_\eta)| = o_p(n^{1/5})$ *because they do not depend on* α.

The results of Theorem 2.1 show the attractive properties of our proposed CF approach. Similar to the results of Härdle et al. (1993) and Xia et al. (1999), Theorem 2.1 shows that the properties of the bandwidth parameter estimators can be studied while assuming α_0 is known. Moreover, the asymptotically optimal bandwidth parameters for estimating the $m(\cdot)$ function are assumed to be used for the $\sqrt{n}$-consistent estimation of α_0. In addition, under-smoothing is not required in estimating the first-stage reduced-form equation, as already stated in Remark 2.1. In particular, Theorem 2.1 suggests that minimizing $\hat{J}(\alpha, h_v, h_{\hat{\eta}})$ simultaneously with respect to α, h_v and $h_{\hat{\eta}}$, is asymptotically equivalent to separately minimizing $\tilde{J}(\alpha)$ with respect to α, $T_1(h_z)$ with respect to h_z, and $T_2(h_v, h_\eta)$ with respect to h_v and h_η, assuming that α_0 and η are known. This is because the remainder terms, namely $R_1(\alpha, h_v, h_\eta)$ and $R_2(\alpha, h_z, h_v, h_\eta)$, are shown to be asymptotically negligible.

Next, the asymptotic properties of $\hat{\alpha}$ and $\hat{\beta}$ are shown as a corollary of Theorem 2.1, given that $\Phi_{U_0} = [\{X - E(X|V_0,\eta)\}\{X - E(X|V_0,\eta)\}']$.

Corollary 2.1. *Under the assumptions of Theorem 2.1, the asymptotic properties of* $\hat{\alpha}$ *and* $\hat{\beta}$ *are as follows*

$$\sqrt{n}(\hat{\beta} - \beta_0) \to_D N(0, Var_1), \tag{2.21}$$

where $Var_1 = \sigma^2\left[\Phi_{U_0}^{-} - \left(m_0^{(1)}\Phi_{U_0}\right)^{-}\Phi_{U_0}\left\{m_0^{(1)}\right\}^2\left(m_0^{(1)}\Phi_{U_0}\right)^{-}\right]$, *and*

$$\sqrt{n}(\hat{\alpha} - \alpha_0) \to_D N(0, Var_2), \tag{2.22}$$

where $Var_2 = \sigma^2\left[\left\{\left(m_0^{(1)}\right)^2\Phi_{U_0}\right\}^{-} - \left\{m_0^{(1)}\Phi_{U_0}\right\}^{-}\Phi_{U_0}\left\{m_0^{(1)}\Phi_{U_0}\right\}^{-}\right]$.

Finally, the asymptotic properties of $\hat{g}(\hat{v})$ are presented in Theorem 2.2 below.

Theorem 2.2. *Under the assumptions of Theorem 2.1, and* $\inf_{z\in\mathcal{A}_z} f_z(z) > 0$ *and* $\inf_{x,\eta\in\mathcal{U}} f(v_0,\eta) > 0$*, the asymptotic results of* $\hat{g}(\hat{v})$ *are as follows*

$$\sqrt{nh_v}\left(\hat{g}(\hat{v}) - g(v_0) - Bias\right) \to_D N(0, Var),$$

where $Bias = h_v^2 B_v(v_0,\eta) + h_\eta^2 B_\eta(v_0,\eta)$ *and* $Var = f_\alpha(v_0)\mathcal{K}_v \int \frac{\sigma^2(V_0,\eta) f_\eta^2(\eta)}{f^2(v_0,\eta)} dQ(\eta)$*, with* $f_\alpha(v_0)$ *and* $f_\eta(\eta)$ *denoting the density functions of* V_0 *and* η*, respectively.*

The mathematical proofs of Theorems 2.1 and 2.2 and Corollary 2.1 are given in the Appendix.

Remark 2.2. *In these results, it is clear the first-stage nonparametric estimation does not contribute to the asymptotic variance of the estimators in the final-stage because the contribution of first-stage nonparametric estimation is asymptotically negligible. This characteristic is common among multi-stage nonparametric estimation procedures (see Su and Ullah (2008) for an example). However, this differs from the work of Li and Wooldridge (2002), which considers parametrically generated covariates in a PL semiparametric regression model.*

Remark 2.3. *It is also interesting to explore the case of performing the CF approach without the presence of nonparametric-endogeneity. The essential stochastic assumption of the CF approach ((2.9)) implies no existence of any endogeneity control function and, hence there is no identification problem in estimating the structural link function. Therefore, performing the CF approach without the presence of endogeneity causes an unnecessary multi-stage nonparametric estimation and the presence of redundant covariates in estimating the structural link function. However, the theoretical results of the proposed estimators particularly Theorems 2.1 and 2.2 and Corollary 2.1, are still valid with minor modifications, especially in terms of* $IMSE_2(h_v,h_\eta)$*,* Var_1 *and* Var_2*, and the bias and the variance of* $\hat{g}(\hat{v})$*. The minor modifications of the theoretical results are as follows*

$$IMSE_2(h_v,h_\eta)^* \asymp \int \left\{ \left[B_v^*(v_0,\eta) h_v^2 h_\eta^2\right]^2 + \frac{\mathcal{K}}{nh_v h_\eta} \frac{\sigma^{*2}(V_0,\eta)}{f(v_0,\eta)} \right\} f(x,\eta)\,dx\,d\eta$$

$$Bias^* = h_v^2 B_v^*(v_0,\eta) \text{ and } Var^* = f_\alpha(v_0)\mathcal{K}_v \int \frac{\sigma^{*2}(V_0) f_\eta^2}{f^2(v_0,\eta)} dQ(\eta),$$

where $B_v^*(v_0,\eta) = \frac{\mathcal{K}_{v,2}}{2f(v_0,\eta)}\left\{2f_{v_0}(v_0,\eta)g^{(1)}(v_0) + f(v_0,\eta)g_{v_0}^{(2)}(v_0,\eta)\right\}$ *and* $\sigma^{*2} = E(\epsilon^2|X,\eta) = E(\epsilon^2|X)$, *and* Var_1^* *and* Var_2^* *are obtained by replacing* $m_0^{(1)}$ *with* $g_0^{(1)}$ *in (2.19) and (2.20) with* $g_0^{(1)}$ *being the first derivative of* $g(v_0)$ *with respect to* V_0.

Remark 2.4. *Our results can also be extended to a more general data structure where a random sample* $\{(X_t', Z_t', Y_t);\ t = 1,\ldots,n\}$ *is a strictly stationary and strongly mixing process under Assumptions 2.6 and 2.7 below in addition to 2.1 to 2.5 above.*

In the rest of this section, we discuss about how to extend these established theoretical results to stationary time series data as in Remark 2.4. First, let $\xi_t \equiv (X_t'\alpha_0, \eta_t)$ and $f_\xi(\xi)$ denote the joint density function of $X'\alpha_0$ and η. The necessary regularity conditions for the strictly stationary and α-mixing case are then as follows.

Assumption 2.6. *(i) The conditional densities satisfy the following conditions:*

$$f_{\xi_1,\xi_l|X_1,X_l}(\xi_1,\xi_l) \le C < \infty;\ f_{\xi_1,\xi_l|Y_1,Y_l}(\xi_1,\xi_l) \le C' < \infty;\ f_{Z_1,Z_l|X_1,X_l}(Z_1,Z_l) \le C'' < \infty$$

for some constants $C, C', C'' > 0$ *and for all* $l \ge 1$. *(ii) The mixing and moment conditions are as follows:*

$$\sum_l l^a[\alpha(l)]^{1-2/l} < \infty,\ E||X_0||^l < \infty \text{ and } f_{\xi_1|X_1}(\xi|X) \le C < \infty;$$

$$\sum_l l^{a'}[\alpha(l)]^{1-2/l} < \infty,\ E|Y_0|^l < \infty \text{ and } f_{\xi_1|Y_1}(\xi|Y) \le C' < \infty;$$

$$\sum_l l^{a''}[\alpha(l)]^{1-2/l} < \infty,\ E||Z_0||^l < \infty \text{ and } f_{Z_1|X_1}(z|X) \le C'' < \infty,$$

where $l > 2$ *and* $a, a', a'' > 1 - 2/l$. *(iii) There is a sequence of positive integer* s_T, *which satisfies* $s_T \to \infty$ *and* $s_T = o\left\{(nh_{z,T}^{q_z})^{1/2}\right\}$, *such that* $(n/h_{z,T}^{q_z})^{1/2}\alpha(s_T) \to 0$ *as* $T \to \infty$.

Assumption 2.7. *(i) Let the density functions* $f_z(z)$ *and* $f(v_0,\eta)$ *satisfy* $\inf_{z\in\mathcal{A}_z} f_z(z) > 0$ *and* $\inf_{x,\eta\in\mathcal{U}} f(v_0,\eta) > 0$. *(ii) In addition, we require the following moment conditions:*

$$E||X||^s < \infty,\ \sup_{\xi\in\mathcal{U}}\int ||X||^s f(x,\xi)dx,\ E|Y|^s < \infty,\ \sup_{\xi\in\mathcal{U}}\int |Y|^s f(y,\xi)dy;\ \sup_{z\in\mathcal{A}_z}\int ||X||^s f(x,z)dx,$$

for some $s > 2$. (iii) The bandwidth sequences, h_v, h_η and h_z, tend to zero as $n \to \infty$ and satisfy the following, for some $\delta > 0$

$$T^{1-2s^{-1}-2\delta} h_z^{q_z} \to \infty;\ T^{1-2s^{-1}-2\delta} h_v h_\eta \to \infty; T^{1-2s^{-1}-2\delta} \left(h_z^{q_z} h_v h_\eta^3\right)^{1/2} \to \infty.$$

In the proof of the $\sqrt{n}$-consistency of $\hat{\alpha}$ and $\hat{\beta}$ in the case of Remark 2.4, Propositions A.1 to A.15 in the Appendix encompass the extra covariance terms caused by the serial dependences in the sample. Under Assumptions 2.1 to 2.5 and 2.6(i)(ii), the covariance terms can be shown to be $o_p(n^{-1/2})$. For instance, the extra covariance term in Proposition A.1 might be derived as $\sum_{l=1}^{n-1}(1-t/n)\mathrm{Cov}(\hat{\varphi}_1, \hat{\varphi}_{l+1}) = o(h_v h_\eta)$. However, the consistency of $\hat{g}(\hat{v})$ requires stronger conditions than the case of $\hat{\alpha}$ and $\hat{\beta}$, namely the uniform convergence of $\hat{f}(v_0, \eta)$, which requires the uniform convergence of Q_j, where $j = 1, \cdots, 5$ in (B.1) in the Appendix. Under Assumptions 2.1 to 2.5, 2.6(i)-(ii) and 2.7, Q_j is shown to be $o_p(1)$ as follows

$$\sup_{\xi \in \mathcal{U}, z \in A_z} |Q_{2i}| = \sup_{\xi \in \mathcal{U}, z \in A_z} |Q_{5i}| = O_p \left\{ \left(\frac{(\ln n)^2}{n^2 h_z^{q_z} h_v h_\eta^3} \right)^{1/2} + h_z^2 (h_v^2 + h_\eta^2) \right\}.$$

Furthermore, the asymptotic normality of $\hat{g}(\hat{v})$ is then obtained by applying Assumption 2.6 (iii) for the standard nonparametric small-block and large-block arguments. Nonetheless, the asymptotic normalities of $\hat{\alpha}$ and $\hat{\beta}$ are obtained by applying parts of Assumption 2.6 (ii), namely $\sum_l l^a [\alpha(l)]^{1-2/l} < \infty$, $E||X_0||^l < \infty$, $\sum_l l^{a'} [\alpha(l)]^{1-2/l} < \infty$ and $E|Y_0|^l < \infty$, to (A.6) and (A.10) in the Appendix for the small-block and large-block arguments of a standard strictly stationary and strongly mixing process.

2.4. Simulation Studies

In this section[5], the finite-sample performances of the estimation procedure proposed above are investigated by making a comparison between the performances of the estimation method introduced in Xia et al. (1999), referred to the XTL procedure, and the CF approach established in Section 2.2 as the KS procedure in the presence of nonparametric-endogeneity. Throughout this section, optimization is

[5]The results of extensive simulation exercises for GPLSI model are available on request from the author.

implemented by using a limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm for the bound-constrained optimization of Byrd et al. (1995). All simulation exercises are conducted in R with the Gaussian kernel function and the number of replications $Q = 200$. To compare and evaluate the finite sample performances of the procedures, the mean and mean absolute errors of the estimates of both coefficients, α_0 and β_0, across Q replications are computed in Tables 2.1 and 2.2. The averaged absolute error of the estimates of the unknown structural function is also computed as follows

$$\mathrm{ae}_{\hat{g}} = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{g}(\hat{V}_i) - g(V_{0i})\right|,$$

where n is the number of samples.

In the analysis that follows, an example model of the following form is considered

$$Y_i = \beta_{01}X_{1i} + \beta_{02}X_{2i} + \beta_{03}X_{3i} + g(V_{0i}) + \epsilon_i, \tag{2.23}$$

where $V_0 = \alpha_{01}X_1 + \alpha_{02}X_2 + \alpha_{03}X_3$, $g(V_0) = \exp\{-2(\alpha_{01}X_1 + \alpha_{02}X_2 + \alpha_{03}X_3)^2\}$, and X_j is independently and uniformly distributed on $[-1, 1]$ for $j = 1, 2$. It is necessary that $\beta_0 \perp \alpha_0$ with $\parallel \alpha_0 \parallel = 1$. In order for these conditions to be satisfied, define $\beta_{02} = 0.4$, $\beta_{03} = 0$, $\alpha_{01} = 0.7$, $\alpha_{02} = -0.6$, then β_{01} and α_{03} are defined as follows

$$\alpha_{03} = \sqrt{1 - \alpha_{01}^2 - \alpha_{02}^2} \quad \text{and} \quad \beta_{01} = -\frac{\beta_{02}\alpha_{02}}{\alpha_{01}}.$$

In this example, nonparametric-endogeneity is introduced by letting $X_3 = Z + \eta$, where Z and η are independently and uniformly distributed on $[-0.5, 0.5]$ and $[-1, 1]$, respectively, and $\epsilon = \eta + e$ and e is independent and standard normally distributed. Tables 2.1 and 2.2 present the estimation results from the XTL and KS procedures, respectively.

The simulation results in Table 2.1 show strong evidence against the use of the XTL procedure in the presence of endogeneity. This evidence is clear when the averaged absolute errors, $\mathrm{ae}_{\hat{g}}$ in Table 2.1 are considered. On the other hand, the simulation results in Table 2.2 suggest that the KS procedure is able to identify the structural link function, namely the $g(\cdot)$ function, in the presence of endogeneity.

Table 2.1. *EGPLSI model with nonparametric-endogeneity and the XTL's procedure.*

n	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$
50	0.3130	0.4332	0.8884	-0.7748	0.5597
150	0.3088	0.4340	0.8993	-0.7671	0.5279
300	0.3142	0.4264	0.8988	-0.7674	0.5225
500	0.3135	0.4288	0.8960	-0.7653	0.5179

n	$\|\hat{\beta}_1 - \beta_{01}\|$	$\|\hat{\beta}_2 - \beta_{02}\|$	$\|\hat{\alpha}_1 - \alpha_{01}\|$	$\|\hat{\alpha}_2 - \alpha_{02}\|$	$\|\hat{\alpha}_3 - \alpha_{03}\|$	$ae_{\hat{g}}$
50	0.0656	0.0714	0.1691	0.1253	0.1586	0.0905
150	0.0428	0.04572	0.0859	0.0559	0.0910	0.0891
300	0.0331	0.03377	0.0629	0.0548	0.0426	0.0895
500	0.0306	0.0319	0.0229	0.0156	0.0181	0.0906

Table 2.2. *EGPLSI model with nonparametric-endogeneity and the KS procedure.*

n	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$
50	0.2645	0.4652	0.9638	-0.8249	0.5483
150	0.3260	0.4135	0.8975	-0.7852	0.4756
300	0.3486	0.3945	0.8090	-0.6997	0.4382
500	0.3555	0.3891	0.7353	-0.6295	0.3992

n	$\|\hat{\beta}_1 - \beta_{01}\|$	$\|\hat{\beta}_2 - \beta_{02}\|$	$\|\hat{\alpha}_1 - \alpha_{01}\|$	$\|\hat{\alpha}_2 - \alpha_{02}\|$	$\|\hat{\alpha}_3 - \alpha_{03}\|$	$ae_{\hat{g}}$
50	0.0816	0.0684	0.1678	0.1389	0.1195	0.0632
150	0.0307	0.0264	0.1244	0.0962	0.0769	0.0265
300	0.0213	0.0183	0.0446	0.0327	0.0285	0.0160
500	0.0189	0.0159	0.0416	0.0319	0.0263	0.0124

3. Semiparametric CF approach to Shape-Invariant Empirical Engel Curves

In this section, a flexible shape-invariant Engel curve system is analyzed within the framework of the EGPLSI model with the newly developed CF approach above. The consumer optimization theory in the empirical demand study literature suggests including a scale and shift parameters within a flexible shape-invariant empirical Engel curve in order to incorporate the individual household heterogeneity (see Pendakur (1999), Blundell and Powell (2003) and Blundell et al. (2007) for examples). In addition, it is also well-known that modelling a shape-invariant Engel curve system involves a critical difficulty, which resides in the endogeneity of total expenditure caused by the two-stage budgeting model (see Blundell et al. (1998) and Blundell et al. (2007) for details). Hence, it is natural to study a shape-invariant Engel curve system within the framework of the EGPLSI model with the newly developed CF approach.

3.1. The Empirical Model and Estimation

Hereafter, let $\{Y_{il}, X_{1i}, X_{2i}\}_{i=1}^{n}$ represent an i.i.d. sequence of n household observations on the budget share Y_{il} of good $l = 1, \ldots, L \geq 1$ for each household i facing the same relative prices, the log of total expenditure X_{1i}, and a vector of household composition variables X_{2i}. For each commodity l, budget shares and total outlay are related by a general stochastic Engel curve, namely $Y_l = G_l(X_1) + \epsilon_l$, where $G_l(\cdot)$ is an unknown function that can be estimated by using a standard nonparametric regression method under the exogeneity assumption of total expenditure (i.e. $E(\epsilon_l|X_1) = 0$). Nonetheless, a number of previous studies have reported that household expenditures typically display great variation with demographic composition. A simple approach for estimating the model is to stratify the data by each distinct discrete outcome of X_2 and then carry out our estimation with nonparametric smoothing within each cell. At some point, however, it may be useful to pool the Engel curves across different household demographic types and to allow X_1 to enter each Engel curve semiparametrically. This idea leads to the specification below

$$Y_{il} = \beta_{0l}' X_{2i} + g_l(X_{1i} - \phi(\gamma_0' X_{2i})) + \epsilon_{il}, \tag{3.1}$$

where $g_l(\cdot)$ is an unknown function and $\phi(\gamma_0' X_{2i})$ is a known function up to a finite set of unknown parameters γ_0, which can be interpreted as the log of general equivalence scales for household i. In the current paper, $\phi(\gamma_0' X_{2i}) = \gamma_0' X_{2i}$ is chosen so that (3.1) is specified as follows

$$Y_{il} = \beta_{0l}' X_{2i} + g_l(X_{1i} - \gamma_0' X_{2i}) + \epsilon_{il}. \tag{3.2}$$

In this application, total expenditure is allowed to be endogenous and a measure of earning of the head of each household is used as an instrument.

Following the CF approach discussed above, the empirical model to be estimated is of the form below

$$Y_{il} = \beta_{01,l} X_{1i} + \beta_{0l}' X_{2i} + g_l(\alpha_{01} X_{1i} + \alpha_{02}' X_{2i}) + \epsilon_{il} \tag{3.3}$$

$$X_{1i} = m_{X1}(Z_i) + \eta_i, \text{ where } E(\eta|Z) = 0 \tag{3.4}$$

$$E(\epsilon_l|Z, \eta) = E(\epsilon_l|\eta) \neq 0, \tag{3.5}$$

where $m_{X1}(Z) = E(X_1|Z)$ and $\{Z_i\}_{i=1}^{n}$ represents an i.i.d. sequence of the measure of earning of n heads of households and (3.3) is a semiparametric model that satisfies all the identification conditions required in the construction of the EGPLSI model. The theoretically consistent model in (3.1) can then be solved based on (3.3). For this end, a similar scaling transformation to that explained in Section 2.1 is used. In the remainder of this section, some specific details about the estimation procedure are discussed. Rather than basing our discussion on (3.3) to (3.5), it is statistically more equivalent to do so based on the following

$$Y_{il} = \beta_{0l}' X_{2i} + g_l(X_{1i} - \gamma_0' X_{2i}) + \epsilon_{il} \tag{3.6}$$

$$X_{1i} = m_{X1}(Z_i) + \eta_i, \text{ where } E(\eta|Z) = 0 \tag{3.7}$$

$$E(\epsilon_l|Z,\eta) = E(\epsilon_l|\eta) \neq 0. \tag{3.8}$$

These models suggest the conditional expectation relationship shown below

$$E\left(Y_l|(X_1 - \gamma_0' X_2), \eta\right) - \beta_{0l}' E\left(X_2|(X_1 - \gamma_0' X_2), \eta\right) = g_l(X_1 - \gamma_0' X_2) + \iota_l(\eta), \tag{3.9}$$

where $E\left(\epsilon_l|(X_1 - \gamma_0' X_2), \eta\right) = E\left(\epsilon_l|\eta\right) \equiv \iota_l(\eta) \neq 0$, which immediately leads to

$$Y_{il} = \beta_{0l}' X_{2i} + g_l(X_{1i} - \gamma_0' X_{2i}) + \iota_l(\eta_i) + e_{il}, \tag{3.10}$$

$$X_{1i} = m_{X1}(Z_i) + \eta_i, \tag{3.11}$$

where $E(e_l|X_1, X_2, \eta) = 0$. Let $m_l\left(\{X_{1i} - \gamma_0' X_{2i}\}, \eta_i\right) = g_l(X_{1i} - \gamma_0' X_{2i}) + \iota_l(\eta_i)$. In order to use (3.10), it is important to note that

$$\begin{aligned} m_{1,l}(X_1 - \gamma_0' X_2) &= \int m_l(\{X_1 - \gamma_0' X_2\}, \eta)\, d\eta \\ g_l(X_1 - \gamma_0' X_2) &= m_{1,l}(X_1 - \gamma_0' X_2) - C, \end{aligned} \tag{3.12}$$

where $C = \int \iota(\eta) dQ(\eta)$ and $E(g_l(\cdot)) = 0$.

If a linear specification is imposed on $\iota_l(\cdot)$, (3.10) would be similar to the extended partially linear model discussed in Blundell et al. (1998). In this case, Blundell et al. (1998) showed that a test of the endogeneity null can be constructed by testing $H_0 : \iota_l = 0$, where ι_l is an unknown parameter. To allow for more flexibility in the functional form between total expenditure and its instrument, as an alternative, one may apply an existing test of a parametric mean-regression model against a nonparametric alternative (see Horowitz and Spokoiny (2001), for example). However, the current paper suggests that it is more convenient to simply construct the variability bands for $\iota_l(\cdot)$ since its estimate is readily available. To do so, the following procedure is used.

Step 3.1.1: Obtain an empirical estimate of $g_l(X_1 - \gamma_0' X_2)$ in (3.12).

Step 3.1.2: Regress (3.10) by using the estimates in Step 3.1.1 to obtain the nonparametric estimates of $\iota_l(\cdot)$.

Step 3.1.3: Compute the bias-corrected confidence bands for the nonparametric smoothing using the procedure introduced by Xia (1998). Finally, the Bonferroni-type variability bands are obtained by using a similar procedure to that discussed by Eubank and Speckman (1993).

To perform Step 3.1.1, the estimation procedure introduced in Section 2 is used. However, some modifications are required to take the vector of index coefficient, γ_0, a general equivalence scale for household i, into account. In this case, the objective function (2.16) is only used for a particular commodity l. The new objective function, $\min_{\gamma \in A_n, h_{v,l}, h_{\hat{\eta},l} \in \mathcal{H}_n} \hat{J}(\gamma, h_{v,l}, h_{\hat{\eta},l})$, is the summation of these individual functions that is minimized with respect to γ and 14 smoothing parameters, particularly two for each commodity. Finally, the estimation procedure is completed by using $\hat{\gamma}$ as well as $\hat{h}_{\hat{v},l}$ and $\hat{h}_{\hat{\eta},l}$.

In addition, the model in (3.10) can also be re-stated as

$$Y_{il}^* = g_l(X_{1i} - \gamma_0' X_{2i}) + e_{il}, \tag{3.13}$$

where $Y_l^* \equiv Y_l - \beta_{0l}' X_2 - \iota_l(\eta)$. The use of (3.13) relies on

$$m_{2,l}(\eta) \;=\; \int m_l(v, \eta)\, dv = \iota_l(\eta) + C' \text{ and } \iota_l(\eta) = m_{2,l}(\eta) - C', \tag{3.14}$$

where $V = X_1 - \gamma' X_2$, $C' = \int g(v) dQ(v)$ and $E(\iota_l(\cdot)) = 0$, which corresponds to (3.12) above. Hence, the model in (3.13) suggests that the estimates of the shape-invariant Engel curves and the related confidence bands are obtained as follows.

Step 3.2.1: Obtain empirical estimates of $\iota_l(\eta)$ in (3.14).

Step 3.2.2: Regress (3.13) using the estimates in Step 3.2.1 to obtain the nonparametric estimates of $g_l(\cdot)$.

Step 3.2.3: Compute the bias-corrected confidence bands about the nonparametric estimator in Step 3.2.2 by using the procedure introduced by Xia (1998).

3.2. The Engel Curve Data

In our application, the data set is drawn from the British Family Expenditure Survey (FES) 1995-96. Seven broad categories of goods are considered as follows: (1) fuel, light and power (fuel hereafter); (2) fares, other travel costs and running motor vehicles (fares); (3) food; (4) alcoholic drink and tobacco (alcohol); (5) leisure goods and services (leisure goods); (6) clothing and footwear (clothing) and (7) personal goods and services (personal goods).

Table 3.1. *Descriptive statistics.*

	Couples with 1 or 2 children		**Couples without children**	
	Mean	**Std. Dev**	**Mean**	**Std. Dev**
Budget shares:				
Fuel	0.0692	0.0011	0.0618	0.0012
Fares	0.1537	0.0025	0.1715	0.0031
Food	0.3235	0.0028	0.2768	0.0031
Alcohol	0.0844	0.0022	0.1144	0.0031
Leisure goods	0.2155	0.0038	0.2298	0.0045
Clothing	0.0926	0.0024	0.0872	0.0029
Personal goods	0.0606	0.0016	0.0581	0.0019
Expenditure and income:				
log (total expenditure)	5.4374	0.0130	5.4524	0.0161
log (income)	5.9205	0.0153	6.0397	0.0166
Sample size	1072		1278	

To maintain some demographic homogeneity, a subset of married or cohabiting couples are selected from the FES, particularly categories 1 and 3 of the variable *ms* in the table *adult*. In addition, those where the head of household is aged between 20 and 55 (i.e. the variable *age* in the table *adult*) and in work (i.e. excluding the category 1 of the variable *fted* in the table *adult* and category 6 of the variable *a093* in the table *set8*) are considered. Finally, all households with three or more children are excluded. Our demographic variable, X_2, is a binary dummy variable that reflects whether a couple has 1 or 2 children (where $X_2 = 1$) or no children (where $X_2 = 0$). Overall, there are 2350 observations, 1278 are couples with one or two children. Table 3.1 shows larger expenditure shares for fuel, food, clothing and personal goods for the households with children as expected. Also as expected, households without children are able to spend higher proportions of their total expenditure on alcohol and leisure goods. Overall, there are clear differences in the consumption patterns between the two demographic groups. The estimates of the scale and the shift coefficients are expected to reflect these differences.

Furthermore, the log of total expenditure on the nondurables and services is our measure of the continuous endogenous explanatory variable, X_1. In our analysis that follows, the log of normal weekly disposable head of household income, specifically the variable *p389* of the table *set3*, is used as an instrument. The two variables show a strongly-positive correlation with correlation coefficients of 0.5660 and 0.5954 for couples with and without children, respectively. Figures 3.1 and 3.2 present plots of the kernel estimates of the joint density for these variables. Finally, in the empirical application the instrument variable $Z = \Phi(\text{log earnings})$ is taken, similar to Blundell et al. (2007).

Figure 3.1. *Kernel joint density estimates for the log of total expenditure and the log of weekly income for couples with 1 or 2 children.*

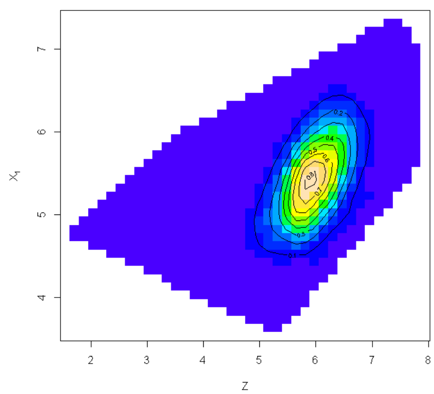
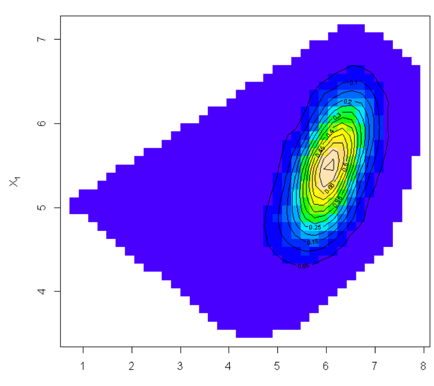


Figure 3.2. *Kernel joint density estimates for the log of total expenditure and the log of weekly income for couples without ch*



3.3. Empirical Findings

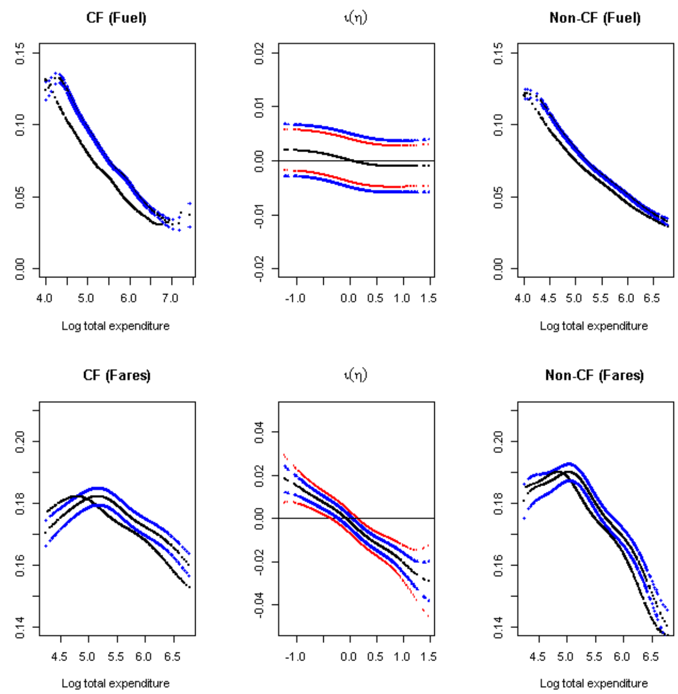
The important empirical findings are now presented and summarized in Table 3.2. Although exact definitions of the data are not given in Blundell et al. (1998), Blundell et al. (1998) estimated the shape-invariant Engel curves for four broad categories of nondurables and services by using the FES data, namely fuel, fares, alcohol and leisure, similar to this paper. The empirical estimate, $\hat{\gamma}$, of 0.36355 reported in the first column is very close to 0.3698 as found in Blundell et al. (1998). Furthermore, the signs of the parameter estimates, $\hat{\beta}_l$, for the four broad categories are all consistent with those of Blundell et al. (1998); specifically they are positive for food and leisure, but negative for alcohol, fares and fuel.

Table 3.2. *Empirical results*

$\hat{\gamma}$	Categories of goods	$\hat{\beta}_l$	$\hat{h}_{v,l}$	$\hat{h}_{\hat{\eta},l}$
0.36355	Fuel, light and power	-0.01401	0.14021	0.93631
	Fares, other travel costs and running of motor vehicles	-0.02027	0.19545	0.26831
	Food	0.00537	0.15120	0.25826
	Alcoholic drink and tobacco	-0.05205	0.30802	0.22569
	Leisure goods and services	0.05077	0.14663	0.40277
	Clothing and footwear	0.02079	0.14846	0.27234
	Personal goods and services	0.00738	0.49331	0.49335

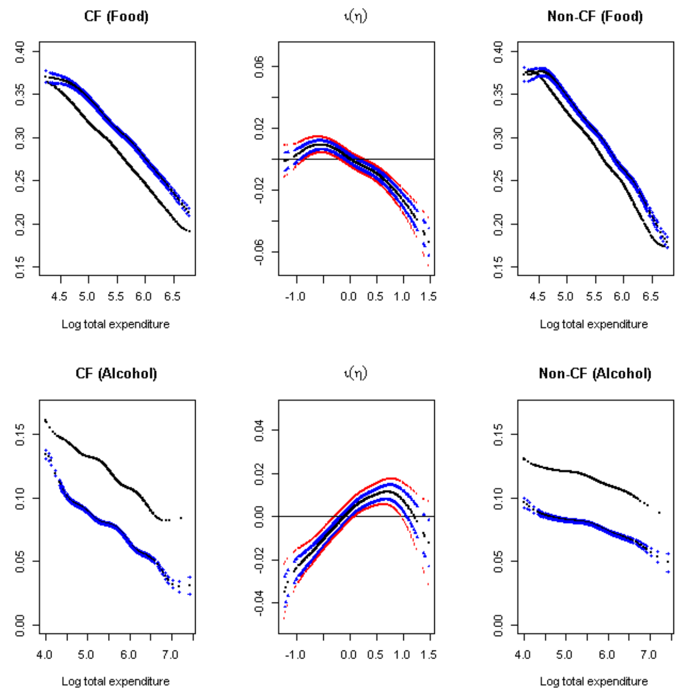
The first columns of Figures 3.3 to 3.6 present the empirical estimates of the Engel curves for seven of the goods in our system based on the CF approach discussed in Section 3.1. For these plots, the smoothing parameters presented in the fourth and fifth columns of Table 3.2 are used. Furthermore, the third columns of these figures show the empirical estimates of the Engel curves computed from the Xia et al. (1999)'s procedure by which the exogeneity assumption is imposed on the total expenditure. Together with the estimated Engel curves, their 90% point-wise confidence bands are also reported. The bands are obtained by using the procedure discussed in Section 3.1. Let us now concentrate on the first columns. For fuel, food and alcohol, the Engel curves appear to demonstrate that the Working-Leser linear logarithmic formulation may provide a reasonable approximation. Nonetheless, for other shares, especially for fares, a nonlinear relationship between the shares and the log of expenditure is evident. A detailed investigation of the data shows that on average, up to 70% of fares belongs to running motor vehicles. Hence, motor vehicles seemed to be a necessity good for a household for which the log of total expenditure is more than around 5.3 for those with children, for those without children, it is up to around 4.8. It seemed that motor vehicles are a superior good for households where the log of total expenditure, is below these levels. The estimated shares for the couples with children are higher than those for couples without children, except for extreme lower quantile of the log of total expenditure. This could lead to the nonlinear relationship witnessed in Figure 3.3.

Figure 3.3. *Fuel and fares (90% confidence bands are drawn for households with children)*



As expected, the estimated shares of fuel and food for households with children are consistently above those for households without children. Couples without children spend around 3% more of their budget on fuel and food than couples with children. In addition, the estimated shares of alcohol, leisure, clothing and personal goods for households with children are consistently below those for households without children. Couples with children spend around 3%, 8% and 2% more of their budget on leisure, clothing and personal goods than couples with children at the same level of expenditure. In all but one case (i.e. fares), there seem to be a broadly parallel shift in the Engel curves from one demographic group to another. Our results suggest that fuel, food and alcohol may be categorized as necessity goods in the sense that the demand for these goods increases proportionally less than the increase in total expenditure. These goods whose demand increases with total expenditure are leisure, clothing and personal goods.

Figure 3.4. *Food and alcohol (90% confidence bands are drawn for households with children)*

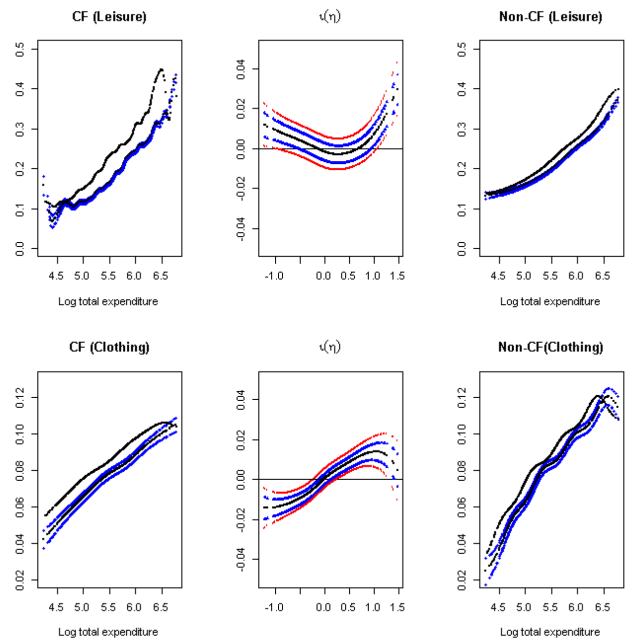


The second column presents the nonparametric estimates of the control functions, $\iota_l(\cdot)$. With the estimated control functions, the two sets of bands, namely the 90% bias-corrected confidence bands for the nonparametric smoothing of Xia (1998) (blue) and the 90% Bonferroni-type variability bands of Eubank and Speckman (1993) (red) are also reported. Regarding fuel and personal goods, $\iota_l(\cdot)$ for these cases do not seem statistically significant. However, the opposite is found for fares, food, leisure and clothing. Hence, neglecting potential endogeneity in the estimation can lead to incorrect estimates of the shape of Engel curves for these goods. This can be seen by comparing the first and the third columns of the figures. For these goods it is clear that the curvature changes significantly as the presence of the endogeneity is allowed.

4. Conclusion

In this paper, the usefulness of the EGPLSI model in its ability to model a flexible shape-invariant specification is elaborated. A shape-invariant specification is beneficial for analyzing an aggregate structural relationship, taking individual heterogeneity into account. A flexible shape-invariant specification is easily studied within the EGPLSI framework because both the scale and shift parameters are easily

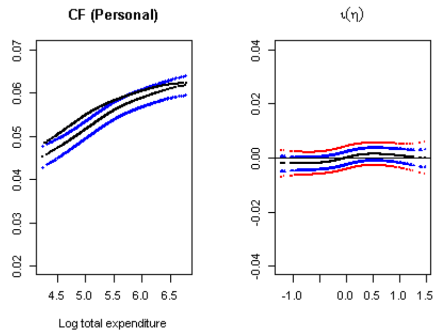
Figure 3.5. *Leisure and clothing (90% confidence bands are drawn for households with children)*



incorporated in the EGPLSI model. Despite the benefits mentioned above, the applicability of the EGPLSI model to an empirical study is limited because of its shortfalls in addressing endogeneity problems. Hence, the current paper develops the CF approach to address the endogeneity problem in the EGPLSI model to enhance its applicability to an empirical study.

The proposed CF approach inherits a few intrinsic features. Firstly, it resembles existing multi-stage nonparametric estimation procedures in the sense that the endogeneity control covariates must be estimated from the first-stage reduced-form equation. Furthermore, the involvement of the nonparametrically generated covariates means that establishment of the CF approach is not straightforward. The optimization technique of Xia et al. (1999) needs to be extended one step further to ensure its theoretical validity. The current paper shows that under-smoothing is not required in the first-stage of our proposed estimation procedure under the relatively mild conditions seen in the literature. The first-stage nonparametric estimation is shown to be statistically negligible. The paper then closes the theoretical discussion by providing an outline of the straightforward extension of the results based on an i.i.d. random sample to a strictly stationary and strongly mixing process. The paper also presents the satisfactory finite sample performance of identifying the structural link function in the EGPLSI model in the presence of nonparametric-endogeneity from a Monte Carlo simulation exercise.

Figure 3.6. *Engel curves for personal (90% confidence bands are drawn for households with children)*



Finally, the semiparametric analysis of a system of shape-invariant empirical Engel curves using the FES (1995-96) data-set within the framework of the EGPLSI model with our proposed CF approach is conducted. Not only are the findings interesting empirically but the accessible applicability of our proposed CF approach is also explored.

Additionally, the development of the CF approach in this paper also provides a foundation for addressing the presence of weak instruments in the EGPLSI model. Han (2012) discussed how the intuitive triangular structure of the CF approach in a simple nonparametric regression model translates the difficult problem (namely the presence of weak instruments in the first-stage reduced-form equation) into a much simpler one, particularly the multicollinearity problem in the second-stage structural equation. Hence it is plausible to develop the current paper further to the case of the presence of weak instruments in the EGPLSI model. However, a thorough investigation is required to examine a number of important issues, particularly examining the $\sqrt{n}$-consistent estimation of the finite-dimensional parameters, namely α_0 and β_0, and the properties of the smoothing parameters in each stage of the proposed estimation procedure and, most importantly, how to address the presence of weak instruments in the relatively general semiparametric model, the EGPLSI model.

References

Ai, C., Chen, X., 2003. Efficient estimation of models with conditional moment restrictions containing unknown functions. Econometrica 71 (6), 1795–1843.

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

Blundell, R., Duncan, A., Pendakur, K., 1998. Semiparametric estimation and consumer demand. Journal of Applied Econometrics, 435–461.

Blundell, R., Powell, J. L., 2003. Endogeneity in nonparametric and semiparametric regression models. Econometric Society Monographs 36, 312–357.

Blundell, R., Stoker, T. M., 2007. Models of aggregate economic relationships that account for heterogeneity. Handbook of Econometrics 6, 4609–4666.

Blundell, R. W., Powell, J. L., 2004. Endogeneity in semiparametric binary response models. The Review of Economic Studies 71 (3), 655–679.

Byrd, R. H., Lu, P., Nocedal, J., Zhu, C., 1995. A limited memory algorithm for bound constrained optimization. SIAM Journal on Scientific Computing 16 (5), 1190–1208.

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

Eubank, R. L., Speckman, P. L., 1993. Confidence bands in nonparametric regression. Journal of the American Statistical Association 88 (424), 1287–1301.

Gao, J., 2007. Nonlinear Time Series: Semiparametric and Nonparametric Methods. CRC Press.

Gao, J., Lu, Z., Tjøstheim, D., et al., 2006. Estimation in semiparametric spatial regression. The Annals of Statistics 34 (3), 1395–1435.

Han, S., 2012. Nonparametric triangular simultaneous equations models with weak instruments. Working paper (UT Austin).

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

Härdle, W., Hall, P., Ichimura, H., 1993. Optimal smoothing in single-index models. The annals of Statistics 21 (1), 157–178.

Härdle, W., Marron, J. S., 1990. Semiparametric comparison of regression curves. The Annals of Statistics 18 (1), 63–89.

Hastie, T., Tibshirani, R., 1990. Generalized Additive Models. Chapman & Hall/CRC.

Horowitz, J. L., Spokoiny, V. G., 2001. An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. Econometrica 69 (3), 599–631.

Ichimura, H., 1993. Semiparametric least squares (sls) and weighted sls estimation of single-index models. Journal of Econometrics 58 (1-2), 71–120.

LaFree, G., Morris, N. A., Dugan, L., 2009. Cross-national patterns of terrorism: Comparing trajectories for total, attributed and fatal attacks, 1970–2006. The British Journal of Criminology 50 (4), 622–649.

Li, Q., Racine, J. S., 2007. Nonparametric Econometrics: Theory and Practice. Princeton University Press, Princeton, NJ.

Li, Q., Wooldridge, J. M., 2002. Semiparametric estimation of partially linear models for dependent data with generated regressors. Econometric Theory 18 (3), 625–645.

Linton, O., Härdle, W., 1996. Estimation of additive regression models with known links. Biometrika 83 (3), 529–540.

Linton, O., Nielsen, J. P., 1995. A kernel method of estimating structured nonparametric regression based on marginal integration. Biometrika 82 (1), 93–100.

Mammen, E., Rothe, C., Schienle, M., 2016. Semiparametric estimation with generated covariates. Econometric Theory 32 (5), 1140–1177.

Nagin, D. S., Odgers, C. L., 2010. Group-based trajectory modeling in clinical research. Annual Review of Clinical Psychology 6, 109–138.

Newey, W. K., Powell, J. L., Vella, F., 1999. Nonparametric estimation of triangular simultaneous equations models. Econometrica 67 (3), 565–603.

Pendakur, K., 1999. Semiparametric estimates and tests of base-independent equivalence scales. Journal of Econometrics 88 (1), 1–40.

Robinson, P., Pinkse, C., 1995. Pooling nonparametric estimates of regression functions with similar shape. In: Advances in Econometrics and Quantitative Economics. Wiley-Blackwell, 172–195.

Robinson, P. M., 1988. Root-n-consistent semiparametric regression. Econometrica 56 (4), 931–954.

Speckman, P., 1988. Kernel smoothing in partial linear models. Journal of the Royal Statistical Society. Series B (Methodological) 50 (3), 413–436.

Su, L., Ullah, A., 2008. Local polynomial estimation of nonparametric simultaneous equations models. Journal of Econometrics 144 (1), 193–218.

Tjøstheim, D., Auestad, B. H., 1994. Nonparametric identification of nonlinear time series: projections. Journal of the American Statistical Association 89 (428), 1398–1409.

Xia, Y., 1998. Bias-corrected confidence bands in nonparametric regression. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 60 (4), 797–811.

Xia, Y., Härdle, W., 2006. Semi-parametric estimation of partially linear single-index models. Journal of Multivariate Analysis 97 (5), 1162–1184.

Xia, Y., Tong, H., Li, W., 1999. On extended partially linear single-index models. Biometrika 86 (4), 831–842.

Appendix

In this Appendix, the mathematical proofs of the main theoretical results of the paper are presented. Note that the proofs in this section are the generalized version, namely the case where more than one endogeneous nonparametric covariates ($q_2 > 1$). The proofs of Theorem 2.1 and Corollary 2.1 are first discussed in two main steps, then the proofs of Theorem 2.2. follow.

For the sake of notational simplicity, let us first introduce the following terms; $m = m(v_0, \eta)$, $m_x = E(X_{\mathcal{A}_x}|v_0, \eta)$, $\tilde{m} = E(m|v, \eta)$, $\tilde{m}_x = E(X_{\mathcal{A}_x}|v, \eta)$, $L_{0,ij} = L_{h_{v_0}h_\eta}(V_{0i} - V_{0j}, \eta_i - \eta_j)$, $L_{ij} = L_{h_v h_\eta}(V_i - V_j, \eta_i - \eta_j)$, $L_{1,ij} = L_{h_v h_\eta}(\hat{V}_i - \hat{V}_j, \eta_i - \eta_j)$, $L_{2,ij} = L_{h_v h_\eta}(V_i - V_j, \hat{\eta}_i - \hat{\eta}_j)$ and $L_{3,ij} = L_{h_v h_\eta}(\hat{V}_i - \hat{V}_j, \hat{\eta}_i - \hat{\eta}_j)$. Let us also assume that $h_{\eta,1} = \cdots = h_{\eta,q_2} = h_\eta$ and $h_{z,1} = \cdots = h_{z,q_z} = h_z$ for the sake of simplicity.

Proofs of Theorem 2.1 and Corollary 2.1

Step 1. Proofs of Theorem 2.1: The proofs of Theorem 2.1 are based on the decomposition of (2.16) in a few interesting terms and by showing the uniform convergence of the remainder terms, namely R_1 and R_2. Let us first denote $\beta_0 - \beta = B$ in this step of the proofs. Given α, β and $\hat{\eta}$, the minimizing objective function in (2.20) is decomposed as shown below

$$\begin{aligned} \hat{J}(\alpha, h_v, h_{\hat{\eta}}) &\equiv \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i - \hat{\delta}_{Yi} - \left\{X_i - \hat{X}_i - \hat{\delta}_{Xi}\right\}'\beta\right)^2 \\ &= \hat{J}^*(\alpha, h_v, h_\eta) + T_1(h_z) + R_2(\alpha, h_v, h_\eta, h_z), \qquad \text{(A.1)} \end{aligned}$$

where $\hat{\delta}_{Yi} \equiv \hat{Y}_{2i} - \hat{Y}_i$; $\hat{\delta}_{Xi} \equiv \hat{X}_{2i} - \hat{X}_i$; $\hat{Y}_{2i} = \hat{m}_y(V_i, \hat{\eta}_i) + \hat{W}_{2i}$; $\hat{W}_{2i} = \frac{\sum_{j\neq i} W_j L_{2,ij}}{\sum_{j\neq i} L_{2,ij}}$; $\hat{X}_{2i} = \hat{m}_x(V_i, \hat{\eta}_i) + \hat{U}_{2i}$; $\hat{U}_{2i} = \frac{\sum_{j\neq i} U_j L_{2,ij}}{\sum_{j\neq i} L_{2,ij}}$; $\hat{Y}_i = \hat{m}_y(V_i, \eta_i) + \hat{W}_i$; $\hat{W}_i = \frac{\sum_{j\neq i} W_j L_{ij}}{\sum_{j\neq i} L_{ij}}$; $\hat{X}_i = \hat{m}_x(V_i, \eta_i) + \hat{U}_i$; $\hat{U}_i = \frac{\sum_{j\neq i} U_j L_{ij}}{\sum_{j\neq i} L_{ij}}$. To obtain the asymptotic equivalence between (2.16) and

(2.20), $\hat{J}^*(\alpha, h_v, h_\eta)$ in (A.1) is further expanded as follows

$$\begin{aligned} \hat{J}^*(\alpha, h_v, h_\eta) &\equiv \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i - \left\{X_i - \hat{X}_i\right\}'\beta\right)^2 \\ &= \tilde{J}(\alpha) + T_2(h_v, h_\eta) + R_1(\alpha, h_v, h_\eta). \end{aligned} \tag{A.2}$$

The two remainder terms, R_1 and R_2 in (A.2) and (A.1), are composed of a number of elementary terms. Firstly, R_1 is decomposed with the following terms: $B'S_{\tilde{m}_x-\hat{m}_x}B$, $B'S_{\hat{U}}B, B'S_{m_x-\tilde{m}_x,\tilde{m}_x-\hat{m}_x}B, B'S_{m_x-\tilde{m}_x,\hat{U}}B, B'S_{\tilde{m}_x-\hat{m}_x,U}B, B'S_{\tilde{m}_x-\hat{m}_x,\hat{U}}B, B'S_{U\hat{U}}B, S_{\tilde{m}-\hat{m}}$, $S_{m-\tilde{m},\tilde{m}-\hat{m}}, S_{e\hat{e}}, S_{\hat{e}}, B'S_{\tilde{m}_x-\hat{m}_x,m-\tilde{m}}, B'S_{\hat{U},m-\tilde{m}}, B'S_{m_x-\tilde{m}_x,\tilde{m}-\hat{m}}, BS_{\tilde{m}_x-\hat{m}_x,\tilde{m}-\hat{m}}, B'S_{U,\tilde{m}-\hat{m}}$, $B'S_{\hat{U},\tilde{m}-\hat{m}}, B'S_{\tilde{m}_x-\hat{m}_x,e}, B'S_{\hat{U}e}, B'S_{U\hat{e}}, B'S_{m_x-\tilde{m}_x,\hat{e}}, B'S_{\tilde{m}_x-\hat{m}_x,\hat{e}}, B'S_{U\hat{e}}, BS_{\hat{U}\hat{e}}, S_{\tilde{m}-\hat{m},e}$, $S_{m-\tilde{m},\hat{e}}, S_{\tilde{m}-\hat{m},\hat{e}}$ and $S_{m-\hat{m}_0}$, where $\hat{m}_0 = \frac{\sum_{j\neq i} m_j L_{0,ij}}{\sum_{j\neq i} L_{0,ij}}$. The uniform consistency of R_1, namely $\sup_{\alpha\in\mathcal{A}_n, h_v, h_\eta\in\mathcal{H}_n} |R_1(\alpha, h_v, h_\eta)| = o_p(n^{-1/2})$, is followed by Propositions A.1-A.3, A.6, A.7, A.9 and A.12-A.14, with $\beta = \beta_0 + O(n^{-1/2})$ as defined in (2.17) and $S_{m-\hat{m}_0} = O_p(n^{-1}h_v^{-1}h_\eta^{-q_2}) + O_p((h_v^2+h_\eta^2)^2)$ by nonparametric analysis. The second remainder term, R_2, is decomposed as follows: $B'S_{\hat{\delta}_X}B, S_{\hat{\delta}_m}, S_{\hat{\delta}_e}, B'S_{\hat{\delta}_X\hat{\delta}_m}, B'S_{\hat{\delta}_X\hat{\delta}_e}, S_{\hat{\delta}_m\hat{\delta}_e}, B'S_{\hat{\delta}_X,m_x-\tilde{m}_x}B$, $B'S_{\hat{\delta}_X,\tilde{m}_x-\hat{m}_x}B, B'S_{\hat{\delta}_X U}B, B'S_{\hat{\delta}_X\hat{U}}B, B'S_{\hat{\delta}_X,m-\tilde{m}}, B'S_{\hat{\delta}_X,\tilde{m}-\hat{m}}, B'S_{\hat{\delta}_X e}, B'S_{\hat{\delta}_X\hat{e}}, B'S_{\hat{\delta}_m,m_x-\tilde{m}_x}$, $BS_{\hat{\delta}_m,\tilde{m}_x-\hat{m}_x}, B'S_{\hat{\delta}_m U}, B'S_{\hat{\delta}_m\hat{U}}, S_{\hat{\delta}_m,m-\tilde{m}}, S_{\hat{\delta}_m,\tilde{m}-\hat{m}}, S_{\hat{\delta}_m e}, S_{\hat{\delta}_m\hat{e}}, BS_{\hat{\delta}_e,m_x-\tilde{m}_x}, B'S_{\hat{\delta}_e,\tilde{m}_x-\hat{m}_x}$, $BS_{\hat{\delta}_e U}, B'S_{\hat{\delta}_e\hat{U}}, S_{\hat{\delta}_e,m-\tilde{m}}, S_{\hat{\delta}_e,\tilde{m}-\hat{m}}, S_{\hat{\delta}_e e}$ and $S_{\hat{\delta}_e\hat{e}} - S_{g_x-\hat{g}_x}$. The uniform consistency of R_2, namely $\sup_{\alpha\in A_n, h_v, h_\eta, h_z\in\mathcal{H}_n} |R_2(\alpha, h_z, h_v, h_\eta)| = o_p(n^{-1/2})$, is followed by Propositions A.4, A.5, A.8, A.10, A.11 and A.15, with $\beta = \beta_0 + O(n^{-1/2})$ and $S_{g_x-\hat{g}_x} = O_p(n^{-1}h_z^{-q_z}) + O_p(h_z^4)$ by nonparametric analysis. Note that the stated orders of the remainder terms are made available by using Chebyshev inequality.

Step 2. Proofs of Corollary 2.1: The proofs of the asymptotic properties of $\hat{\alpha}$ and $\hat{\beta}$ are now ready to be discussed. Firstly, by using the condition in (2.17), particularly $A_n = \{\alpha : ||\alpha - \alpha_0|| \leq C_1 n^{-1/2}\}$, and given the bounded values of X, the conditional expectation relationships are written as follows

$$m(v_0, \eta) = m(v, \eta) - X'(\alpha - \alpha_0)m_0^{(1)} + O(n^{-1}), \tag{A.3}$$

$$m(v_0, \eta | v, \eta) = m(v, \eta) - \tilde{m}_x(x|V, \eta)'(\alpha - \alpha_0)m_0^{(1)} + O(n^{-1}). \tag{A.4}$$

The asymptotic properties of $\hat{\alpha}$ by obtained using (A.3) and (A.4) with the expansion of $\tilde{J}(\alpha)$ are then considered as follows

$$\begin{aligned} \tilde{J}(\alpha) = {} & (\alpha_0 - \alpha)'\left(\frac{1}{n}\sum_{i=1}^{n}\left\{m_0^{(1)}\right\}^2 U_{0i}U_{0i}'\right)(\alpha_0 - \alpha) + \frac{2}{n}\sum_{i=1}^{n} m_0^{(1)}U_{0i}'(\alpha_0 - \alpha)e_i \\ & +2(\beta_0 - \beta)'\left(\frac{1}{n}\sum_{i=1}^{n} m_0^{(1)}U_{0i}U_{0i}'\right)(\alpha_0 - \alpha) + o_p(1) + O_p(n^{-1/2}), \end{aligned} \tag{A.5}$$

where $U_{0i} \equiv \{X_i - E(X_i|V_{0i}, \eta_i)\}$. Given α_0 and η, $\beta_0 - \beta$ is $\left(\frac{1}{n}\sum_{i=1}^n U_{0i}U_{0i}'\right)^- \frac{1}{n}\sum_{i=1}^n U_{0i}e_i$ and hence

$$\begin{aligned}\tilde{J}(\alpha) &= (\alpha_0 - \alpha)' \left\{m_0^{(1)}\right\}^2 S_{U_0}(\alpha_0 - \alpha) + 2m_0^{(1)} S_{eU_0}(\alpha_0 - \alpha) \\ &\quad -2\left\{(S_{U_0})^- S_{eU_0}\right\}\left\{m_0^{(1)} S_{U_0}(\alpha_0 - \alpha)\right\} + o_p(1). \qquad \text{(A.6)}\end{aligned}$$

The asymptotic properties of $\hat{\beta}$ are obtained by considering the linear reduced form (see Robinson (1988) for details), given $\hat{\eta}$ and $\hat{\alpha}$, as follows

$$Y_i - \hat{Y}_{3i} = (X_i - \hat{X}_{3i})'\beta_0 + (m_i - \hat{m}_{3i}) + (e_i - \hat{e}_{3i}), \qquad \text{(A.7)}$$

where $\hat{Y}_{3i} = \hat{m}_y(\hat{V}_i, \hat{\eta}_i) + \hat{W}_{3i}$, $\hat{W}_{3i} = \frac{\sum_{j\neq i} W_j L_{3,ij}}{\sum_{i\neq j} L_{3,ij}}$, $\hat{X}_{3i} = \hat{m}_x(\hat{V}_i, \hat{\eta}_i) + \hat{U}_{3i}$, $\hat{U}_{3i} = \frac{\sum_{j\neq i} U_j L_{3,ij}}{\sum_{i\neq j} L_{3,ij}}$, $\hat{m}_{3i} = \frac{\sum_{j\neq i} \tilde{m}_j L_{3,ij}}{\sum_{i\neq j} L_{3,ij}}$, $\hat{e}_{3i} = \frac{\sum_{j\neq i} e_j L_{3,ij}}{\sum_{i\neq j} L_{3,ij}}$. By using (A.7), we obtain

$$\hat{\beta} - \beta_0 = S_{X-\hat{X}_3}^{-1}\left(S_{X-\hat{X}_3, m-\hat{m}_3} + S_{X-\hat{X}_3, e-\hat{e}_3}\right). \qquad \text{(A.8)}$$

Further decomposition of (A.7) is required as shown below

$$\begin{aligned}Y_i - \tilde{Y}_i + \tilde{Y}_i - \hat{Y}_{1i} - \breve{\delta}_{Yi} &= (X_i - \tilde{X}_i + \tilde{X}_i - \hat{X}_{1i} - \breve{\delta}_{Xi})'\beta_0 + (m_i - \tilde{m}_i + \tilde{m}_i - \hat{m}_{1i} - \breve{\delta}_{mi}) \\ &\quad + (e_i - \hat{e}_{1i} - \breve{\delta}_{ei}), \qquad \text{(A.9)}\end{aligned}$$

where $\breve{\delta}_{Yi} \equiv \hat{Y}_{3i} - \hat{Y}_{1i}$, $\breve{\delta}_{Xi} \equiv \hat{X}_{3i} - \hat{X}_{1i}$, $\breve{\delta}_{mi} \equiv \hat{m}_{3i} - \hat{m}_{1i}$, $\breve{\delta}_{ei} \equiv \hat{e}_{3i} - \hat{e}_{1i}$, $\hat{Y}_{1i} = \hat{m}_y(\hat{V}_i, \eta_i) + \hat{W}_{1i}$, $\hat{W}_{1i} = \frac{\sum_{j\neq i} W_j L_{1,ij}}{\sum_{j\neq i} L_{1,ij}}$, $\hat{X}_{1i} = \hat{m}_x(\hat{V}_i, \eta_i) + \hat{U}_{1i}$, $\hat{U}_{1i} = \frac{\sum_{j\neq i} U_j L_{1,ij}}{\sum_{j\neq i} L_{1,ij}}$, $\hat{m}_{1i} = \frac{\sum_{j\neq i} \tilde{m}_j L_{1,ij}}{\sum_{j\neq i} L_{1,ij}}$ and $\hat{e}_{1i} = \frac{\sum_{j\neq i} e_j L_{1,ij}}{\sum_{j\neq i} L_{ij}}$.

The terms in (A.8) is decomposed further by using (A.9) as follows: $S_{m_x-\tilde{m}_x}$, $S_{\tilde{m}_x-\hat{m}_{x_1}}$, S_U, $S_{\hat{U}_1}$, $S_{\breve{\delta}_X}$, $S_{m_x-\tilde{m}_x,\tilde{m}_x-\hat{m}_{x_1}}$, $S_{m_x-\tilde{m}_x,U}$, $S_{m_x-\tilde{m}_x,\hat{U}_1}$, $S_{m_x-\tilde{m}_x,\breve{\delta}_X}$, $S_{\tilde{m}_x-\hat{m}_{x_1},U}$, $S_{\tilde{m}_x-\hat{m}_{x_1},\hat{U}_1}$, $S_{m_x-\tilde{m}_x,\breve{\delta}_X}$, $S_{U\hat{U}_1}$, $S_{U\breve{\delta}_X}$, $S_{\hat{U}_1\breve{\delta}_X}$, $S_{m-\tilde{m},e}$, $S_{m-\tilde{m},\hat{e}_1}$, $S_{m-\tilde{m},\breve{\delta}e}$, $S_{\tilde{m}-\hat{m}_1,e}$, $S_{\tilde{m}-\hat{m}_1,\hat{e}_1}$, $S_{\tilde{m}-\hat{m}_1,\breve{\delta}_e}$, $S_{\breve{\delta}_m e}$, $S_{\breve{\delta}_m\hat{e}_1}$, $S_{m-\tilde{m}}$, $S_{\tilde{m}-\hat{m}_1}$, $S_{\breve{\delta}_m}$, $S_{m-\tilde{m},\tilde{m}-\hat{m}_1}$, $S_{m-\tilde{m},\breve{\delta}_m}$, $S_{\tilde{m}-\hat{m}_1,\breve{\delta}_m}$, $S_{\breve{\delta}_e\breve{\delta}_m}$, S_e, $S_{\hat{e}_1}$, $S_{\breve{\delta}_e}$, $S_{e\hat{e}_1}$, $S_{e\breve{\delta}_e}$, $S_{\hat{e}_1\breve{\delta}e}$, $S_{m_x-\tilde{m}_x,e}$, $S_{m_x-\tilde{m}_x,\hat{e}_1}$, $S_{m_x-\tilde{m}_x,\breve{\delta}e}$, $S_{\tilde{m}_x-\hat{m}_{x_1},e}$, $S_{\tilde{m}_x-\hat{m}_{x_1},\hat{e}_1}$, $S_{\tilde{m}_x-\hat{m}_{x_1}}$, S_{Ue}, $S_{U\hat{e}_1}$, $S_{U\breve{\delta}_e}$, $S_{\hat{U}_1 e}$, $S_{\hat{U}_1\hat{e}_1}$, $S_{\hat{U}_1\breve{\delta}_e}$, $S_{\breve{\delta}_X e}$, $S_{\breve{\delta}_X\hat{e}_1}$, $S_{\breve{\delta}X\breve{\delta}e}$, $S_{m_x-\tilde{m}_x,m-\tilde{m}}$, $S_{m_x-\tilde{m}_x,\tilde{m}-\hat{m}_1}$, $S_{m_x-\tilde{m}_x,\breve{\delta}_m}$, $S_{\tilde{m}_x-\hat{m}_{x_1},m-\tilde{m}}$, $S_{\tilde{m}_x-\hat{m}_{x_1},\tilde{m}-\hat{m}_1}$, $S_{\hat{U}_1\breve{\delta}_m}$, $S_{\tilde{m}_x-\hat{m}_{x_1},\breve{\delta}_m}$, $S_{m-\tilde{m},U}$, $S_{\tilde{m}-\hat{m}_1,U}$, $S_{U\breve{\delta}_m}$, $S_{m-\tilde{m},\hat{U}_1}$, $S_{\tilde{m}-\hat{m}_1,\hat{U}_1}$, $S_{m-\tilde{m},\breve{\delta}_X}$, $S_{\tilde{m}-\hat{m}_1,\breve{\delta}_X}$ and $S_{\breve{\delta}_X\breve{\delta}_m}$. Note that the two kernel functions are approximated such that $L_{3,ij} = L_{2,ij} + O_p(n^{-1/2}h_v^{-1})$ and $L_{1,ij} = L_{ij} + O_p(n^{-1/2}h_v^{-1})$ uniformly in i. Hence, $L_{2,ij}$ and L_{ij} are used instead of $L_{3,ij}$ and $L_{1,ij}$, respectively, for the case of $\hat{\beta}$ in Propositions A.1 to A.15. By Propositions A.1-A.15, and (A.3) and (A.4), (A.8) becomes

$$\begin{aligned}&\hat{\beta} - \beta_0 \\ &= \left(\frac{1}{n}\sum_{i=1}^n U_{0i}U_{0i}'\right)^- \left\{\frac{1}{n}\sum_{i=1}^n U_{0i}e_i - \frac{1}{n}\sum_{i=1}^n m_0^{(1)} U_{0i}U_{0i}'(\alpha_0 - \alpha)\right\} + O_p(n^{-1/2}) + o_p(1).\end{aligned}$$

Given β_0 and by using (A.5), (A.8) is further simplified as shown below

$$\hat{\beta}-\beta_0=(S_{U_0})^{-}\left\{S_{U_0e}-m_0^{(1)}S_{U_0}\left(\left\{m_0^{(1)}\right\}^2S_{U_0}\right)^{-}m_0^{(1)}S_{eU_0}\right\}+o_p(1). \qquad (A.10)$$

Given both $\hat{\beta}$ and $\hat{\alpha}$, the variance of e is

$$\begin{aligned}\hat{\sigma}^2 &= S_{e-\hat{e}_3}+S_{m-\hat{m}_3}+(\hat{\beta}-\beta_0)'S_{X-\hat{X}_3}(\hat{\beta}-\beta_0)-2(\hat{\beta}-\beta_0)'S_{X-\hat{X}_3,e-\hat{e}_3}\\ &\quad -2(\hat{\beta}-\beta_0)'S_{X-\hat{X}_3,m-\hat{m}_3}+2S_{m-\hat{m}_3,e-\hat{e}_3} \qquad (A.11)\\ &= S_e+o_p(1)\xrightarrow{p}\sigma_2^2\end{aligned}$$

by Propositions A.1-A.15 below, the law of large numbers and the *i.i.d.* assumption of e_i . The other nine terms, $(\hat{\beta}-\beta_0)'S_{m_x-\tilde{m}_x}(\hat{\beta}-\beta_0)$; $(\hat{\beta}-\beta_0)'S_{m_x-\tilde{m}_x,U}(\hat{\beta}-\beta_0)$; $(\hat{\beta}-\beta_0)'S_U(\hat{\beta}-\beta_0)$; $S_{m-\tilde{m}}$; $S_{m_x-\tilde{m}_x,m-\tilde{m}}$; $S_{m-\tilde{m},U}$; $S_{m_x-\tilde{m}_x,e}$; S_{Ue}; $S_{m-\tilde{m},e}$, are $o_p(n^{-1/2})$. Therefore, by using the central limit theorem and the law of large numbers, the asymptotic normalities of $\hat{\alpha}$ and $\hat{\beta}$ are as follows

$$\begin{aligned}\sqrt{n}(\hat{\beta}-\beta_0) &= \sqrt{n}\,(S_{U_0})^{-}\left\{S_{U_0e}-m_0^{(1)}S_{U_0}\left(\left\{m_0^{(1)}\right\}^2S_{U_0}\right)^{-}m_0^{(1)}S_{eU_0}\right\}+o_p(1)\\ &\to_D N\left(0,\sigma^2\left[\Phi_{U_0}^{-}-\left(m_0^{(1)}\Phi_{U_0}\right)^{-}\Phi_{U_0}\left\{m_0^{(1)}\right\}^2\left(m_0^{(1)}\Phi_{U_0}\right)^{-}\right]\right)\\ \sqrt{n}(\hat{\alpha}-\alpha_0) &= \sqrt{n}\left(\left\{m_0^{(1)}\right\}^2S_{U_0}\right)^{-}\left\{m_0^{(1)}S_{eU_0}-m_0^{(1)}S_{U_0}\,(S_{U_0})^{-}\,S_{eU_0}\right\}+o_p(1)\\ &\to_D N\left(0,\sigma^2\left[\left(\left\{m_0^{(1)}\right\}^2\Phi_{U_0}\right)^{-}-\left\{m_0^{(1)}\Phi_{U_0}\right\}^{-}\Phi_{U_0}\left\{m_0^{(1)}\Phi_{U_0}\right\}^{-}\right]\right).\end{aligned}$$

■

Next, the proofs of Propositions A.1 to A.15, are shown.

Proposition A.1. *$\sqrt{n}S_{\tilde{m}_x-\hat{m}_x}$ and $\sqrt{n}S_{\tilde{m}-\hat{m}}$ are $O_p(n^{-1/2}h_v^{-1}h_\eta^{-q_2})+O_p(n^{1/2}(h_v^2+h_\eta^2)^2)$.*

Proof: *Let $\varphi(\cdot)$ and $\tilde{\varphi}(\cdot)$ denote $m(\cdot)$ and $m_x(\cdot)$, and $\tilde{m}(\cdot)$ and $\tilde{m}_x(\cdot)$, respectively. Then, uniformly in i, (A.3) and (A.4) are used to deduce the following*

$$\tilde{\varphi}_i-\hat{\varphi}_i = \frac{(nh_vh_\eta^{q_2})^{-1}\sum_{j\neq i}\left\{\tilde{\varphi}_i-\tilde{\varphi}_j+U_j'(\alpha-\alpha_0)\varphi_0^{(1)}\right\}L_{ij}}{f(V,\eta)}\left(1-\frac{\hat{f}(V,\eta)-f(V,\eta)}{\hat{f}(V,\eta)}\right)+o(1),$$

where $\varphi_0^{(1)}=\partial\varphi(V_0,\eta)/\partial V_0$. Note that $\left(\hat{f}(V,\eta)-f(V,\eta)\right)=O_p(nh_vh_\eta^{q_2})^{-1/2}+O_p(h_v^2+h_\eta^2)$ so that $\left(1-\frac{\hat{f}(V,\eta)-f(V,\eta)}{\hat{f}(V,\eta)}\right)$ can be dropped and, hence, only the numerator term is *considered in the rest of this section. By identical distribution,* $E(S_{\tilde{\varphi}-\hat{\varphi}}) = E\left\{(\tilde{\varphi}_i - \hat{\varphi}_i)^2\right\}$, *where* $E(\tilde{\varphi}_i - \hat{\varphi}_i) = O(h_v^2 + h_\eta^2)$ *and* $\mathrm{Var}(\tilde{\varphi}_i - \hat{\varphi}_i) = O(nh_v h_\eta^{q_2})^{-1}$. *Because*

$$\mathrm{Var}(\tilde{\varphi}_i - \hat{\varphi}_i) = \mathrm{Var}\left(\frac{1}{nh_v h_\eta^{q_2}}\sum_{j\neq i}(\tilde{\varphi}_i - \tilde{\varphi}_j)L_{ij}\right) + \mathrm{Var}\left(\frac{1}{nh_v h_\eta^{q_2}}\sum_{j\neq i}U_j'(\alpha-\alpha_0)\varphi_0^{(1)}L_{ij}\right)$$
$$+\ 2\mathrm{Cov}\left(\frac{1}{nh_v h_\eta^{q_2}}\sum_{j\neq i}(\tilde{\varphi}_i - \tilde{\varphi}_j)L_{ij}, \frac{1}{nh_v h_\eta^{q_2}}\sum_{j\neq i}U_j'(\alpha-\alpha_0)\varphi_0^{(1)}L_{ij}\right),$$

where

$$\mathrm{Var}\left(\frac{1}{nh_v h_\eta^{q_2}}\sum_{j\neq i}U_j'(\alpha-\alpha_0)\varphi_0^{(1)}L_{ij}\right) = O(n^2 h_v h_\eta^{q_2})^{-1}$$
$$\mathrm{Cov}\left(\frac{1}{nh_v h_\eta^{q_2}}\sum_{j\neq i}(\tilde{\varphi}_i - \tilde{\varphi}_j)L_{ij}, \frac{1}{nh_v h_\eta^{q_2}}\sum_{j\neq i}U_j'(\alpha-\alpha_0)\varphi_0^{(1)}L_{ij}\right) = O(n^{3/2}h_v h_\eta^{q_2})^{-1}.$$

Hence $E(S_{\tilde{\varphi}-\hat{\varphi}}) = O(nh_v h_\eta^{q_2})^{-1} + O((h_v^2 + h_\eta^2)^2)$. ■

Proposition A.2. $\sqrt{n}S_{\tilde{m}_x - \hat{m}_x, \tilde{m}-\hat{m}}$ *is* $O_p(n^{-1/2}h_v^{-1}h_\eta^{-q_2}) + O_p(n^{1/2}(h_v^2 + h_\eta^2)^2)$.

Proof: *By using Proposition A.1 (i)-(ii), and the Cauchy inequality.* ■

Proposition A.3. $\sqrt{n}S_{\hat{U}}$ *and* $\sqrt{n}S_{\hat{e}}$ *are* $O_p(n^{-1/2}h_v^{-1}h_\eta^{-q_2})$.

Proof: *Let* ϱ *denote* U *and* e. *By using the assumptions of* $E(\varrho|\mathcal{L}) = 0$ *almost surely, where* $\mathcal{L} = (X, \eta)$ *and the i.i.d. property of* ϱ, *we obtain the results of*$E(S_{\hat{\varrho}}) = E(\hat{\varrho}_i^2)$ *and*

$$E(\hat{\varrho}_i^2) = \frac{1}{n^2 h_v^2 h_\eta^{2q_2}} E\left(\sum_{j\neq i}\varrho_j^2 L_{ij}^2\right) = O(nh_v h_\eta^{q_2})^{-1}.$$

■

Proposition A.4. $\sqrt{n}S_{\hat{\delta}_X}, \sqrt{n}S_{\hat{\delta}_m}$ *and* $\sqrt{n}S_{\hat{\delta}_e}$ *are* $O_p\left(n^{-3/2}h_z^{-q_z}h_v^{-1}h_\eta^{-(q_2+2)}\right) + O_p\left(n^{-1/2}h_z^4 h_v^{-1}h_\eta^{-(q_2+2)}\right) + O_p(n^{3/2}h_z^4(h_v^2 + h_\eta^2)^2)$.

Proof: *Let* δ *denote* δ_X, δ_m *and* δ_e, *and* $\hat{\delta}_i = \hat{\delta}_{2i} - \hat{\delta}_{1i} \equiv \frac{\sum_{j\neq i}\delta_j L_{2,ij}}{\sum_{j\neq i}L_{2,ij}} - \frac{\sum_{j\neq i}\delta_j L_{ij}}{\sum_{j\neq i}L_{ij}}$. *By using Taylor expansion,* $L_{2,ij} = L_{ij} + L_{ij}^{(1)}\left(\frac{\triangle_{ij}}{h_\eta}\right) + L_{ij}^{(2)}(\tau)\left(\frac{\triangle_{ij}}{h_\eta}\right)^2$, *where* $L_{ij}^{(r)}$ *is the* r^{th} *derivative of* L_{ij} *with respect to* η *with* $r = 1$ *or* 2, $\triangle_{ij} = \{\hat{g}_x(Z_j) - g_x(Z_j)\} - \{\hat{g}_x(Z_i) - g_x(Z_i)\}$ *and* τ *is between the segment line of* $\eta_j - \eta_i$ *and* $\hat{\eta}_j - \hat{\eta}_i$. *Hence, the denominator of* $\hat{\delta}_{2i}$ *is*

$$\frac{1}{nh_v h_\eta^{q_2}}\sum_{j\neq i}L_{2,ij} = \frac{1}{nh_v h_\eta^{q_2}}\sum_{j\neq i}L_{ij} + \frac{1}{nh_v h_\eta^{q_2+1}}\sum_{j\neq i}L_{ij}^{(1)}\triangle_{ij} + R_{ij},$$

where R_{ij} *is the remainder term and the second term on the right-hand side is* $o_p(n^{-1/2})$, *because*

$$
\begin{aligned}
& E\left(\frac{1}{nh_v h_\eta^{q_2+1}}\sum_{j\neq i} L_{ij}^{(1)}\triangle_{ij}\right)^2 \\
& = \frac{1}{n^4 h_z^{2q_z} h_v^2 h_\eta^{2(q_2+1)}} E\left(\sum_{j\neq i}\left(L_{ij}^{(1)}\right)^2\left\{\sum_{l\neq j} C_{(l,j;K)} - \sum_{l\neq i} C_{(l,i;K)}\right\}^2\right) \\
& + \frac{2}{n^4 h_z^{2q_z} h_v^2 h_\eta^{2(q_2+1)}} E\left(\sum_{j\neq i}\left(L_{ij}^{(1)}\right)^2\left\{\sum_{l\neq j} C_{(l,j;K)} - \sum_{l\neq i} C_{(l,i;K)}\right\}\right. \\
& \left.\times\left\{\sum_{m\neq j,l} C_{(m,j;K)} - \sum_{m\neq i,l} C_{(m,i;K)}\right\}\right) \\
& + \frac{2}{n^4 h_z^{2q_z} h_v^2 h_\eta^{2(q_2+1)}} E\left(\sum_{j\neq i}\sum_{k\neq i,j} L_{ij}^{(1)} L_{ik}^{(1)}\left\{\sum_{l\neq j} C_{(l,j;K)} - \sum_{l\neq i} C_{(l,i;K)}\right\}\right. \\
& \left.\times\left\{\sum_{m\neq k,l} C_{(m,k;K)} - \sum_{m\neq i,l} C_{(m,i;K)}\right\}\right) \\
& = O\left(n^{-2}h_z^{-q_z}h_v^{-1}h_\eta^{-(q_2+2)}\right) + O\left(n^{-1}h_z^4 h_v^{-1}h_\eta^{-(q_2+2)}\right) + O\left(h_z^4(h_v^2+h_\eta^2)^2\right),
\end{aligned}
$$

where $C_{(l,j;K)} \equiv \{g_x(Z_l) - g_x(Z_j)\} K_{jl}$. *Hence* $\hat{\delta}_i = \frac{h_\eta^{-1}\sum_{j\neq i}\delta_j L_{ij}^{(1)}\triangle_{ij}}{\sum_{j\neq i} L_{ij}}$. *Next, consider* $E(S_{\hat{\delta}}) = \frac{1}{n}\sum_{i=1}^n E(\hat{\delta}_i^2) + \frac{2}{n}\sum_{i=1}^n\sum_{j=1,\neq i}^n E(\hat{\delta}_i\hat{\delta}_j)$. *Using a similar argument to the above, the two terms on the right-hand side of* $E(S_{\hat{\delta}})$ *are*

$$
\begin{aligned}
E\left(\hat{\delta}_i^2\right) & = \frac{1}{n^4 h_z^{2q_z} h_v^2 h_\eta^{2(q_2+1)}} E\left(\sum_{j\neq i}\delta_j^2\left(L_{ij}^{(1)}\right)^2\left\{\sum_{l\neq j} C_{(l,j;K)} - \sum_{l\neq i} C_{(l,i;K)}\right\}^2\right) \\
& + \frac{2}{n^4 h_z^{2q_z} h_v^2 h_\eta^{2(q_2+1)}} E\left(\sum_{j\neq i}\delta_j^2\left(L_{ij}^{(1)}\right)^2\left\{\sum_{l\neq j} C_{(l,j;K)} - \sum_{l\neq i} C_{(l,i;K)}\right\}\right. \\
& \left.\times\left\{\sum_{m\neq j,l} C_{(m,j;K)} - \sum_{m\neq i,l} C_{(m,i;K)}\right\}\right) \\
& = O\left(n^{-2}h_z^{-q_z}h_v^{-1}h_\eta^{-(q_2+2)}\right) + O\left(n^{-1}h_z^4 h_v^{-1}h_\eta^{-(q_2+2)}\right)
\end{aligned}
$$

$$
\begin{aligned}
E\left(\hat{\delta}_i\hat{\delta}_j\right) & = \frac{1}{n^4 h_z^{2q_z} h_v^2 h_\eta^{2(q_2+1)}} \\
& \times\sum_{l\neq i}\sum_{l\neq j} E\left(\delta_l^2 L_{il}^{(1)} L_{jl}^{(1)}\left\{\sum_{k\neq l} C_{(k,l;K)} - \sum_{k\neq i} C_{(k,i;K)}\right\}\left\{\sum_{m\neq l} C_{(m,l;K)} - \sum_{m\neq j} C_{(m,j;K)}\right\}\right) \\
& = O\left(h_z^4(h_v^2+h_\eta^2)^2\right).
\end{aligned}
$$

■

Proposition A.5. $\sqrt{n}S_{\hat{\delta}_X\hat{\delta}_m}, \sqrt{n}S_{\hat{\delta}_X\hat{\delta}_e}$ *and* $\sqrt{n}S_{\hat{\delta}_m\hat{\delta}_e}$ *are* $O_p\left(n^{-3/2}h_z^{-q_z}h_v^{-1}h_\eta^{-(q_2+2)}\right) + O_p\left(n^{-1/2}h_z^4 h_v^{-1}h_\eta^{-(q_2+2)}\right) + O_p(n^{3/2}h_z^4(h_v^2+h_\eta^2)^2)$.

Proof: *By using Proposition A.4 (i)-(iii), and the Cauchy inequality.* ■

Proposition A.6. $\sqrt{n}S_{U\hat{U}}, \sqrt{n}S_{\hat{U}e}, \sqrt{n}S_{e\hat{e}}$ *and* $\sqrt{n}S_{U\hat{e}}$ *are* $O_p(n^{-1/2}h_v^{-1/2}h_\eta^{-q_2/2})$.

Proof: *Since* $E(\varrho|\mathcal{L})=0$ *almost surely and* ϱ *is i.i.d., we have*

$$E\left(\sqrt{n}S_{\varrho\hat{\varrho}}\right)^2 = \frac{1}{n}\sum_{i=1}^n E\left(\varrho_i^2\hat{\varrho}_i^2\right), \text{ where } E(\varrho_i^2\hat{\varrho}_i^2) = \frac{1}{n^2h_v^2h_\eta^{2q_2}}E\left(\varrho_i^2\sum_{j\neq i}\varrho_j^2L_{ij}^2\right) = O(nh_vh_\eta^{q_2})^{-1}.$$

■

Proposition A.7. $\sqrt{n}S_{\tilde{m}_x-\hat{m}_x,U}, \sqrt{n}S_{\tilde{m}-\hat{m},U}, \sqrt{n}S_{\tilde{m}_x-\hat{m}_x,e}$ *and* $\sqrt{n}S_{\tilde{m}-\hat{m},e}$ *are* $O_p(n^{-1/2}h_v^{-1/2}h_\eta^{-q_2/2}) + O_p(h_v^2+h_\eta^2)$.

Proof: *Because* $E(\varrho|\mathcal{L})=0$ *almost surely and* ϱ *is i.i.d., we have*

$$E(\sqrt{n}S_{\tilde{\varphi}-\hat{\varphi},\varrho})^2 = \frac{1}{n}\sum_{i=1}^n E\left\{(\tilde{\varphi}_i-\hat{\varphi}_i)^2\varrho_i^2\right\},$$

where

$$\begin{aligned}
E\left\{(\tilde{\varphi}_i-\hat{\varphi}_i)^2\varrho_i^2\right\} &= \frac{1}{n^2h_v^2h_\eta^{2q_2}}E\left(\sum_{j\neq i}\left(C^*_{(i,j;L)}\right)^2\varrho_i^2\right) + \frac{2}{n^2h_v^2h_\eta^{2q_2}}E\left(\sum_{l\neq i}\sum_{l\neq i,j}C^*_{(i,l;L)}C^*_{(j,l;L)}\varrho_i^2\right) \\
&= O(n^{-1}h_v^{-1}h_\eta^{-q_2}) + O((h_v^2+h_\eta^2)^2)
\end{aligned}$$

with $C^*_{(i,l;L)} = \left\{\tilde{\varphi}_i - \tilde{\varphi}_l + U_l'(\alpha-\alpha_0)\varphi_0^{(1)}\right\}L_{il}$. ■

Proposition A.8. $\sqrt{n}S_{U\hat{\delta}_X}, \sqrt{n}S_{U\hat{\delta}_m}, \sqrt{n}S_{U\hat{\delta}_e}, \sqrt{n}S_{e\hat{\delta}_X}, \sqrt{n}S_{e\hat{\delta}_m}$ *and* $\sqrt{n}S_{e\hat{\delta}_e}$ *are* $O_p\left(n^{-1}h_z^{-q_z/2}h_v^{-1/2}h_\eta^{-(q_2+2)/2}\right) + O_p\left(n^{-1/2}h_z^2h_v^{-1/2}h_\eta^{-(q_2+2)/2}\right)$.

Proof: *Because* $E(\varrho|\mathcal{L})=0$ *almost surely and* ϱ *is i.i.d., we have*

$$E(\sqrt{n}S_{\varrho\hat{\delta}})^2 = \frac{1}{n}\sum_{i=1}^n E\left(\varrho_i^2\hat{\delta}_i^2\right),$$

where

$$\begin{aligned}
&E(\varrho_i^2\hat{\delta}_i^2) \\
&= \frac{1}{n^4h_z^{2q_z}h_v^2h_\eta^{2(q_2+1)}}E\left(\varrho_i^2\sum_{j\neq i}\delta_j^2\left(L_{ij}^{(1)}\right)^2\left\{\sum_{l\neq j}C_{(l,j;K)} - \sum_{l\neq i}C_{(l,i;K)}\right\}^2\right) \\
&\quad + \Bigg\{\frac{2}{n^4h_z^{2q_z}h_v^2h_\eta^{2(q_2+1)}} \\
&\quad \times E\left(\varrho_i^2\sum_{j\neq i}\delta_j^2\left(L_{ij}^{(1)}\right)^2\left\{\sum_{l\neq j}C_{(l,j;K)} - \sum_{l\neq i}C_{(l,i;K)}\right\}\left\{\sum_{m\neq j,l}C_{(m,j;K)} - \sum_{m\neq i,l}C_{(m,i;K)}\right\}\right)\Bigg\}
\end{aligned}$$

and hence

$$E(\varrho_i^2\hat{\delta}_i^2) = \; O\left(n^{-2}h_z^{-q_z}h_v^{-1}h_\eta^{-(q_2+2)}\right) + O\left(n^{-1}h_z^4h_v^{-1}h_\eta^{-(q_2+2)}\right),$$

using similar arguments to those in Proposition A.4. ■

Proposition A.9. $\sqrt{n}S_{\tilde{m}_x-\hat{m}_x,\hat{U}}$, $\sqrt{n}S_{\tilde{m}-\hat{m},\hat{U}}$, $\sqrt{n}S_{\tilde{m}_x-\hat{m}_x,\hat{e}}$ *and* $\sqrt{n}S_{\tilde{m}-\hat{m},\hat{e}}$ *are* $O_p(nh_vh_\eta^{q_2})^{-1} + O_p(n^{1/2}(h_v^2+h_\eta^2)^2)$.

Proof:

$$E\left(\sqrt{n}S_{\tilde{\varphi}-\hat{\varphi},\hat{\varrho}}\right)^2 = \frac{1}{n}\sum_{i=1}^n E\left\{(\tilde{\varphi}_i-\hat{\varphi}_i)^2\hat{\varrho}_i^2\right\} + \frac{2}{n}\sum_{i=1}^n\sum_{j=1,\neq i}^n E\left\{(\tilde{\varphi}_i-\hat{\varphi}_i)(\tilde{\varphi}_j-\hat{\varphi}_j)\hat{\varrho}_i\hat{\varrho}_j\right\},$$

where

$$E\left\{(\tilde{\varphi}_i-\hat{\varphi}_i)^2\hat{\varrho}_i^2\right\} = \frac{1}{n^4h_v^4h_\eta^{4q_2}}E\left(\sum_{l\neq i}\left(C^*_{(i,l;L)}\right)^2\sum_{j\neq i}L_{ij}^2\varrho_j^2\right)$$
$$+\;\frac{2}{n^4h_v^4h_\eta^{4q_2}}E\left(\sum_{l\neq i}\sum_{k\neq i,l}C^*_{(i,l;L)}C^*_{(i,k;L)}\sum_{j\neq i}L_{ij}^2\varrho_j^2\right)$$
$$= O(n^{-2}h_v^{-2}h_\eta^{-2q_2}) + O(n^{-1}h_v^{-1}h_\eta^{-q_2}(h_v^2+h_\eta^2)^2)$$
$$E\left\{(\tilde{\varphi}_i-\hat{\varphi}_i)(\tilde{\varphi}_j-\hat{\varphi}_j)\hat{\varrho}_i\hat{\varrho}_j\right\} = \frac{1}{n^4h_v^4h_\eta^{4q_2}}E\left(\sum_{k\neq i}\sum_{m\neq j}C^*_{(i,k;L)}C^*_{(j,m;L)}\sum_{l\neq i}\sum_{l\neq j}L_{il}L_{jl}\varrho_l^2\right)$$
$$= O((h_v^2+h_\eta^2)^4).$$

■

Proposition A.10. $\sqrt{n}S_{\hat{U}\hat{\delta}_X}$, $\sqrt{n}S_{\hat{U}\hat{\delta}_m}$, $\sqrt{n}S_{\hat{U}\hat{\delta}_e}$, $\sqrt{n}S_{\hat{e}\hat{\delta}_X}$, $\sqrt{n}S_{\hat{e}\hat{\delta}_m}$ *and* $\sqrt{n}S_{\hat{e}\hat{\delta}_e}$ *are* $O_p\left(n^{-3/2}h_z^{-q_z/2}h_v^{-1}h_\eta^{-(q_2+1)}\right) + O_p\left(n^{-1}h_z^2h_v^{-1}h_\eta^{-(q_2+1)}\right) + O_p\left(n^{1/2}h_z^2(h_v^2+h_\eta^2)\right)$.

Proof:

$$E\left(\sqrt{n}S_{\hat{\varrho}\hat{\delta}}\right)^2 = \frac{1}{n}\sum_{i=1}^n E\left(\hat{\varrho}_i^2\hat{\delta}_i^2\right) + \frac{2}{n}\sum_{i=1}^n\sum_{j=1,\neq i}^n E\left(\hat{\varrho}_i\hat{\varrho}_j\hat{\delta}_i\hat{\delta}_j\right),$$

where

$$E\left(\hat{\varrho}_i^2\hat{\delta}_i^2\right)$$
$$= \frac{1}{n^6h_z^{2q_z}h_v^4h_\eta^{2(2q_2+1)}}E\left(\sum_{j\neq i}\varrho_j^2L_{ij}^2\sum_{l\neq i}\delta_l^2\left\{L_{il}^{(1)}\right\}^2\left\{\sum_{k\neq l}C_{(k,l;K)}-\sum_{k\neq i}C_{(k,i;K)}\right\}^2\right)$$
$$+\left\{\frac{2}{n^6h_z^{2q_z}h_v^4h_\eta^{2(2q_2+1)}}E\left(\sum_{j\neq i}\varrho_j^2L_{ij}^2\sum_{l\neq i}\delta_l^2\left(L_{il}^{(1)}\right)^2\left\{\sum_{k\neq l}C_{(k,l;K)}-\sum_{k\neq i}C_{(k,i,;K)}\right\}\right.\right.$$
$$\left.\left.\times\;\left\{\sum_{m\neq l,k}C_{(m,l;K)}-\sum_{m\neq i,k}C_{(m,i;K)}\right\}\right)\right\}$$

and hence

$$E\left(\hat{\varrho}_i^2\hat{\delta}_i^2\right) = O\left(n^{-3}h_z^{-q_z}h_v^{-2}h_\eta^{-(2q_2+2)}\right) + O\left(n^{-2}h_z^4 h_v^{-2}h_\eta^{-(2q_2+2)}\right),$$

and the cross-product term, $E\left(\hat{\varrho}_i\hat{\varrho}_j\hat{\delta}_i\hat{\delta}_j\right)$, *is* $\left(n^6 h_z^{2q_z}h_v^4 h_\eta^{2(2q_2+1)}\right)^{-1}$ *times*

$$E\left(\sum_{s\neq i}\sum_{s\neq i,j}\varrho_s^2 L_{is}L_{js}\sum_{l\neq i}\sum_{l\neq i,j}\delta_l^2 L_{il}^{(1)}L_{jl}^{(1)}\left\{\sum_{k\neq l}C_{(k,l;K)} - \sum_{k\neq j}C_{(k,j;K)}\right\}\right.$$
$$\left.\times\left\{\sum_{m\neq l,k}C_{(m,l;K)} - \sum_{m\neq i,k}C_{(m,i;K)}\right\}\right).$$

Hence the cross-product term is $O\left(h_z^4(h_v^2+h_\eta^2)^2\right)$. ■

Proposition A.11. $\sqrt{n}S_{\tilde{m}_x-\hat{m}_x,\hat{\delta}_X}$, $\sqrt{n}S_{\tilde{m}_x-\hat{m}_x,\hat{\delta}_m}$, $\sqrt{n}S_{\tilde{m}_x-\hat{m}_x,\hat{\delta}_e}$, $\sqrt{n}S_{\tilde{m}-\hat{m},\hat{\delta}_X}$, $\sqrt{n}S_{\tilde{m}-\hat{m},\hat{\delta}_m}$, *and* $\sqrt{n}S_{\tilde{m}-\hat{m},\hat{\delta}_e}$ *are* $O_p\left(n^{-3/2}h_z^{-q_z/2}h_v^{-1}h_\eta^{-(q_2+1)}\right)+O_p\left(n^{-1}h_z^2 h_v^{-1}h_\eta^{-(q_2+1)}\right)+O_p\left(n^{1/2}h_z^2(h_v^2+h_\eta^2)\right)$.

Proof:

$$E\left(\sqrt{n}S_{\tilde{\varphi}-\hat{\varphi},\hat{\delta}}\right)^2 = \frac{1}{n}\sum_{i=1}^n E\left((\tilde{\varphi}_i-\hat{\varphi}_i)^2\hat{\delta}_i^2\right) + \frac{2}{n}\sum_{i=1}^n\sum_{j=1,\neq i}^n E\left((\tilde{\varphi}_i-\hat{\varphi}_i)(\tilde{\varphi}_j-\hat{\varphi}_j)\hat{\delta}_i\hat{\delta}_j\right),$$

where

$$\begin{aligned}
&E\left((\tilde{\varphi}_i-\hat{\varphi}_i)^2\hat{\delta}_i^2\right)\\
&= \frac{1}{n^6 h_z^{2q_z}h_v^4 h_\eta^{2(2q_2+1)}}E\left(\sum_{j\neq i}\left(C^*_{(i,j;L)}\right)^2\sum_{l\neq j}\delta_l^2\left(L_{il}^{(1)}\right)^2\left\{\sum_{k\neq l}C_{(k,l:K)} - \sum_{k\neq i}C_{(k,i;K)}\right\}^2\right)\\
&\quad+\frac{2}{n^6 h_z^{2q_z}h_v^4 h_\eta^{2(2q_2+1)}}E\left(\sum_{j\neq i}\left(C^*_{(i,j;L)}\right)^2\sum_{l\neq j}\delta_l^2\left(L_{il}^{(1)}\right)^2\left\{\sum_{k\neq l}C_{(k,l:K)} - \sum_{k\neq i}C_{(k,i;K)}\right\}\right.\\
&\quad\left.\times\left\{\sum_{m\neq l,k}C_{(m,l:K)} - \sum_{m\neq i,k}C_{(m,i;K)}\right\}\right)\\
&= O\left(n^{-3}h_z^{-q_z}h_v^{-2}h_\eta^{-(2q_2+2)}\right) + O\left(n^{-2}h_z^4 h_v^{-2}h_\eta^{-(2q_2+2)}\right),
\end{aligned}$$

and the cross-product term, $E\left((\tilde{\varphi}_i-\hat{\varphi}_i)(\tilde{\varphi}_j-\hat{\varphi}_j)\hat{\delta}_i\hat{\delta}_j\right)$, *is* $\left(n^6 h_z^{2q_z}h_v^4 h_\eta^{2(2q_2+1)}\right)^{-1}$ *times*

$$E\left(\sum_{s\neq i}\sum_{s\neq j,i}C^*_{(i,s;L)}C^*_{(j,s;L)}\sum_{l\neq i}\sum_{l\neq j,i}\delta_l^2 L_{il}^{(1)}L_{jl}^{(1)}\left\{\sum_{k\neq l}C_{(l,k:K)} - \sum_{k\neq j}C_{(k,j;K)}\right\}\right.$$
$$\left.\times\left\{\sum_{m\neq l,k}C_{(m,l:K)} - \sum_{m\neq i,k}C_{(m,i;K)}\right\}\right).$$

Hence the cross-product term is $O\left(h_z^4(h_v^2+h_\eta^2)^4\right)$. ■

Proposition A.12. *$\sqrt{n}S_{\hat{U}\hat{e}}$ is $O_p(nh_vh_\eta^{q_2})^{-1}+O_p(n^{1/2}(h_v^2+h_\eta^2)^2)$.*

Proof:

$$E\left(\sqrt{n}S_{\hat{U}\hat{e}}\right)^2=\frac{1}{n}\sum_{i=1}^{n}E\left\{\hat{U}_i^2\hat{e}_i^2\right\}+\frac{2}{n}\sum_{i=1}^{n}\sum_{j=1,\neq i}^{n}E\left\{\hat{U}_i\hat{U}_j'\hat{e}_i\hat{e}_j\right\},$$

where

$$E\left\{\hat{U}_i^2\hat{e}_i^2\right\}=\frac{1}{n^4h_v^4h_\eta^{4q_2}}E\left\{\sum_{j\neq i}U_jU_j'L_{ij}^2\sum_{l\neq i}e_l^2L_{il}^2\right\}=O(n^{-2}h_v^{-2}h_\eta^{-2q_2})$$

$$E\left\{\hat{U}_i\hat{U}_j'\hat{e}_i\hat{e}_j\right\}=\frac{1}{n^4h^4h_\eta^{4q_2}}E\left\{\sum_{l\neq i}\sum_{l\neq j}U_lU_l'L_{il}L_{jl}\sum_{k\neq i}\sum_{k\neq j}e_k^2L_{ik}L_{jk}\right\}=O((h_v^2+h_\eta^2)^4).$$

■

Proposition A.13. *$\sqrt{n}S_{m_x-\tilde{m}_x,\tilde{m}_x-\hat{m}_x}$, $\sqrt{n}S_{m-\tilde{m},\tilde{m}-\hat{m}}$, $\sqrt{n}S_{m_x-\tilde{m}_x,\tilde{m}-\hat{m}}$ and $\sqrt{n}S_{m-\tilde{m},\tilde{m}_x-\hat{m}_x}$ are $O_p\left(n^{-1}h_v^{-1/2}h_\eta^{-q_2/2}\right)$.*

Proof: *By (A.3) and (A.4) we deduce that, uniformly in i, we have*

$$\varphi_i-\tilde{\varphi}_i=U_i'(\alpha_0-\alpha)\varphi_0^{(1)}(V_{0i},\eta_i)+O(n^{-1}). \tag{A.12}$$

By using (A.12), we have $(\varphi_i-\tilde{\varphi}_i)(\tilde{\varphi}_i-\hat{\varphi}_i)=\frac{1}{nh_vh_\eta^{q_2}}\sum_{j\neq i}t_i\left\{\tilde{\varphi}_i-\tilde{\varphi}_j+U_j'(\alpha-\alpha_0)\varphi_0^{(1)}\right\}L_{ij}$, where $t_i=U_i'(\alpha_0-\alpha)\varphi_0^{(1)}$. For the rest of proofs, we use similar arguments to those in Proposition A.7 because $E(U|\mathcal{L})=0$ almost surely and U is i.i.d.. Hence we have

$$E\left(\sqrt{n}S_{\varphi-\tilde{\varphi},\tilde{\varphi}-\hat{\varphi}}\right)^2 \;=\; \frac{1}{n}\sum_{i=1}^{n}E\left(t_i^2(\tilde{\varphi}_i-\hat{\varphi}_i)^2\right),$$

where

$$\begin{aligned}E\left(t_i^2(\tilde{\varphi}_i-\hat{\varphi}_i)^2\right) &= \frac{1}{n^2h_v^2h_\eta^{2q_2}}E\left\{t_i^2\sum_{j\neq i}\left(C^*_{(i,j;L)}\right)^2\right\}+\frac{2}{n^2h_v^2h_\eta^{2q_2}}E\left\{t_i^2\sum_{j\neq i}\sum_{l\neq i,j}C^*_{(i,j;L)}C^*_{(i,l;L)}\right\}\\ &= O\left(n^{-2}h_v^{-1}h_\eta^{-q_2}\right)+O\left(n^{-1}(h_v^2+h_\eta^2)^2\right).\end{aligned}$$

■

Proposition A.14. *$\sqrt{n}S_{m_x-\tilde{m}_x,\hat{U}}$, $\sqrt{n}S_{m_x-\tilde{m}_x,\hat{e}}$, $\sqrt{n}S_{m-\tilde{m},\hat{U}}$ and $\sqrt{n}S_{m-\tilde{m},\hat{e}}$ are $O_p\left(n^{-1}h_v^{-1/2}h_\eta^{-q_2/2}\right)$.*

Proof: *By (A.12), $E(U|\mathcal{L})=0$ almost surely and because of the i.i.d. assumption of U, similar arguments to those in Proposition A.6 are used for the rest of the proof.*

$$E(\sqrt{n}S_{\varphi-\tilde{\varphi},\hat{\varrho}})^2=\frac{1}{n}\sum_{i=1}^{n}E\left(t_i^2\hat{\varrho}_i^2\right),$$

where

$$E\left(t_i^2\hat{\varrho}_i^2\right)=\frac{1}{n^2h_v^2h_\eta^{2q_2}}E\left(t_i^2\sum_{j\neq i}\varrho_j^2L_{ij}^2\right)=O\left(n^{-2}h_v^{-1}h_\eta^{-q_2}\right).$$

■

Proposition A.15. $\sqrt{n}S_{m_x-\tilde{m}_x,\hat{\delta}_X}$, $\sqrt{n}S_{m_x-\tilde{m}_x,\hat{\delta}_m}$, $\sqrt{n}S_{m_x-\tilde{m}_x,\hat{\delta}_e}$, $\sqrt{n}S_{m-\tilde{m},\hat{\delta}_X}$, $\sqrt{n}S_{m-\tilde{m},\hat{\delta}_m}$ *and* $\sqrt{n}S_{m-\tilde{m},\hat{\delta}_e}$ *are* $O_p\left(n^{-3/2}h_z^{-q_z/2}h_v^{-1/2}h_\eta^{-(q_2+2)/2}\right)$.

Proof: *By (A.12),* $E(U|\mathcal{L})=0$ *almost surely and because of the i.i.d. assumptions, the rest of the proofs is similar to that of Proposition A.8.*

$$E(\sqrt{n}S_{\varphi-\tilde{\varphi},\hat{\delta}})^2=\frac{1}{n}\sum_{i=1}^{n}E\left(t_i^2\hat{\delta}_i^2\right),$$

where

$$\begin{aligned}E\left(t_i^2\hat{\delta}_i^2\right)&=\frac{1}{n^4h_z^{2q_z}h_v^2h_\eta^{2(q_2+1)}}E\left(t_i^2\sum_{j\neq i}\delta_j^2\left\{L_{ij}^{(1)}\right\}^2\left\{\sum_{l\neq j}C_{(l,j;K)}-\sum_{l\neq i}C_{(l,i;K)}\right\}^2\right)\\&+\frac{2}{n^4h_z^{2q_z}h_v^2h_\eta^{2(q_2+1)}}E\left(t_i^2\sum_{j\neq i}\delta_j^2\left(L_{ij}^{(1)}\right)^2\left\{\sum_{l\neq j}C_{(l,j;K)}-\sum_{l\neq i}C_{(l,i;K)}\right\}\right.\\&\times\left.\left\{\sum_{k\neq j,l}C_{(k,j;K)}-\sum_{k\neq i,l}C_{(k,i;K)}\right\}\right)\\&=O(n^{-3}h_z^{-q_z}h_v^{-1}h_\eta^{-(q_2+2)})+O(n^{-2}h_z^4h_v^{-1}h_\eta^{-(q_2+2)}).\end{aligned}$$

■

Proof of Theorem 2.2

Given $\hat{\beta}$ and $\hat{\alpha}$, fix the observation V_{0i} at v_0, then at the observation v_0 and η_i

$$\begin{aligned}\hat{m}(\hat{v},\hat{\eta}_i)-m(v_0,\eta_i)&=\{\hat{m}_{y**}(\hat{v},\eta_i)-\widetilde{m}_{y**}(\hat{v},\eta_i)\}\\&+\left\{Q_{1i}+Q_{2i}-(Q_{3i}+Q_{4i}+Q_{5i})'(\hat{\beta}-\beta_0)\right\},\end{aligned}\tag{B.1}$$

where $Y_i^{**}\equiv Y_i-X_i'\beta_0$; $\widetilde{m}_{y^{**}}(\hat{v},\eta_i)=E(m_{y^{**}}|\hat{v},\eta_i)$; $\widetilde{m}_x(\hat{v},\eta_i)=E(m_x|\hat{v},\eta_i)$; $\breve{\delta}_{m_y^{**},i}\equiv\hat{m}_{y^{**}}(\hat{v},\hat{\eta}_i)-\hat{m}_{y^{**}}(\hat{v},\eta_i)$; $\breve{\delta}_{m_X,i}\equiv\hat{m}_x(\hat{v},\hat{\eta}_i)-\hat{m}_x(\hat{v},\eta_i)$; $Q_{1i}=\tilde{m}_{y^{**}}(\hat{v},\eta_i)-m_{y^{**}}(v_0,\eta_i)$; $Q_{2i}=\breve{\delta}_{m_y^{**},i}$; $Q_{3i}=\hat{m}_x(\hat{v},\eta_i)-\tilde{m}_x(\hat{v},\eta_i)$; $Q_{4i}=\tilde{m}_x(\hat{v},\eta_i)-m_x(v_0,\eta_i)$; $Q_{5i}=\breve{\delta}_{m_x,i}$. As the results of a standard nonparametric analysis (see Hansen (2008) for example), the last five terms on the right-hand-side of (B.1) are $o_p(1)$ uniformly in i. In particular, $\sup_{z\in A_z}|Q_{2i}|=\sup_{z\in A_z}|Q_{5i}|=O_p\left\{\left(\frac{(\ln n)^2}{n^2h_z^{q_z}h_vh_\eta^{q2+2}}\right)^{1/2}+h_z^2(h_v^2+h_\eta^2)\right\}$. Hence (B.1) is

$$\hat{m}(\hat{v},\hat{\eta}_i)-m(v_0,\eta_i)=\hat{m}_{y^{**}}(\hat{v},\eta_i)-\widetilde{m}_{y^{**}}(\hat{v},\eta_i)+o_p(1)\equiv\hat{m}(\hat{v},\eta_i)-\widetilde{m}(\hat{v},\eta_i)+o_p(1),\tag{B.2}$$

where

$$\hat{m}(\hat{v},\eta_i) - \widetilde{m}(\hat{v},\eta_i) \quad = \quad \frac{\sum_{j\neq i}\{m(v_0,\eta_j) - m(v_0,\eta_i)\}\left\{L_{0,ij} + O_p(n^{-1/2}h_v^{-1})\right\}}{\sum_{j\neq i} L_{0,ij} + o_p(1)} + O_p(n^{-1/2}).$$

Hence (B.2) is

$$\hat{m}(\hat{v},\hat{\eta}_i) - m(v_0,\eta_i) = \hat{m}(v_0,\eta_i) - m(v_0,\eta_i) + o_p(1). \tag{B.3}$$

Let us define $\check{m}(v_0,\eta_i) = \hat{m}(v_0,\eta_i)\hat{f}(v_0,\eta_i)$. We can then rewrite the first term on the right-hand side of (B.3) as follows

$$\hat{m}(v_0,\eta_i) - m(v_0,\eta_i) \quad = \quad \frac{\check{m}(v_0,\eta_i) - m(v_0,\eta_i)\hat{f}(v_0,\eta_i)}{f(v_0,\eta_i)}\left(1 - \frac{\hat{f}(v_0,\eta_i) - f(v_0,\eta_i)}{\hat{f}(v_0,\eta_i)}\right),$$

where $\sup_{x,\eta\in\mathcal{U}}\left|\hat{f}(v_0,\eta_i) - f(v_0,\eta_i)\right| = O_p\left\{\left(\frac{\ln n}{nh_v h_\eta^{q_2}}\right)^{1/2} + (h_v^2 + h_\eta^2)\right\}$. First, we consider the bias term as follows

$$E(\hat{m}(v_0,\eta_i) - m(v_0,\eta_i)) = \frac{E\check{m}(v_0,\eta_i) - m(v_0,\eta_i)E(\hat{f}(v_0,\eta_i))}{f(v_0,\eta_i)},$$

where

$$\begin{aligned}
E\check{m}(v_0,\eta_i) \quad &= \quad E\left(E_{v_0,\eta_i}\left\{\frac{1}{nh_v h_\eta^{q_2}}\sum_{j=1,\neq i}^{n} k_v\left(\frac{V_{0j}-v_0}{h_v}\right)K_\eta\left(\frac{\eta_j-\eta_i}{h_\eta}\right)Y_j^{**}\right\}\right) \\
&= \quad E\left(\frac{1}{nh_v h_\eta^{q_2}}\sum_{j=1,\neq i}^{n} k_v\left(\frac{V_{0j}-v_0}{h_v}\right)K_\eta\left(\frac{\eta_j-\eta_i}{h_\eta}\right)m(v_0,\eta_i)\right) \\
&= \quad f(v_0,\eta_i)m(v_0,\eta_i) + h_v^2 B_v(v_0,\eta_i) + \sum_{l=1}^{q_2} h_{\eta,l}^2 B_{\eta,l}(v_0,\eta_i) + o(1).
\end{aligned}$$

In the expression above, E_{v_0,η_i} denotes the conditional expectation at the observation v_0 and η_i. Hence it is as follows

$$E(\hat{m}(v_0,\eta_i) - m(v_0,\eta_i)) = \left\{h_v^2 B_v(v_0,\eta_i) + \sum_{l=1}^{q_2} h_{\eta,l}^2 B_{\eta,l}(v_0,\eta_i)\right\} + o(1). \tag{B.4}$$

The single sum of (B.4) converges to its population mean by Chebyshev's law of large numbers (see Linton and Härdle (1996) for details).

Now let us consider the variance term. Note that $f(v_0,\eta_i) = f(v_0,\eta) + O_p(n^{-1/2})$ and $m(v_0,\eta_i) = m(v_0,\eta) + O_p(n^{-1/2})$ by the law of large numbers, since both functions satisfy the bounded moment conditions. We then have

$$\begin{aligned}
&\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{m}(v_0,\eta_i)\right) = \frac{1}{f(v_0,\eta)^2}\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n}\left\{\check{m}(v_0,\eta_i) - m(v_0,\eta_i)\hat{f}(v_0,\eta_i)\right\}\right) \\
&= \frac{1}{f(v_0,\eta)^2}\left\{\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n}\check{m}(v_0,\eta_i)\right) + m(v_0,\eta)^2\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{f}(v_0,\eta_i)\right)\right. \\
&\left. -2m(v_0,\eta)\mathrm{Cov}\left(\frac{1}{n}\sum_{i=1}^{n}\check{m}(v_0,\eta_i), \frac{1}{n}\sum_{i=1}^{n}\hat{f}(v_0,\eta_i)\right)\right\},
\end{aligned}$$

where

$$\begin{aligned}
&\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n}\check{m}(v_0,\eta_i)\right) = E\left(\mathrm{Var}_{v_0,\eta_i}\left\{\frac{1}{n}\sum_{i=1}^{n}\check{m}(v_0,\eta_i)\right\}\right) + \mathrm{Var}\left(E_{v_0,\eta_i}\left\{\frac{1}{n}\sum_{i=1}^{n}\check{m}(v_0,\eta_i)\right\}\right) \\
&= \sigma^2 f_\eta(\eta)^2 E\left(\frac{1}{nh_v}\sum_{j=1}^{n}k_v\left(\frac{V_{0j}-v_0}{h_v}\right)\right)^2 + f_\eta(\eta)^2\mathrm{Var}\left(\frac{1}{nh_v}\sum_{j=1}^{n}k_v\left(\frac{V_{0j}-v_0}{h_v}\right)m(V_{0j},\eta_j)\right) \\
&= \frac{\sigma^2 f_\alpha(v_0)f_\eta(\eta)^2}{nh_v}\mathcal{K}_v + \frac{m(v_0,\eta)^2 f_\alpha(v_0)f_\eta(\eta)^2}{nh_v}\mathcal{K}_v + O(n^{-1}), \\
&\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{f}(v_0,\eta_i)\right) = \frac{f_\alpha(v_0)f_\eta(\eta)^2\mathcal{K}_v}{nh_v} + O(n^{-1})
\end{aligned}$$

and

$$\begin{aligned}
&\mathrm{Cov}\left(\frac{1}{n}\sum_{i=1}^{n}\check{m}(v_0,\eta_i), \frac{1}{n}\sum_{i=1}^{n}\hat{f}(v_0,\eta_i)\right) \\
&= E\left\{\frac{1}{n^2}\sum_{i=1}^{n}\check{m}(v_0,\eta_i)\sum_{i=1}^{n}\hat{f}(v_0,\eta_i)\right\} - E\left\{\frac{1}{n}\sum_{i=1}^{n}\check{m}(v_0,\eta_i)\right\}E\left\{\frac{1}{n}\sum_{i=1}^{n}\hat{f}(v_0,\eta_i)\right\} \\
&= \frac{m(v_0,\eta)f_\alpha(v_0)f_\eta(\eta)^2\mathcal{K}_v}{nh_v} + O(n^{-1})
\end{aligned}$$

with $\mathrm{Var}_{v_0,\eta_i}$ denoting the conditional variance at v_0 and η_i. Hence we have

$$\sqrt{nh_v}(\hat{m}(\hat{v}) - m(v_0) - Bias) \to_D N(0, Var).$$

The consistency of $\hat{g}(\hat{v})$ and its asymptotic normality are argued in a similar way to the above because $m(v_0) = g(v_0) + C$. ■