Estimation in Partially Linear Semiparametric Models with Parametric and/or Nonparametric Endogeneity

Namhyun Kim†[1] and Patrick W. Saart‡[2]
University of Exeter† and *Cardiff University*‡

Abstract

Partially linear semiparametric models are advantageous to use in empirical studies of various economic problems due to a special feature that allows the parametric and nonparametric components to exist simultaneously in the model. However, systematic estimation procedures and methods have not yet been satisfactorily developed to deal effectively with a well-known endogeneity problem that may be present in some empirical applications. In the current paper, we aim to address endogeneity comprehensively, which may take place in either a parametric or a nonparametric component or both, and to provide guidance to an appropriate estimation procedure and method in the presence of such a problem. A significant difficulty we must overcome before such goals can be achieved is a generated regressor problem which arises because a critical part, known in the literature as the "control variables", is not observable in practice and hence must be estimated. We show theoretically (i.e. through the derivation of a set of important asymptotic properties) and experimentally (i.e. through the use of simulation exercises) that our newly introduced method can help in overcoming the above-mentioned endogeneity problem. For the sake of completeness, we also discuss an adaptive data-driven method of bandwidth selection and show its asymptotic optimality.

JEL Classification: C12, C14, C22
Abbreviated Title: Semiparametric Models with Endogeneity

1. Introduction

As is known in the literature, a partially linear semiparametric (PL) model is:

$$Y_i = X_i'\beta + g(V_i) + \epsilon_i \quad \text{for} \quad i = 1, \ldots, n$$
$$E(\epsilon|x, v) = 0, \tag{1.1}$$

where (V, X, Y) is a $\mathbb{R}^q \times \mathbb{R}^p \times \mathbb{R}$-valued observable random vector, β is a $\mathbb{R}^p$-valued unknown parameter vector and $g(\cdot)$ is an unknown real function such that $g : \mathbb{R}^q \to \mathbb{R}$. To estimate the model in (1.1), Robinson (1988) proposed a two-step estimation procedure, which is to first obtain consistent estimators of the unknown parameters and then use them in order to identify an unknown structural function; see also a discussion in Speckman (1988). Based on

[1]We thank Jiti Gao for his invaluable suggestions and comments.
[2]Patrick Saart, Cardiff University, United Kingdom: https://sites.google.com/site/patrickwsaart/home

an independently and identically distributed (i.i.d.) random sample (Y_i, X_i', V_i'), it has been shown that the parameter vector β in various versions of (1.1) can be consistently estimated at a rate of $\sqrt{n}$ (see Robinson (1988), Fan et al. (1995), Härdle et al. (2000), Gao (2007), and Li and Racine (2007), for examples).

It should be noted that the exogeneity condition of the regressors in the model (as stated mathematically in the second line in (1.1)) is crucial for obtaining consistent estimators and identifying an unknown structural function. However, it is not difficult to find circumstances by which such a fragile condition may breakdown in practice.

For instance, consider cases of modeling mis-specifications, e.g. omitted variables or/and interaction terms. For the sake of illustration, let us assume that $V = (V_1, V_2)$, where (V_1, V_2) is a $\mathbb{R}^{q^*} \times \mathbb{R}^{q-q^*}$-valued vector with $1 \leq q^* \leq q-1$, and that the first line of the model in (1.1) is mistakenly replaced by the following:

$$Y_i = X_i'\beta + g(V_{1,i}) + \epsilon_i. \tag{1.2}$$

Assuming for the time being that $E(\epsilon|x) = 0$, in this case the exogeneity assumption is satisfied only if $E(y|v_1) = E(y|v)$ and $E(x|v_1) = E(x|v)$; see also a model specification test discussed in Li (1999). Hereafter, let us refer to the cases where $E(\epsilon|v) \neq 0$ as *"nonparametric endogeneity"*. Furthermore, let us assume the following:

$$X_i = m_x(V_i) + U_i, \tag{1.3}$$

where $E(u|v) = 0$. The corresponding *"parametric endogeneity"* may occur if the equation below is employed instead of (1.3):

$$X_i = m_x(V_{1,i}) + U_i, \tag{1.4}$$

where $E(u|v_1) = 0$. Given that the PL model in (1.1) is correctly specified, then the use of (1.2) and (1.4) leads to the violation of the exogeneity condition, i.e. $E(\epsilon|x, v) \neq 0$.

Regarding the latter, let us illustrate such a problem with an empirical example of the relationship between the logarithm of wages, and the covariates of education (in years) and working experience (in years). While the individual effects of covariates should be examined, their interaction effects should also be considered. For instance:

$$\log(wage_i) = age_i\beta + g(edu_i) + H(agedu_i) + e_i, \tag{1.5}$$

where $agedu_i = age_i \times edu_i$ and $E(e|age, edu) = 0$. Although the identification of each function in (1.5) can be dealt with as shown by Linton and Nielsen (1995), the endogeneity problem could be present if the interaction term $H(\cdot)$ was mistakenly omitted from the model. We present a detailed review of the effects of the various types of endogeneity on the estimation of the PL model in Section 2.2.

The goal of the current paper is to develop a systematic approach for addressing the endogeneity issues in the PL model. Although the PL model has been extensively studied in the literature as reviewed earlier (see also Fan and Li (1999), and Härdle et al. (2000) among others), the endogeneity issues have only recently been considered in pure nonparametric and semiparametric models (for example Blundell and Powell (2003)). A systematic estimation procedure and method that are capable of satisfactorily addressing endogeneity problems in the PL model have yet to be developed owing to its relative complexity in the sense that it contains both parametric and nonparametric components. In this paper, we intend to comprehensively discuss the issues that are essential in dealing with endogeneity problems in the PL model, such as identifying whether they originated from the parametric component, the nonparametric component or both, and appropriate estimation procedures to deal with different types of the problems. While we will summarise of the key contributions of the current paper at the end of this introduction, in the next few paragraphs we will give a brief overview of the methods to be discussed in this paper.

In principle, the methods considered in this paper closely follow the logic of the Robinson's (1988) two-step estimation procedure mentioned previously, i.e. first to obtaining consistent estimators of the unknown parameters and then using them in order to identify an unknown structural function. If the parametric regressors are exogenous, the LS estimation is consistent as also reviewed previously. Otherwise, if the parametric-endogeneity is present, then the parametric instrumental variable (PIV) estimation can be used. Although the PIV estimation has been considered in the literature (see Chapter 16 of Li and Racine (2007), for example) we feel that there are still some outstanding issues which are worth discussing within the context of our study. After all, special attention should be given to constructing consistent estimators of the unknown parameters especially given the dominance of the parametric component of the model. The consistency of parametric estimators is important not only in its own right but also for identifying an unknown structural function (more details can be found in Sections 2.2 and 2.3).

In addition, the presence of nonparametric endogeneity induces further complicates the identification of the unknown structural function (see Section 2.4 for details). There are two alternative methods in the literature which may be helpful in identifying the unknown structural function in such a case, namely the nonparametric instrumental variable (NpIV) estimation and the control function (CF) approach. Newey and Powell (2003), Hall and

Horowitz (2005), and Darolles et al. (2011) developed the NpIV estimation for a pure nonparametric model, while Ai and Chen (2003) did so for semi-parametric models, which included the PL model as a special case. An important difficulty with using NpIV estimation resides in the well-known "ill-posed inverse" problem; see O'Sullivan (1986) for example. To overcome such an obstacle, Ai and Chen (2003) based their estimation on a complex sieve estimation under some regularity conditions on the inversion matrix and a constraint on the space of the reduced relation to keep it compact. On the other hand, Newey et al. (1999), Pinkse (2000), and Su and Ullah (2008) considered the CF approach in a pure nonparametric model, while Blundell and Powell (2004) did so for a special case of a single index model, i.e. a case where the discrete dependent variable only was considered. With regard to the nonparametric estimation employed, Newey et al. (1999) and Pinkse (2000) relied on the series approximation, while Su and Ullah (2008) used the local polynomial estimation of Fan and Gijbels (1996). Blundell and Powell (2004), on the other hand, relied on the local constant kernel estimation.

This paper addresses nonparametric endogeneity in the estimation and inference of the PL model in a simple but widely-used framework of nonparametric simultaneous equations specifically, a nonparametric triangular model. Although the full details will be presented later, let us discuss this briefly here. We consider a model $y = x'\beta + g(v) + \epsilon$ such that x might be either exogenous or endogenous, and v is endogenous. In addition, a nonparametric reduced-form equation $v = m_v(z) + \eta$, where z is a vector of the instrumental variables such that $E(\eta|z) = 0$ and $E(\epsilon|z, \eta) = E(\epsilon|\eta) \neq 0$. In order to identify and to estimate the structural function $g(\cdot)$, we take the standard control function approach, as in Newey et al. (1999), namely $E(y|v, \eta) = E(x|v, \eta)'\beta + g(v) + \iota(\eta)$, where the endogeneity (i.e. $E(\epsilon|\eta) = \iota(\eta) \neq 0$) is controlled by introducing an additional unknown function. Such a structure enables us to write the model as a simple nonparametric additive structure and therefore to employ the local constant kernel estimation and the marginal integration technique of Linton and Nielsen (1995), and Tjøstheim and Austad (1996) to identify the unknown structural function.

To summarise, in this paper we comprehensively study the estimation procedures and methods which provide and identify consistent estimators of the unknowns in the PL model, namely the parametric parameters and the nonparametric structural function, when parametric endogeneity or/and nonparametric endogeneity is/are present. Firstly, we extend the CF approach suggested in Newey et al. (1999) to the PL model with nonparametric endogeneity. We also provide the asymptotic properties of the estimator of the nonparametric function. Furthermore, we show the $\sqrt{n}$-consistency and asymptotic normality results for the estimators of the unknown parameters under parametric endogeneity. Here, an important difficulty resides in a generated regressor issue, which arises due to the fact that the control regressor, η, (or the so–called control variable as referred to in Blundell and Powell (2004)) is not observable. The generated regressor issue must be taken into account when studying the properties and inference of the estimation procedure.

The remaining of the paper is organised as follows. In Section 2, we first review the PL model without endogeneity, introduce parametric endogeneity and nonparametric endogeneity into the model, then discuss the various issues including identification of endogeneity in the model, and appropriate estimation methods and procedures. Section 3 discusses the adaptive data-driven method to bandwidth selection and derives its asymptotic optimality. In Section 4, we conduct an experimental study to investigate the finite sample properties of the estimators introduced in the current study. Finally, Section 5 concludes the paper, while mathematical proofs of the main results are presented in Appendix.

2. Endogeneity in a Partially Linear Semiparametric Model

In this section, we first introduce the PL model and Robinson's two-step estimation procedure as usually seen in the literature. We then introduce endogeneity into the model and discuss the various issues caused by endogeneity in details. Finally, we discuss and propose appropriate estimation methods for addressing endogeneity in Sections 2.3 and 2.4.

2.1. The PL Model

To estimate the model in (1.1), Robinson (1988) proposed a two-step estimation procedure, which is first obtains consistent estimators of the unknown parameters and then uses them to identify an unknown structural function. In particular, the first step of Robinson's estimation procedure is a simple LS estimation, which is tenable after the unknown $g(\cdot)$-function is partialled-out. That is, we obtain the conditional expectation relation by applying the conditional expectation operator to (1.1), since it satisfies the exogeneity condition:

$$g(v) = E(y|v) - E(x|v)'\beta. \tag{2.1}$$

Subtracting the conditional expectation relation (2.1) from the structural one (1.1) produces a simple linear reduced form:

$$W_i^* = U_i^{*\prime}\beta + \epsilon_i, \tag{2.2}$$

where $E(\epsilon u^*) = 0$. We then have, by defining $m_y^*(v) = E(y|v)$ and $m_x^*(v) = E(x|v)$, $Y_i = m_y^*(V_i) + W_i^*$ and $X_i = m_x^*(V_i) + U_i^*$ with $E(w^*|v) = 0$ and $E(u^*|v) = 0$. Equation (2.2) immediately suggests an infeasible estimator for β by a LS estimation of W_i^* on U_i^*:

$$\overline{\beta}_{LS}^* = \left(\bar{S}_{U^*}\right)^{-1} \bar{S}_{U^*W^*}, \tag{2.3}$$

where the notation for scalar and column vector sequences A_i and B_i are $\bar{S}_{AB} = \frac{1}{n}\sum_{i=1}^n A_i B_i'$ and $\bar{S}_A = \bar{S}_{AA}$. The second step of Robinson's (1988) estimation procedure is to identify the structural $g(\cdot)$-function based on the conditional expectation relation in (2.1):

$$\overline{g}_{LS}^*(v) = m_y^*(v) - m_x^*(v)'\overline{\beta}_{LS}^*. \tag{2.4}$$

Nonetheless, the estimators in (2.3) and therefore in (2.4) are infeasible due to the unknown functions of $m_y^*(v)$ and $m_x^*(v)$. Robinson (1988) suggests that $m_y^*(\cdot)$ and $m_x^*(\cdot)$ should be estimated first by a local constant kernel estimation and these can be used in order to obtain feasible estimators. Let us introduce the even functions $k : \mathbb{R} \to \mathbb{R}$ and $K : \mathbb{R}^q \to \mathbb{R}$ related by:

$$K_s(s) = \prod_{j=1}^q k(s_j),$$

where s_j is the jth element in s and k is a univariate kernel function. Now, the above-mentioned feasible estimators are:

$$\hat{\beta}_{LS}^* = \left(S_{\hat{U}^*}\right)^{-1} S_{\hat{U}^*\hat{W}^*} \quad \text{and} \quad \hat{g}_{LS}^*(v) = \hat{m}_y^*(v) - \hat{m}_x^*(v)'\hat{\beta}_{LS}^*$$

by which $S_{AB} = \frac{1}{n}\sum_{i=1}^n A_i B_i' I_i$ and $S_A = S_{AA}$ for the scalar and column vector sequences A_i and B_i, and a constant $b > 0$, $I_i = I(|\hat{f}(V_i)| > b)$, where $\hat{f}(v)$ is the estimate of the probability density function of v with a random argument V_i, I is the usual indicator function, and $\hat{U}_i^* = X_i - \hat{m}_x^*(V_i)$ and $\hat{W}_i^* = Y_i - \hat{m}_y^*(V_i)$ with:

$$\hat{m}_x^* \equiv \hat{E}(x|v) = \frac{\sum_{i=1}^n X_i K_v\left(\frac{v-V_i}{h_v}\right)}{\sum_{j=1}^n K_v\left(\frac{v-V_j}{h_v}\right)} \quad \text{and} \quad \hat{m}_y^* \equiv \hat{E}(y|v) = \frac{\sum_{i=1}^n Y_i K_v\left(\frac{v-V_i}{h_v}\right)}{\sum_{j=1}^n K_v\left(\frac{v-V_j}{h_v}\right)}.$$

Note that I is introduced in order to trim out small values of $\hat{f}(v)$ that is in order to overcome the random denominator problem (see Fan et al. (1995) and Li and Woodridge (2002), for alternative methods). Based on i.i.d. random sample (Y_i, X_i', V_i'), it has been shown that the parameter vector β in various versions of (1.1) can be consistently estimated at $\sqrt{n}$-rate; see Robinson (1988) and Fan et al. (1995), for example.

Remark 2.1. *Robinson (1988) introduced two factors which are essential in the establishment of $\sqrt{n}$-consistency for $\hat{\beta}^*$. These are the higher-order kernel function and the local*

Lipschitz type of condition for smoothness on the functions. The higher-order kernel function reduces bias when the sufficient smoothness condition imposed on the functions and hence they ensure $\sqrt{n}$-consistency. We restate these definitions in Section 2.4 in the context of the estimation method introduced in this paper. ■

2.2. Endogeneity in the PL Model

Prior to presenting a more detailed discussion of the method studied in this paper, let us present a review of the effects that various types of endogeneity have on the model and some suggested remedies. Let us begin with the linear reduced form of the model in (2.2):

$$W_i^* = U_i^{*\prime}\beta + e_i^*, \tag{2.5}$$

where $e^* = \epsilon - E(\epsilon|v) \equiv \epsilon - \iota(v)$. If nonparametric regressors are endogenous, then $\iota(v) \neq 0$. Hence, it is apparent that nonparametric endogeneity induces the problem of identifying a structural $g(\cdot)$-function as follows:

[1.A] $E(y|v) - E(x|v)'\beta = g(v)$, when nonparametric regressors are exogenous;

[1.B] $E(y|v) - E(x|v)'\beta = g(v) + \iota(v)$, when nonparametric regressors are endogenous.

The exogeneity moment condition of (2.5), i.e., $E(e^*u^*) = 0$, is satisfied, unless parametric-endogeneity is present. This moment condition implies two possible cases, namely:

[2.A] $E(\epsilon u^*) = 0$ and $E[\iota(v)u^*] = 0$, i.e. when $\iota(v) = 0$;

[2.B] $E(\epsilon u^*) \neq 0$ and $E(\epsilon u^*) = E[\iota(v)u^*]$, i.e. when $\iota(v) \neq 0$.

While the conditions in [2.A] suggest that the model is endogeneity-free, those in [2.B] suggest that only nonparametric endogeneity is present since the linear reduced form satisfies the moment condition. The fact that the nonparametric endogeneity is partialled-out in the Robinson's transformation suggests that the LS estimation of the unknown parameters is applicable for both cases, i.e. in [2.A] and [2.B]. However, [2.B] is similar to [1.B] in the sense that we must address the nonparametric-endogeneity in order to identify an unknown structural function.

The final remaining case is the presence of parametric-endogeneity such that:

[3.A] $E(\epsilon|x) \neq 0$ so that $E(e^*u^*) \neq 0$ when $\iota(v) \neq 0$ and $E(\epsilon u^*) \neq 0$ otherwise.

In the other words, if parametric regressors are endogenous, then the moment condition of the linear reduced form is not satisfied, i.e. $E(e^*u^*) \neq 0$. In this case, the LS estimation results in inconsistent estimators for both of the unknowns. This stresses the dominance of the parametric part of the model. Although the consistency of a nonparametric estimator is unnecessary for obtaining consistent estimators of the parametric ones (due to Robinson's partialling-out process), the opposite is not true. Let us define $m_0(v) = g(v)$ and $m_1(v) = g(v) + \iota(v)$ with $\iota(v) \neq 0$ in order to illustrate the argument more conveniently. We have:

$$\overline{\beta}^*_{LS} = \beta + \left(\bar{S}_{U^*}\right)^{-1} \bar{S}_{U^* e^*} \overset{p}{\nrightarrow} \beta$$

and:

$$\overline{m}^*_{LS,s}(v) = E(y|v) - E(x|v)' \left\{\beta + \left(\bar{S}_{U^*}\right)^{-1} \bar{S}_{U^* e^*}\right\} \overset{p}{\nrightarrow} m_s(v),$$

where $\overset{p}{\nrightarrow}$ denotes non-convergence in probability, and $m^*_{LS,s}(v) = m^*_{LS,0}(v)$ or $m^*_{LS,1}(v)$, since $\bar{S}_{U^* e^*} \overset{p}{\nrightarrow} 0$. Hence, various issues regarding the identification of endogeneity in parametric regressors and obtaining tenable and consistent parametric estimators are nontrivial, and we discuss these in details in the next section.

Remark 2.2. *In this section, we consider only the moment condition, $E(e^* u^*)$, rather than the conditional moment condition, $E(e^*|u^*)$, of the linear reduced form (2.5), since we discuss the endogeneity issues in terms of infeasible estimators such as (2.3) and (2.4). Note that the cost of having the former rather than the latter is that a more restrictive moment bound on the dependant regressor than that of Robinson (1988) is imposed to establish the $\sqrt{n}$–consistency. If we consider the former then $E(y)^4 < \infty$ is required rather than $E(y)^2 < \infty$ due to the remainder terms, $S_{\hat{u}^* e^*}$ and $S_{u^* \hat{e}^*}$; see Propositions 10 and 11 in appendix of Robinson (1988), for example. Hereafter, we consider the conditional moment condition.* ■

Table 2.1 below summarises the effects and remedies of various sources of endogeneity as discussed above.

Table 2.1. *The effects of endogeneity on the PL model and the appropriate estimation methods*

Types of Endogeneity	Estimators	Effects on estimators	Remedies
Parametric	LSoPP*	Inconsistent	PIV or 2SLS
	NSF**	Unidentifiable	(automatically resolved)
Nonparametric	LSoPP	Consistent	
	NSF	Unidentifiable	NpIV or CF
Both	LSoPP	Inconsistent	PIV or 2SLS
	NSF	Unidentifiable	NpIV or CF

∗ *LS of Parametric Parameters (LSoPP);* ∗∗ *Nonparametric Structural Function (NSF)*

2.3. Parametric-Endogeneity

Let us first consider the case [3.A] above, i.e. the presence of parametric-endogeneity, which may be associated with the linear reduced form model below:

$$W_i^* = U_i^{*\prime}\beta + e_i^*, \tag{2.6}$$

where $E(e^*|u^*) \neq 0$ when $\iota(v) \neq 0$, and $E(\epsilon|u^*) \neq 0$ otherwise. Let us consider the Robinson (1988) type of an IV estimation as follows. Suppose that ϱ^* is an IV vector for U^* such that:

$$\mathcal{Z}_i = m_{\mathcal{Z}}^*(V_i) + \varrho_i^*, \tag{2.7}$$

where $\mathcal{Z}$ is a $\mathbb{R}^p$-valued IV vector for X, $m_{\mathcal{Z}}^*(v) = E(\mathcal{Z}|v)$ and $E(\varrho^*|v) = 0$. Furthermore, we assume that $E(x\mathcal{Z}) \neq 0$ suggests $E(\varrho^* u^*) \neq 0$ and $E(\varrho^*|\epsilon) = 0$ implies $E(\varrho^*|e^*) = 0$, where nonparametric regressors are exogenous; otherwise, they are endogenous. Unlike the NpIV estimation, which requires the conditional moment condition, the PIV estimation also allows for the moment condition as stated in Remark 2.2; $E(\varrho^*\epsilon) = 0$ implies that $E(\varrho^* e^*) = 0$. In this case, we replace $E(y)^2 < \infty$ with $E(y)^4 < \infty$ in Assumption A.0.4 due to the remainder terms $S_{\hat{\varrho}^* e^*}$ and $S_{\varrho^* \hat{e}^*}$.

The Robinson (1988) type of IV estimators have the form:

$$\hat{\beta}_{IV}^* = (S_{\hat{\varrho}^*\hat{U}^*})^{-1} S_{\hat{\varrho}^*\hat{W}^*} \quad \text{and} \quad \hat{m}_{IV,s}^*(v) = \hat{E}(y|v) - \hat{E}(x|v)'\hat{\beta}_{IV}^*,$$

where $\hat{\varrho}_i^* = \mathcal{Z}_i - \hat{E}(\mathcal{Z}_i|V_i)$, $\hat{E}(\mathcal{Z}|v) = \frac{\sum_{j=1}^n \mathcal{Z}_j K_v\left(\frac{v-V_j}{h_v}\right)}{\sum_{l=1}^n K_v\left(\frac{v-V_l}{h_v}\right)}$, and $\hat{m}_{IV,s}^*(v) = \hat{m}_{IV,0}^*(v)$ or $\hat{m}_{IV,1}^*(v)$; see Chapter 16 of Li and Racine (2007), for its asymptotic normality and $\sqrt{n}$-consistency.

Furthermore, note that the PIV estimation above requires a similar rank condition to that of a conventional parametric case. If the rank condition is not satisfied, i.e. if $rank(\mathcal{Z}) > p$, then it can be shown that two-stage LS (2SLS) estimation is the most optimal, as in a conventional parametric case; see Chapter 5 of Sargan (1998), for example. The most optimal candidate for an IV vector is a projection matrix of a parametric regressor vector in the space of an IV vector:

$$\tilde{\varrho}^* = \varrho^*(\varrho^{*\prime}\varrho^*)^{-1}\varrho^{*\prime}U^*.$$

Then, the 2SLS estimators are:

$$\hat{\beta}_{2SLS}^* = (S_{\hat{\tilde{\varrho}}^*\hat{U}^*})^{-1} S_{\hat{\tilde{\varrho}}^*\hat{W}^*} \quad \text{and} \quad \hat{m}_{2SLS,s}^*(v) = \hat{E}(y|v) - \hat{E}(x|v)'\hat{\beta}_{2SLS}^*,$$

where $S_{\hat{\tilde{\varrho}}^*\hat{U}^*} = S_{\hat{U}^*\hat{\varrho}^*}\left(S_{\hat{\varrho}^*}\right)^{-1} S_{\hat{\varrho}^*\hat{U}^*}$, $S_{\hat{\tilde{\varrho}}^*\hat{W}^*} = S_{\hat{U}^*\hat{\varrho}^*}\left(S_{\hat{\varrho}^*}\right)^{-1} S_{\hat{\varrho}^*\hat{W}^*}$, and $\hat{m}_{2SLS,s}^*(v) = \hat{m}_{2SLS,0}^*(v)$ or $\hat{m}_{2SLS,1}^*(v)$. The asymptotic normality and $\sqrt{n}$-consistency of the 2SLS estimator can be established similarly to those of the PIV ones.

Remark 2.3. *The proofs of $\sqrt{n}$-consistency and the asymptotic normality of $\hat{\beta}^*_{IV}$ and $\hat{\beta}^*_{2SLS}$ are similar to those of $\hat{\beta}^*_{LS}$. For instance, in the case of the PIV estimation, we need to establish that $S_{\mathcal{Z}-\hat{\mathcal{Z}},X-\hat{X}} \xrightarrow{p} \Phi_{\varrho^* U^*}$ with $\Phi_{\varrho^* U^*} \equiv E[(\mathcal{Z}-E(\mathcal{Z}|V))'(X-E(X|V))]$, that $S_{\mathcal{Z}-\hat{\mathcal{Z}}} \xrightarrow{p} \Phi_{\varrho^*}$ with $\Phi_{\varrho^*} \equiv E[(\mathcal{Z}-E(\mathcal{Z}|V))'(\mathcal{Z}-E(\mathcal{Z}|V))]$, and particularly that $\sqrt{n}S_{\mathcal{Z}-\hat{\mathcal{Z}},m_s-\hat{m}^*_s} \xrightarrow{p} 0$ and $\sqrt{n}S_{\mathcal{Z}-\hat{\mathcal{Z}},e-\hat{e}^*} \xrightarrow{D} N(0,\sigma^2\Phi_{\varrho^*})$ with $\sigma^2 = E(e^*)^2$; for the definitions of $S_{\mathcal{Z}-\hat{\mathcal{Z}},X-\hat{X}}$, $S_{\mathcal{Z}-\hat{\mathcal{Z}}}$, $S_{\mathcal{Z}-\hat{\mathcal{Z}},m-\hat{m}^*}$ and $S_{\mathcal{Z}-\hat{\mathcal{Z}},e-\hat{e}^*}$, see the Appendix. The set of additional conditions required for this establishment comprise the moment condition on $\mathcal{Z}$, the smoothness condition on the function $m^*_{\mathcal{Z}}(v)$ and the corresponding regularity condition for the bandwidth. We present these conditions in Appendix 5.A.0 for the sake of convenience. A similar argument is also true for the 2SLS estimation case.* ■

Furthermore, the above-mentioned asymptotic normality of the LS and the PIV estimators make it possible to establish a Hausman (1978) type of mis-specification testing. Given the results of Lemma 2.1 and Corollary 2.6 of Hausman (1978), this can be done as follows. Let us define $\hat{d}^* = \hat{\beta}^*_{IV} - \hat{\beta}^*_{LS}$, $V(\hat{\beta}^*_{LS})$ and $V(\hat{\beta}^*_{IV})$ as the asymptotic variances of $\hat{\beta}^*_{LS}$ and $\hat{\beta}^*_{IV}$, respectively. Then a mis-specification testing in the PL model can be implemented as

$$H_0 : \hat{d}^* \xrightarrow{p} 0 \text{ i.e. } \textit{parametric-exogeneity};\ H_1 : \hat{d}^* \not\xrightarrow{p} 0 \text{ i.e. } \textit{parametric-endogeneity}. \quad (2.8)$$

Under the null hypothesis in (2.8), we have:

$$\hat{\beta}^*_{LS} = \beta + O_p(n^{1/2}) \text{ and } \hat{\beta}^*_{IV} = \beta + O_p(n^{-1/2}),$$

which imply that $\hat{d}^* = o_p(1)$. However, under the alternative hypothesis in (2.8), we have:

$$\hat{\beta}^*_{LS} = \beta + O_p(1) \text{ and } \hat{\beta}^*_{IV} = \beta + O_p\left(n^{-1/2}\right),$$

which suggest that $\hat{d}^* = O_p(1)$. Under the null hypothesis, the asymptotic distribution of the difference of two estimators is:

$$\sqrt{n}(\hat{d}^* - d) \xrightarrow{D} N[0, V(\hat{d}^*)],$$

where $V(\hat{d}^*) = V(\hat{\beta}^*_{IV}) - V(\hat{\beta}^*_{LS})$. As $n \to \infty$, the test statistic is:

$$t = \hat{d}^{*\prime}\left(\hat{V}(\hat{d}^*)\right)^{-1}\hat{d}^* \xrightarrow{D} \chi^2_p,$$

where $\hat{V}(\hat{d}^*)$ is the estimate of $V(\hat{d}^*)$ and p is the number of unknown parameters in the model.

2.4. Nonparametric-Endogeneity

In Section 2.2, we have briefly discussed the effects of nonparametric-endogeneity such that $\iota(v) \equiv E(\epsilon|v) \neq 0$ in the model; see cases [1.B] and [2.B], for example. It is apparent in the conditional expectation relation, $m_y^*(v) - m_x^*(v)'\overline{\beta}_\tau = \overline{m}_{\tau,1}^*(v) = \overline{g}_\tau^*(v) + \iota(v)$, that, in these cases, the structural $g(\cdot)$-function is unidentifiable, where τ is found by using LS, IV and $2SLS$. The procedure considered in this section rests on a semiparametric simultaneous equation model, for which the corresponding nonparametric version has previously been considered by Newey et al. (1999).

Suppose that nonparametric-endogeneity is present in the model and that Z is an instrumental variable vector for V, and let us consider a simultaneous equation model:

$$Y_i = X_i'\beta + g(V_i) + \epsilon_i \tag{2.9}$$

$$V_i = m_v(Z_i) + \eta_i, \tag{2.10}$$

$$E(\epsilon|z,\eta) = E(\epsilon|\eta) \text{ a.s.} \tag{2.11}$$

$$E(\eta|z) = 0 \text{ a.s.} \tag{2.12}$$

where a.s. denotes for almost surely, (2.9) is as defined in (1.1), $m_v(z) \equiv E(v|z)$ is a $q \times 1$ vector of the functions of the instruments, Z is a $\mathbb{R}^{q_z}$-valued vector with $q_z \geq q$ and η is a $q \times 1$ vector of disturbances. It should be noted here that while the stochastic conditions stated in (2.11) and (2.12), which are often referred to as the "control function" assumptions, are more general than assuming full independence between (ϵ, η) and Z, they are neither stronger nor weaker than $E(\epsilon|z) = 0$, which is usually required in the NpIV estimation.

Based on (2.9) to (2.12), we have:

$$E(y|v,\eta) = E(x|v,\eta)'\beta + E(g(v)|v,\eta) + E(\epsilon|v,\eta) = E(x|v,\eta)'\beta + g(v) + E(\epsilon|\eta).$$

This ultimately leads to:

$$g(v) + \iota(\eta) = E(y|v,\eta) - E(x|v,\eta)'\beta, \tag{2.13}$$

where $E(\epsilon|\eta) \equiv \iota(\eta)$ is referred to in the literature as the endogeneity control function.

The first key step in the CF approach in this paper is to estimate the endogeneity control regressors from a structural relation between the endogenous regressors and their instrumental variables, (i.e. expression (2.10) above). Such a structural relation is referred to as "a reduced form" in Blundell and Powell (2004). The next step is to control the endogeneity in the structural relation (2.9) by introducing an endogeneity control function, $\iota(\eta)$. Finally, the nonparametric additive structure derived in (2.13) suggests that the unknown structural function can be identified by using the marginal integration technique of Linton and Nielsen (1995), and Tjøstheim and Austad (1996).

The procedure described above can be implemented in a few estimation steps. Hereafter, let us collectively refer to such estimation steps as the "two-step control function (2SCF) procedure", which can be described as follows:

The 2SCF Procedure

Step 2.1: Estimate the endogeneity control regressor, η_i, from (2.10).

Step 2.2: Obtain consistent estimators $\hat{\beta}_\tau$ of the unknown parameters as in Section 2.3.

Step 2.3: Given the consistent parametric estimators in Step 2.2, estimate the conditional expectation relation.

Step 2.4: Perform the marginal integration technique on the resulting estimated conditional expectation relation in Step 2.3 to estimate the structural $g(\cdot)$-function.

In the remainder of this section, let us discuss each of these steps in more detail. *Step 2.1* estimates the endogeneity control regressors, η, from the reduced form in (2.10) since they are not observable in practice, where $m_v(z)$ is a vector of unknown real functions such that $m_v \equiv (m_v)(Z_i, \ldots, Z_i)'$, $i = 1, \ldots, n$, $m_{v,l} : \mathbb{R}^{q_z} \to \mathbb{R}$ and $l = 1, \ldots, q$. The kernel estimation of $m_{v,l}(Z_i)$ is:

$$\hat{m}_{v,l}(Z_i) = \frac{\sum_{j=1}^{n} V_j K_z \left(\frac{Z_i - Z_j}{h_z} \right)}{\sum_{l=1}^{n} K_z \left(\frac{Z_i - Z_l}{h_z} \right)}, \tag{2.14}$$

where $h_z = (h_{z1}, h_{z2}, \ldots, h_{zq})'$, which leads to:

$$\hat{\eta}_i = V_i - \hat{m}_{v,l}(Z_i). \tag{2.15}$$

The next estimation step is to transform the structural model into a linear reduced form to obtain consistent parametric estimators. This can be done by first decomposing the dependent and independent regressors into two components. By defining $m_y(v, \eta) = E(y|v, \eta)$ and $m_x(v, \eta) = E(x|v, \eta)$, the above-mentioned decompositions are:

$$Y_i = m_y(V_i, \eta_i) + W_i \quad \text{and} \quad X_i = m_x(V_i, \eta_i) + U_i,$$

where $E(w|v, \eta) = 0$ and $E(u|v, \eta) = 0$. Now we can obtain the conditional expectation relation of the structural model on the nonparametric and endogeneity control regressors:

$$m_y(v, \eta) = m_x(v, \eta)'\beta + g(v) + \iota(\eta) \tag{2.16}$$

such that $\iota(\eta) \neq 0$ controls the endogeneity. Finally, if we subtract the conditional expectation relation (2.16) from the structural one, the transformed simple linear reduced form is then:

$$W_i = U_i'\beta + e_i, \tag{2.17}$$

where $W_i = Y_i - E(Y_i|V_i, \eta_i)$, $U_i = X_i - E(X_i|V_i, \eta_i)$ and $e_i = \epsilon_i - \iota(\eta_i)$.

In order to obtain consistent estimators of the unknown parameters, it must be ensured that an appropriate estimation method (i.e. among LS, IV or 2SLS as discussed in Section 2.3) is applied to (2.17). If parametric regressors are exogenous, then it is appropriate to simply apply the LS estimation; otherwise, we must apply the PIV estimation with the following vector of parametric instruments:

$$\mathcal{Z}_i = m_{\mathcal{Z}}(V_i, \eta_i) + \varrho_i,$$

where $m_{\mathcal{Z}}(v, \eta) = E(\mathcal{Z}|v, \eta)$ and $E(\varrho|v, \eta) = 0$, $E(x\mathcal{Z}) \neq 0$ implies that $E(u\varrho) \neq 0$ and $E(\epsilon|\varrho) = 0$ suggests that $E(e|\varrho) = 0$. Furthermore, if the rank of the vector $\mathcal{Z}$ is greater than p, then the 2SLS estimation is applied. Hence, the potential consistent parametric estimators can be summarised as:

$$\hat{\beta}_{LS} = (S_{\hat{U}_2})^{-1} S_{\hat{U}_2\hat{W}_2},\ \hat{\beta}_{IV} = (S_{\hat{\varrho}_2\hat{U}_2})^{-1} S_{\hat{\varrho}_2\hat{W}_2} \text{ and } \hat{\beta}_{2SLS} = (S_{\hat{\hat{\varrho}}_2\hat{U}_2})^{-1} S_{\hat{\hat{\varrho}}_2\hat{W}_2}, \tag{2.18}$$

where:

$$\hat{U}_{2,i} = X_i - \hat{E}(X_i|V_i, \hat{\eta}_i),\ \hat{W}_{2,i} = Y_i - \hat{E}(Y_i|V_i, \hat{\eta}_i),\ \hat{\varrho}_{2,i} = \mathcal{Z}_i - \hat{E}(\mathcal{Z}_i|V_i, \hat{\eta}_i),$$
$$S_{\hat{\hat{\varrho}}_2\hat{U}_2} = S_{\hat{U}_2\hat{\varrho}_2} \left(S_{\hat{\varrho}_2}\right)^{-1} S_{\hat{\varrho}_2\hat{U}_2},\ S_{\hat{\hat{\varrho}}_2\hat{W}_2} = S_{\hat{U}_2\hat{\varrho}_2} \left(S_{\hat{\varrho}_2}\right)^{-1} S_{\hat{\varrho}_2\hat{W}_2},$$

by which $\hat{E}(x|v, \hat{\eta})$, $\hat{E}(y|v, \hat{\eta})$ and $\hat{E}(\mathcal{Z}|v, \hat{\eta})$ are kernel estimators with $\hat{\eta}_i$, i.e.:

$$\hat{E}(x|v, \hat{\eta}) = \frac{\sum_{i=1}^n X_i K_v\left(\frac{v-V_i}{h_v}\right) K_\eta\left(\frac{\hat{\eta}-\hat{\eta}_i}{h_\eta}\right)}{\sum_{j=1}^n K_v\left(\frac{v-V_j}{h_v}\right) K_\eta\left(\frac{\hat{\eta}-\hat{\eta}_j}{h_\eta}\right)}, \hat{E}(y|v, \hat{\eta}) = \frac{\sum_{i=1}^n Y_i K_v\left(\frac{v-V_i}{h_v}\right) K_\eta\left(\frac{\hat{\eta}-\hat{\eta}_i}{h_\eta}\right)}{\sum_{j=1}^n K_v\left(\frac{v-V_j}{h_v}\right) K_\eta\left(\frac{\hat{\eta}-\hat{\eta}_j}{h_\eta}\right)}, \tag{2.19}$$

and:

$$\hat{E}(\mathcal{Z}|v, \hat{\eta}) = \frac{\sum_{i=1}^n \mathcal{Z}_i K_v\left(\frac{v-V_i}{h_v}\right) K_\eta\left(\frac{\hat{\eta}-\hat{\eta}_i}{h_\eta}\right)}{\sum_{j=1}^n K_v\left(\frac{v-V_j}{h_v}\right) K_\eta\left(\frac{\hat{\eta}-\hat{\eta}_j}{h_\eta}\right)}. \tag{2.20}$$

Similar to the first stage, a trimming parameter is employed along the way to minimize the impact of a random denominator problem. For a constant $b_2 > 0$, let $I_{2,i} = I(|\hat{f}(V_i, \eta_i)| > b_2)$, where $\hat{f}(v, \eta)$ is the estimate of the probability density function $f(v, \eta)$ with a random argument (V_i, η_i).

The essential factors which helps ensuring the asymptotic consistency as discussed in Remark 2.1 are defined below.

Definition 2.1: *Let the even functions* $k_z : \mathbb{R} \to \mathbb{R}$, $k_v : \mathbb{R} \to \mathbb{R}$, $k_\eta : \mathbb{R} \to \mathbb{R}$ *and* $k_\eta^{(r)} : \mathbb{R} \to \mathbb{R}$, *which is the* r*th derivative of* k_η. *Let* $K_z : \mathbb{R}^m \to \mathbb{R}$, $K_v : \mathbb{R}^q \to \mathbb{R}$, $K_\eta^{(r)} : \mathbb{R}^q \to \mathbb{R}$ *and* $L^{(r)} : \mathbb{R}^{2q} \to \mathbb{R}$, *be related by* $K_z = \prod_{j=1}^{q_z} k_z(s_j)$, $K_v = \prod_{j=1}^{q} k_v(s_j)$, $K_\eta^{(r)} = \prod_{j=1}^{q} k_\eta^{(r)}(s_j)$, *and* $L^{(r)}(s) = K_v(s)K_\eta^{(r)}(s)$, *where* $r = 0, 1, \ldots, \omega_2 - 1$ *for some* $\omega_2 > 0$.

Definition 2.2: $\mathcal{G}_\mu^\alpha$, *where* $\alpha > 0$ *and* $\mu > 0$, *is the class of functions* $g : \mathbb{R}^q \to \mathbb{R}$ *that satisfy the following conditions:* g *is* $(l-1)$ *times partially differentiable for* $l - 1 \le \mu \le l$; *for some* $\rho > 0$, $\sup_{y \in \phi_{z\rho}} |g(y) - g(z) - Q_g(y,z)|/||y - z||^\mu \le G_g(z)$ *for all* z, *where* $\phi_{z\rho} = \{y : ||y - z|| < \rho\}$; $Q_g = 0$ *when* $l = 1$; Q_g *is a* $(l-1)$*th degree homogeneous polynomial in* $y - z$ *with the coefficients the partial derivatives of* g *at* z *of orders 1 through* $l - 1$; *and* $g(z)$ *are its partial derivatives of order* $l - 1$ *and less, and* $G_g(z)$ *has finite* α*th moments.* $\mathcal{G}_\mu^\infty$ *contains the bounded and* $(l-1)$ *times boundedly differentiable functions whose* $(l-1)$*th partial derivatives are in Lip(*$\mu - l + 1$*), i.e. the Lipschitz class of degree* $\mu - l + 1$.

The main theoretical results for this particular step is the $\sqrt{n}$-consistency and the asymptotic normality of the Robinson-type of LS, PIV and 2SLS estimators as stated below.

Theorem 2.1. *Under Assumptions A.1.1-A.1.6, the condition that* Φ_U *is positive and definite is necessary and sufficient for:*

$$\begin{aligned} \sqrt{n}(\hat{\beta}_{LS} - \beta) &\overset{D}{\to} N[0, \Phi_U^{-1}\sigma^2] \\ S_{\hat{U}_2}^{-1}\hat{\sigma}^2 &\overset{p}{\to} \Phi_U^{-1}\sigma^2, \end{aligned} \tag{2.21}$$

where $E(e^2) = \sigma^2 < \infty$, $\Phi_U = E[\{X - E(X|V,\eta)\}'\{X - E(X|V,\eta)\}]$. ■

Corollary 2.1. *Under Assumptions A.1.1, A.1.3 - A.1.7, A.1.8 and A.1.9, the conditions that* $\Phi_{\varrho U}, \Phi_\varrho$ *and* $\Phi_{U\varrho}$ *are positive and definite are necessary and sufficient for:*

$$\begin{aligned} \sqrt{n}(\hat{\beta}_{IV} - \beta) &\overset{D}{\to} N[0, (\Phi_{\varrho U})^{-1}\sigma^2\Phi_\varrho(\Phi_{U\varrho})^{-1}] \\ (S_{\hat{\varrho}_2\hat{U}_2})^{-1}\hat{\sigma}^2 S_{\hat{\varrho}_2}(S_{\hat{U}_2\hat{\varrho}_2})^{-1} &\overset{p}{\to} (\Phi_{\varrho U})^{-1}\sigma^2\Phi_\varrho(\Phi_{U\varrho})^{-1}. \end{aligned} \tag{2.22}$$

When $rank(\mathcal{Z}) > p$ *we have:*

$$\begin{aligned} \sqrt{n}(\hat{\beta}_{2SLS} - \beta) &\overset{D}{\to} N\left[0, \sigma^2\left(\Phi_{U\varrho}\left(\Phi_\varrho\right)^{-1}\Phi_{\varrho U}\right)^{-1}\right] \\ \hat{\sigma}^2\left(S_{\hat{U}_2\hat{\varrho}_2}\left(S_{\hat{\varrho}_2}\right)^{-1}S_{\hat{\varrho}_2\hat{U}_2}\right)^{-1} &\overset{p}{\to} \sigma^2\left(\Phi_{U\varrho}\left(\Phi_\varrho\right)^{-1}\Phi_{\varrho U}\right)^{-1}, \end{aligned}$$

where $\Phi_{\varrho U} = E[\{\mathcal{Z}-E(\mathcal{Z}|V,\eta)\}'\{X-E(X|V,\eta)\}]$, $\Phi_{\varrho} = E[\{\mathcal{Z}-E(\mathcal{Z}|V,\eta)\}'\{\mathcal{Z}-E(\mathcal{Z}|V,\eta)\}]$ *and* $\Phi_{U\varrho} = E[\{X - E(X|V,\eta)\}'\{\mathcal{Z} - E(\mathcal{Z}|V,\eta)\}]$. ■

Remark 2.4. *The proofs of* $\sqrt{n}$*-consistency and the asymptotic normality of* $\hat{\beta}_{IV}$ *and* $\hat{\beta}_{2SLS}$ *are similar to that of* $\hat{\beta}_{LS}$*. For instance, in the case of the PIV estimation, we need to establish* $S_{\mathcal{Z}-\hat{\mathcal{Z}},X-\hat{X}} \xrightarrow{p} \Phi_{\varrho U}$ *and* $S_{\mathcal{Z}-\hat{\mathcal{Z}}} \xrightarrow{p} \Phi_{\varrho}$*, particularly for* $\sqrt{n}S_{\mathcal{Z}-\hat{\mathcal{Z}},m_s-\hat{m}_s} \xrightarrow{p} 0$ *and* $\sqrt{n}S_{\mathcal{Z}-\hat{\mathcal{Z}},e-\hat{e}} \xrightarrow{D} N(0,\sigma^2\Phi_{\varrho})$*; see the Appendix for the definitions of the notations for* $S_{\mathcal{Z}-\hat{\mathcal{Z}},X-\hat{X}}$*,* $S_{\mathcal{Z}-\hat{\mathcal{Z}}}$*,* $S_{\mathcal{Z}-\hat{\mathcal{Z}},m-\hat{m}}$*, and* $S_{\mathcal{Z}-\hat{\mathcal{Z}},e-\hat{e}}$*.* ■

The implications of these results on the above-mentioned Hausman (1978) type of misspecification testing as follows. Under the null hypothesis of no parametric-endogeneity, we have, $\hat{\beta}_{LS} = \beta + O_p(n^{-1/2})$ and $\hat{\beta}_{IV} = \beta + O_p(n^{-1/2})$, which suggest that $\hat{d} = o_p(1)$, where $\hat{d} = \hat{\beta}_{IV} - \hat{\beta}_{LS}$. However, under the alternative hypothesis of the presence of parametric-endogeneity, we have, $\hat{\beta}_{LS} = \beta + O_p(1)$ and $\hat{\beta}_{IV} = \beta + O_p(n^{-1/2})$, which lead to $\hat{d} = O_p(1)$. Under the null hypothesis, the asymptotic distribution of the difference between two estimators is:

$$\sqrt{n}(\hat{d} - d) \xrightarrow{D} N[0, V(\hat{d})], \tag{2.23}$$

where $V(\hat{d}) = V(\hat{\beta}_{IV}) - V(\hat{\beta}_{LS})$. These asymptotic variances are given above as (2.21) and (2.22), respectively. As $n \to \infty$, the test statistic is:

$$t = \hat{d}' \left(\hat{V}(\hat{d})\right)^{-1} \hat{d} \to_D \chi^2_p, \tag{2.24}$$

where $\hat{V}(\hat{d})$ is the estimate of $V(\hat{d})$.

The objective of the final two steps in this estimation procedure, i.e. Steps 2.3 and 2.4, is to identify the structural $g(\cdot)$-function, given the consistent parametric estimators obtained in the earlier step. Let us first recall the conditional expectation of (2.16):

$$m(v,\eta) = m_y(v,\eta) - m_x(v,\eta)'\beta = g(v) + \iota(\eta). \tag{2.25}$$

Clearly the right-hand side can be treated as a general nonparametric additive model for which a standard identification condition is $E(g(v)) = E(\iota(\eta)) = 0$; see Hastie and Tibshirani (1991), and Gao (2007) for example. Since (2.25) is a simple nonparametric additive specification, an implementation of the so-called marginal integration technique identifies the $g(\cdot)$-function up to some constant value, i.e.:

$$m(v) = \int m(v,\eta)dQ(\eta) = g(v) + c_1 \quad \text{and} \quad m(\eta) = \int m(v,\eta)dQ(v) = \iota(\eta) + c_2, \tag{2.26}$$

where $c_1 = \int \iota(\eta)dQ(\eta)$, $c_2 = \int g(v)dQ(v)$ and Q is a deterministic weighting function with $\int dQ(\eta) = \int dQ(v) = 1$. Linton and Nielsen (1995) allow for both discrete and continuous values of Q, while the integrals should be interpreted in the Stieltjes sense. To this end, the functions $m(v)$ and $g(v)$ can be estimated by the sample versions of (2.26):

$$\hat{m}_\tau(v) = \frac{1}{n}\sum_{i=1}^{n} \hat{m}_\tau(v, \hat{\eta}_i) \tag{2.27}$$

and:

$$\hat{g}_\tau(v) = \hat{m}_\tau(v) - \hat{c}_{\tau,1},$$

where $\hat{m}_\tau(v, \hat{\eta}_i) = \hat{E}(y|v, \hat{\eta}_i) - \hat{E}(x|v, \hat{\eta}_i)'\hat{\beta}_\tau$ and $\hat{c}_{\tau,1} = \frac{1}{n}\sum_{i=1}^{n} \hat{m}_\tau(V_i)$, such that (2.27) is estimated by keeping V_i at v and taking an average over the remaining regressor, $\hat{\eta}_i$. We state the asymptotic properties of the nonparametric estimator below.

Theorem 2.2. *Under Assumptions A.1.1 - A.1.6 when the parametric regressors are exogenous, or else, under Assumptions A.1.1, A.1.3 - A.1.7, A.1.8 and A.1.9, we have:*

$$\sqrt{nh_v^q}(\hat{g}_\tau(v) - g(v) - bias) \to_D N(0, var),$$

where $bias = h_v^{p_2} B_v(v, \eta) + h_\eta^{p_2} B_\eta(v, \eta)$ *with* $B_v(v, \eta) = \frac{\mathcal{K}_{v,p_2}}{f(v,\eta)} \sum_{r=1}^{p_2} f_v^{(r)}(v, \eta) m^{(p_2-r)}(v)$, $B_\eta(v, \eta) = \frac{\mathcal{K}_{\eta,p_2}}{f(v,\eta)} \sum_{r=1}^{p_2} f_\eta^{(r)}(v, \eta) m^{(p_2-r)}(\eta)$, $\mathcal{K}_{v,p_2} = \int v^{p_2} K_v(v) dv$, $\mathcal{K}_{\eta,p_2} = \int \eta^{p_2} K_\eta(\eta) d\eta$, $f_v^{(r)}(v, \eta)$ *and* $f_\eta^{(r)}(v, \eta)$ *are the* r*th derivatives of the joint probability density functions of* (v, η) *with respect to* v *and* η *respectively, and* $var = \sigma^2(v, \eta)\mathcal{K}_v f(v) \frac{f(\eta)^2}{f(v,\eta)^2}$ *with* $\mathcal{K}_v = \int K(v)^2 dv$. ■

3. Adaptive Data-Driven Estimation

In this section, we apply the cross-validation (CV) criterion to the adaptive data-driven approach to bandwidth selection. Without loss of generality, let us first assume that only nonparametric-endogeneity is present in the model. Under such an assumption, the adaptive data-driven estimation problem, which is the main focus of the study in this section, directly involves Steps 2.2 and 2.3 of the 2SCF procedure discussed in Section 2. In order to define the CV function, we need to introduce a number of leave-one-out estimators corresponding to those in (2.18) and (2.19). Hereafter, let $\hat{E}_{-i}(y|v, \hat{\eta})$ and $\hat{E}_{-i}(x|v, \hat{\eta})$ be the corresponding leave-one-out estimators of $\hat{E}(y|v, \hat{\eta})$ and $\hat{E}(x|v, \hat{\eta})$ in (2.19), respectively. Hence, by following a similar analysis to that discussed in Section 2.4, we are also able to derive the leave-one-out counterparts of the LS estimators $\left(\hat{\beta}_{LS}\right)$ based on $\hat{E}_{-i}(y|v, \hat{\eta})$ and $\hat{E}_{-i}(x|v, \hat{\eta})$; hereafter, let us call these estimators $\hat{\beta}_{-i,LS}$.

The relevant CV function can now be written as:

$$\hat{CV}(h_v, h_{\hat{\eta}}) = \frac{1}{n}\sum_{i=1}^{n} \left[Y_i - X_i\hat{\beta}_{-i,LS} - \hat{E}_{-i}(y|v, \hat{\eta}) + \hat{E}_{-i}(x|v, \hat{\eta})\hat{\beta}_{-i,LS}\right]^2. \tag{3.1}$$

The CV criterion consists of selecting the value $\hat{h}_v$ of h_v and $\hat{h}_{\hat{\eta}}$ of $h_{\hat{\eta}}$ that achieve:

$$\hat{CV}(\hat{h}_v, \hat{h}_{\hat{\eta}}) = \inf_{\hat{h}_v, \hat{h}_{\hat{\eta}} \in H_n} \hat{CV}(h_v, h_{\hat{\eta}}). \tag{3.2}$$

The estimation performance can be evaluated using the average squared error below:

$$\hat{D}(h_v, h_{\hat{\eta}}) = \frac{1}{n} \sum_{i=1}^{n} \left[Y_i - X_i \hat{\beta}_{LS} - \hat{E}(y|v, \hat{\eta}) + \hat{E}(x|v, \hat{\eta}) \hat{\beta}_{LS} \right]^2 . \tag{3.3}$$

In the current paper, we take the view that the nonparametric prediction algorithm is asymptotically optimal if:

$$\frac{\hat{D}(\hat{h}_v, \hat{h}_{\hat{\eta}})}{\inf_{h_v, h_\eta \in H_n} D(h_v, h_\eta)} \xrightarrow{P} 1, \tag{3.4}$$

where $D(h_v, h_\eta)$ is the average squared error in the case of observed η. The main result of this section is the following asymptotic optimality.

Proposition 3.1. *Under Assumptions A.1.1 - A.1.6, the selected value of the bandwidths $\hat{h}_v$ of h_v and $\hat{h}_{\hat{\eta}}$ of $h_{\hat{\eta}}$ that satisfy the decision rule presented in (3.2) are asymptotically optimal in the sense of (3.4).*

The proof of such an asymptotic optimality in Proposition 3.1 is quite tedious even for a standard kernel regression case and itself warrants a separated paper; see Hart and Vieu (1990), Härdle and Vieu (1992) for some well-known examples. Nonetheless, since this discussion is only included for the sake of completeness of the methods presented in the current paper, we will simplify the task by making use of some results that we have established elsewhere. Recently, Saart et al. (2013) considered optimal smoothing in the PL model with a nonparametrically-generated regressor. This is, in fact, closely related the procedure described in expressions (3.1) to (3.4). Hence, to keep the current paper short, in Section 6.A.4, we will show the key results that must be satisfied in order for Proposition 3.1 to hold. Interested readers are referred to discussion in Saart et al. (2013) for more details.

4. Simulations

In this section, we discuss Monte Carlo simulation exercises, the objective of which is twofold. Firstly, we would like to investigate the finite sample performance of our newly developed approach in dealing with *nonparametric-endogeneity* and/or *parametric-endogeneity* in the estimation of the PL model, as discussed above. Generally, our learning strategy involves establishing an exogenous PL model then systematically introducing (parametric and/or nonparametric) endogeneity into the model; applying the existing estimation procedure (e.g.

Robinson's (1988) procedure as discussed in Section 2.1) in order to investigate its effectiveness in the presence of endogeneity; and finally applying our newly developed approach to the same endogenous models in order to investigate its effectiveness as an alternative method in the presence of endogeneity. Secondly, we would to examine whether experimental evidence can be established in support of the asymptotic optimality established in Proposition 3.1.

All simulations are conducted in R with the number of replications set at 1000. The normal kernel function defined as $K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right)$ is used throughout this section, while bandwidth selection is implemented as explained in Section 3. For convenience, let us summarise some important notations and abbreviations in this paragraph, which will be used throughout the remaining of this section. Hereafter, *2SR* and *2SR–PIV* refer to Robinson's (1988) procedure as discussed in Section 2.3 with the LS and IV estimators of the unknown parameter, respectively. Furthermore, *2SCF* refers to the control function approach as explained in Section 2.4. In the tables that follow, $\hat{\beta}$, "Bias", "Var" and $|\hat{\beta} - \beta|$ refer to the estimate of the unknown parameter, bias, variance and the absolute error, respectively. Moreover, $\mathrm{ae}_{\hat{g}}$ denotes the average absolute error for the estimation of the nonparametric structural function. The averages of these over the above-stated number of replications are tabulated in Tables 4.1 to 4.9.

Let us focus first on the introduction and modelling of nonparametric endogeneity.

Nonparametric endogeneity

For the sake of comparisons, we will employ the PL model in (4.1) as a baseline model:

$$\begin{aligned} Y_i = 1.2X_i + 0.5\left(\frac{V_i}{1+V_i^2}\right) + \epsilon_i, \text{ where} \\ V_i = Z_i + \eta_i, \quad X_i = \sin(V_i - \eta_i) + U_i, \\ \epsilon_i = \iota(\eta_i) + e_i, \\ Z_i \sim \mathcal{U}(0,3),\ \eta_i \sim \mathcal{U}(-1,1),\ e_i, U_i \sim N(0,1). \end{aligned} \tag{4.1}$$

Defining the PL model as in (4.1) gives rise to three related types of model, namely the *"exogenous model"*, *"linear-endogenous model"* and the *"nonlinear-endogenous model"*, simply by specifying, for example:

$$\iota(\eta) = 0 \times \eta,\ \iota(\eta) = 1 \times \eta \text{ and } \iota(\eta) = \frac{\eta}{1+\eta^2}, \tag{4.2}$$

respectively. While Tables 4.1, 4.2 and 4.3 present the estimation results of the exogenous model, the linear-endogeneity model and the nonlinear-endogeneity model, respectively, based on the 2SR procedure, Tables 4.4 and 4.5 summarise those obtained based on the 2SCF procedure.

Table 4.1. *Exogenous model with 2SR*

n	$\hat{\beta}$	Bias	Var	$\|\hat{\beta}-\beta\|$	$ae_{\hat{g}}$	$\hat{h}_v$
100	1.2007	0.0007	0.0001	0.0103	0.0231	0.2058
300	1.2012	0.0012	0.0000	0.0063	0.0145	0.1817
500	1.1998	-0.0001	0.0000	0.0054	0.0112	0.1637
700	1.2005	0.0005	0.0000	0.0043	0.0102	0.1525
900	1.2002	0.0001	0.0000	0.0037	0.0099	0.1458
1,100	1.2002	0.0002	0.0000	0.0034	0.0090	0.1387

Table 4.2. *Linear-Endogeneity model with 2SR*

n	$\hat{\beta}$	Bias	Var	$\|\hat{\beta}-\beta\|$	$ae_{\hat{g}}$	$\hat{h}_v$
100	1.1908	-0.0092	0.0045	0.0536	0.2465	0.2755
300	1.1955	-0.0044	0.0013	0.0293	0.2311	0.2129
500	1.1958	-0.0042	0.0009	0.0259	0.2321	0.1938
700	1.1967	-0.0032	0.0006	0.0198	0.2280	0.1795
900	1.1960	-0.0040	0.0004	0.0172	0.2310	0.1678
1,100	1.1969	-0.0031	0.0003	0.0166	0.2289	0.1646

Table 4.3. *Nonlinear-Endogeneity model with 2SR*

n	$\hat{\beta}$	Bias	Var	$\|\hat{\beta}-\beta\|$	$ae_{\hat{g}}$	$\hat{h}_v$
100	1.1964	-0.0036	0.0006	0.0205	0.0848	0.2304
300	1.1990	-0.0010	0.0002	0.0114	0.0762	0.1905
500	1.1970	-0.0030	0.0001	0.0105	0.0765	0.1708
700	1.1973	-0.0027	0.0000	0.0076	0.0748	0.1614
800	1.1983	-0.0017	0.0000	0.0066	0.0750	0.1542
1,100	1.1976	-0.0024	0.0000	0.0063	0.0748	0.1484

Table 4.4. *Linear-Endogeneity model with 2SCF*

n	$\hat{\beta}$	Bias	Var	$\|\hat{\beta}-\beta\|$	$ae_{\hat{g}}$	$\hat{h}_v$
100	1.1988	-0.0011	0.0003	0.0153	0.0841	0.3433
300	1.2008	0.0008	0.0000	0.0074	0.0557	0.2445
500	1.1998	-0.0001	0.0000	0.0056	0.0470	0.2100
700	1.2007	0.0007	0.0000	0.0044	0.0424	0.1970
900	1.2000	0.0000	0.0000	0.0040	0.0396	0.1835
1,100	1.2004	0.0004	0.0000	0.0036	0.0307	0.1759

Table 4.5. *Nonlinear-Endogeneity model with 2SCF*

n	$\hat{\beta}$	Bias	Var	$\lvert\hat{\beta}-\beta\rvert$	$ae_{\hat{g}}$	$\hat{h}_v$
100	1.2019	0.0019	0.0004	0.0114	0.0403	0.4703
300	1.2013	0.0013	0.0001	0.0066	0.0273	0.1965
500	1.2005	0.0005	0.0001	0.0055	0.0213	0.2130
700	1.2002	0.0002	0.0000	0.0043	0.0183	0.1793
800	1.1994	-0.0006	0.0000	0.0037	0.0165	0.1625
1,100	1.2001	0.0001	0.0000	0.0035	0.0104	0.1536

Let us now discuss some important findings as follows. While the *2SR* procedure performs well for the exogenous model, the presence of nonparametric-endogeneity (either linear-endogeneity or nonlinear-endogeneity) can cause a significant problem in the estimation of the nonparametric structural function; see the sixth column of Tables 4.2 and 4.3 in particular. Judging from the tendency of the average of $|\hat{\beta}-\beta|$ to converge to zero as $n \to \infty$, the presence of nonparametric-endogeneity in the model does not seem to cause a significant problem in the LS estimation of the unknown parametric parameter. For a given instrument with a specific explanatory power, represented by Z, it is interesting to see that the linear-endogeneity seems to have a greater impact on the *2SR* estimation than its nonlinear-endogeneity counterpart. Compared to the results shown in Tables 4.2 and 4.3, those in Tables 4.4 and 4.5 suggest that the *2SCF* procedure is able to provide a much better estimation of the nonparametric structural function in the presence of nonparametric-endogeneity. Furthermore, the results are robust across the various types of endogeneity considered.

Parametric-endogeneity

In this section, let us shift our focus to the introduction and modelling of parametric-endogeneity. Let us consider a PL model such that:

$$Y_i = 1.2X_i + 0.5\left(\frac{V_i}{1+V_i^2}\right) + \epsilon_i, \text{ where} \tag{4.3}$$
$$X_i = \mathcal{Z}_i + \eta_i,\ \mathcal{Z}_i = \sin(V_i) + \varrho_i,\ \epsilon_i = \iota(\eta)_i + e_i,$$
$$\eta_i \sim \mathcal{U}(-1,1,), V_i \sim \mathcal{U}(0,3) \text{ and } e_i \sim N(0,1).$$

Clearly, the model in (4.3) implies that $E(\epsilon|x) \neq 0$, $E(\epsilon|v) = 0$, $E(\mathcal{Z}x) \neq 0$ and $E(\varrho\epsilon) = 0$. Tables 4.6 and 4.7 present the estimation results of the linear parametric-endogeneity model and the nonlinear parametric-endogeneity model respectively, based on the *2SR* procedure; Tables 4.8 and 4.9 provide those based on the *2SR–PIV* procedure.

Table 4.6. *Linear parametric-endogeneity model with 2SR*

n	$\hat{\beta}$	Bias	Var	$\lvert\hat{\beta}-\beta\rvert$	$\text{ae}_{\hat{g}}$	$\hat{h}_v$
100	1.4496	0.2496	0.0016	0.2496	0.1740	0.8056
300	1.4511	0.2511	0.0005	0.2511	0.1724	0.5561
500	1.4502	0.2502	0.0003	0.2502	0.1724	0.5343
700	1.4490	0.2490	0.0002	0.2490	0.1696	0.4432
800	1.4481	0.2481	0.0001	0.2481	0.1682	0.3982
1,100	1.4493	0.2493	0.0001	0.2493	0.1683	0.3849

Table 4.7. *Nonlinear parametric-endogeneity model with 2SR*

n	$\hat{\beta}$	Bias	Var	$\lvert\hat{\beta}-\beta\rvert$	$\text{ae}_{\hat{g}}$	$\hat{h}_v$
100	1.2832	0.0832	0.0002	0.0832	0.0661	0.2650
300	1.2822	0.0822	0.0000	0.0822	0.0613	0.3305
500	1.2812	0.0812	0.0000	0.0812	0.0588	0.2356
700	1.2808	0.0808	0.0000	0.0808	0.0571	0.1916
800	1.2800	0.0800	0.0000	0.0800	0.0560	0.1680
1,100	1.2808	0.0808	0.0000	0.0808	0.0561	0.1645

Table 4.8. *Linear parametric-endogeneity model with 2SR–PIV*

n	$\hat{\beta}$	Bias	Var	$\lvert\hat{\beta}-\beta\rvert$	$\text{ae}_{\hat{g}}$	$\hat{h}_v$
100	1.2038	0.0038	0.0041	0.0515	0.0851	0.8049
300	1.2036	0.0036	0.0010	0.0261	0.0546	0.5459
500	1.2013	0.0013	0.0007	0.0202	0.0469	0.5107
700	1.1991	-0.0009	0.0004	0.0179	0.0367	0.4227
900	1.1995	-0.0005	0.0003	0.0112	0.0316	0.3746
1,100	1.1996	-0.0004	0.0003	0.0101	0.0301	0.3293

Table 4.9. *Nonlinear parametric-endogeneity model with 2SR-PIV*

n	$\hat{\beta}$	Bias	Var	$\lvert\hat{\beta}-\beta\rvert$	$\text{ae}_{\hat{g}}$	$\hat{h}_v$
100	1.2023	0.0023	0.0004	0.0176	0.0403	0.4703
300	1.2019	0.0019	0.0001	0.0098	0.0273	0.1965
500	1.2005	0.0005	0.0001	0.0081	0.0213	0.2130
700	1.2002	0.0002	0.0000	0.0066	0.0180	0.1793
900	1.1998	-0.0002	0.0000	0.0052	0.0161	0.1625
1,100	1.2001	0.0001	0.0000	0.0050	0.0102	0.1536

Let us now discuss some important findings as follows. When compared to the estimation results of an exogenous model in Table 4.1, those in Tables 4.6 and 4.7 suggest that parametric-endogeneity, whether it belongs to the linear or nonlinear-endogenous model, can cause a severe problem in the estimation of both the parametric unknown parameter and the nonparametric structural function using the *2SR* procedure. Compared to the results shown in Tables 4.6 and 4.7, those in Tables 4.4 and 4.5 suggest that the *2SR-PIV* procedure is able to provide a much better estimation of the nonparametric structural function in the presence of parametric-endogeneity.

Bandwidth Selection

In Section 3, we apply the CV criterion to perform the adaptive data-driven approach to bandwidth selection. Furthermore, we have argued for an asymptotic optimality of such a nonparametric prediction algorithm, i.e.

$$\frac{\hat{D}(\hat{h}_v, \hat{h}_{\hat{\eta}})}{\inf_{h_v, h_\eta \in H_n} D(h_v, h_\eta)} \xrightarrow{P} 1; \tag{4.4}$$

see also (3.4) for details. To this end, a slight expansion of (4.4) leads directly to:

$$\frac{\hat{D}(\hat{h}_v, \hat{h}_{\hat{\eta}})}{\inf_{h_v, h_\eta \in H_n} D(h_v, h_\eta)} = \frac{\hat{D}(\hat{h}_v, \hat{h}_\eta)}{\inf_{h_v, h_\eta \in H_n} D(h_v, h_\eta)} + \frac{[\hat{D}(\hat{h}_v, \hat{h}_{\hat{\eta}}) - \hat{D}(\hat{h}_v, \hat{h}_\eta)]}{\inf_{h_v, h_\eta \in H_n} D(h_v, h_\eta)}, \tag{4.5}$$

which shows that the first term on the right-hand side converges to 1 in probability based on the discussion in the standard PL literature (see Härdle et al. (2000), for example). In order to establish experimental evidence in support of this asymptotic optimality, we only have to show that the second term converge to zero in probability.

In this section, we use the model examples from (4.1). For the sake of clarity, we show experimentally that:

$$\hat{D} \text{ ratio} = \frac{\hat{D}(\hat{h}_v, \hat{h}_{\hat{\eta}})}{\hat{D}(\hat{h}_v, \hat{h}_\eta)} \xrightarrow{P} 1 \tag{4.6}$$

as $n \to \infty$. Furthermore, we also compute:

$$\hat{CV} \text{ ratio} = \frac{\hat{CV}(h_v, h_{\hat{\eta}})}{\hat{CV}(h_v, h_\eta)} \tag{4.7}$$

as a complement to (4.6). The means of these are tabulated in Table 4.10, which shows that the $\hat{D}$ ratio and the $\hat{CV}$ratio tend to converge to 1 as $n \to \infty$.

Table 4.10. *Bandwidth selection for the 2SCF procedure in the presence of endogeneity*

Linear-endogeneity			Nonlinear-endogeneity		
n	$\hat{D}$ ratio	$\hat{CV}$ ratio	n	$\hat{D}$ ratio	$\hat{CV}$ ratio
100	0.8513	1.6536	100	0.8479	1.3098
300	0.8651	1.2870	300	0.9264	1.1122
500	0.8882	1.1620	500	0.9422	1.0699
700	0.9126	1.1168	700	0.9437	1.0488
900	0.9239	1.0867	900	0.9510	1.0370
1,100	0.9572	1.0475	1,100	0.9577	1.0337

5. Conclusions

In this paper, we introduce new procedures that comprehensively address endogeneity issues, i.e. parametric endogeneity and/or nonparametric endogeneity, in a partially linear semiparametric model. On the one hand, the dominance of the parametric part of the model highlights the importance of consistent estimation (and hence the estimators) of the unknown parameters. Therefore, identification of the parametric endogeneity and construction of consistent parametric estimators under such an endogeneity are essential. We thoroughly discuss these issues in Sections 2.2 and 2.3. On the other hand, nonparametric endogeneity may cause a serious problem in identifying the structural function in question. In the current paper, we established the 2SCF estimation procedure to address nonparametric endogeneity based on the two-step estimation procedure of Robinson (1988) and the CF approach by imposing the well-known triangular structure on the model. The imposition of such a structure enables us to use the marginal integration technique to identify the unknown structural function in a similar fashion to the case of a nonparametric additive model. Nonetheless, the computation of the control regressor in practice leads to a generated regressor problem which we successfully address in the current paper. Furthermore, we derive the asymptotic properties of both the parametric and nonparametric estimators involved. Among these various properties, a particular interest in the literature is the $\sqrt{n}$ consistency of the parametric estimators. For the sake of completion and practicability of the new method, we also introduce and discuss in some details the asymptotic optimality of the adaptive data-driven estimation of bandwidth selection. Finally, we conduct the Monte Carlo simulation exercises. We find strong evidence in support for the dominance of the parametric component in the model. Moreover, we find substantial evidence which indicates that our newly proposed 2SCF estimation procedure performs well and is able to overcome the endogeneity problem in the estimation of the PL model.

6. Appendix

In this Appendix, we present detailed discussion of the theoretical analysis of the main results of the current paper. We firstly list two sets of conditions which are required for the case with and without the presence of nonparametric endogeneity, respectively. The rest of this Appendix presents the mathematical proofs of Theorems 2.1 and 2.2, and a brief outline of that of Proposition 3.1.

6.A.0. Conditions for the PIV estimator

We state a set of conditions for the PIV and P2SLS estimations when the nonparametric regressors are exogenous. In particular, the conditions on the parametric instrumental variables are the moment condition on $\mathcal{Z}$, the smoothness condition on the function $m_{\mathcal{Z}}(v)$ and the regularity conditions of the bandwidth parameter. It is useful to compare this set of conditions with those in the next section, where we address nonparametric-endogeneity.

Assumption A.0.1. *$(V_i, X_i, Y_i, \mathcal{Z}_i)$, $i = 1, 2, \ldots, n$ are i.i.d. observations.*

Assumption A.0.2. *$E(\epsilon|x, v) \neq 0$ and $E(\epsilon|\mathcal{Z}, v) = 0$.*

Assumption A.0.3. *$E(\epsilon^2|v, \mathcal{Z}) = \sigma^2(v, \mathcal{Z})$ is continuous in $(v, \mathcal{Z})$.*

Assumption A.0.4. *All Y_i have a finite second moment, and all X_i and $\mathcal{Z}_i$ have finite fourth moments.*

Assumption A.0.5. *V_i admits a density function $f \in \mathcal{G}_{\lambda}^{\infty}$ for some $\lambda > 0$.*

Assumption A.0.6.

(1) $m_x(v) \in \mathcal{G}_{\mu_0}^4$ for some $\mu_0 > 0$;

(2) $m_{\mathcal{Z}}(v) \in \mathcal{G}_{\upsilon_0}^4$ for some $\upsilon_0 > 0$;

(3) $g(\cdot) \in \mathcal{G}_{\nu_0}^4$ for some $\nu_0 > 0$.

Assumption A.0.7. *As $n \to \infty$,*

(1) $nh_v^{2q} b^4 \to \infty$;

(2) $nh_v^{2\min(\lambda,\mu_0)+2\min(\lambda+1,\upsilon_0)} b^{-4} \to 0$;

(3) $nh_v^{\min(\lambda,\upsilon_0)} b^{-4} \to 0$;

(4) $h_v^{\min(\lambda,2\lambda,\upsilon_0,\nu_0)} b^{-2} \to 0$;

(5) $b \to 0$.

Assumption A.0.8. $\sup_{v\in\mathbb{R}^q} |K(v)| + \int |v^p K(v)| dv < \infty$ *and* $\int v^{p-i} K(v) dv = 0$ *for $i = 1, \ldots, p-1$, where $p = \max(\lambda + \mu_0, \lambda + \upsilon_0, \lambda + \nu_0)$.*

Assumption A.0.9. *$E(\epsilon|v, x) = 0$ and $E(\epsilon^2|v, x)$ is continuous in (v, x).*

Assumption A.0.10. *As* $n \to \infty$,

(1) $nh_v^{2q}b^4 \to \infty$;

(2) $nh_v^{2\min(\lambda,\mu_0)+2\min(\lambda+1,\nu_0)}b^{-4} \to 0$;

(3) $h_v^{\min(\lambda,2\lambda,\mu_0,\nu_0)}b^{-2} \to 0$;

(4) $b \to 0$.

Assumption A.0.2 indicates that the model suffers from parametric endogeneity. If this is not the case then we consider Assumption A.0.9 instead. Furthermore, if endogeneity is present, then we impose the condition in Assumption A.0.7 on the bandwidth parameter; otherwise, Assumption A.0.10 is used. Assumption A.0.4 provides the moment conditions on the regressors. Assumptions A.0.5 and A.0.6 collectively provide the moment bounds and the smoothness of the density and regression functions. By Assumption A.0.8, the kernel function is bounded, integrable and high-order. Assumptions A.0.7 and A.0.8 should be satisfied simultaneously (see Robinson (1988), for example) in the case of parametric endogeneity, while Assumptions A.0.8 and A.0.10 are used instead for the case where there is no parametric endogeneity.

6.A.1. Conditions for Theorems 2.1 and 2.2

Assumption A.1.1.

$(V_i, X_i, Y_i, Z_i, \mathcal{Z}_i)$ *where* $i = 1, \ldots, n$ *are i.i.d. observations.*

Assumption A.1.2.

$E(\epsilon|x, v) \neq 0$, $E(\epsilon|x, \eta) \neq 0$, $E(\epsilon|x, z) = 0$ *and* $E(e^2) = \sigma^2(x, v, \eta) < \infty$.

Assumption A.1.3.

All X_i, Y_i *and* $\mathcal{Z}_i$ *have finite eight moments.*

Assumption A.1.4.

(1) $f(v, \eta) \in \mathcal{G}_{\lambda_2}^{\infty}$ *for some* $\lambda_2 = r + p_2 \geq 0$;

(2) $f(z) \in \mathcal{G}_{\lambda_1}^{\infty}$ *for some* $\lambda_1 = r + p_1 \geq 0$.

Assumption A.1.5.

(1) $m_x(v, \eta) \in \mathcal{G}_{\mu}^{8}$ *for some* $\mu > 0$;

(2) $m_v(z) \in \mathcal{G}_{\nu_1}^{8}$ *for some* $\nu_1 > 0$;

(3) $m(v, \eta) \in \mathcal{G}_{\nu_2}^{8}$ *for some* $\nu_2 > 0$.

Assumption A.1.6. *As* $n \to \infty$,

(1) $n^5 h_v^{6q} h_\eta^{6q+4} h_z^{q_z} b_1^4 b_2^8 \to \infty$;

(2) $n^3 h_v^{2q} h_\eta^{2q+4} h_z^{3q_z} b_1^4 b_2^4 \to \infty$;

(3) $n^{1/2} h_\eta^{-2} h_z^{2\min(\lambda_1,\nu_1)} b_1^{-2} b_2^{-2} \to 0$;

(4) $n^{-1} h_v^{-3q} h_\eta^{-3q-2} h_z^{2\min(\lambda_1,\nu_1)} b_1^{-2} b_2^{-4} \to 0$;

(5) $n^{-3}h_v^{4\min(\lambda_2,\mu,\nu_2)}h_\eta^{4\min(\lambda_2,\mu,\nu_2)-4}h_z^{-3q_z}b_1^{-4}b_2^{-8} \to 0;$

(6) $h_v^{\min(\lambda_2,\mu,\nu_2)}h_\eta^{\min(\lambda_2,\mu,\nu_2)}h_z^{\min(\lambda_1,\nu_1)}b_1^{-1}b_2^{-2} \to 0;$

(7) $b_1 \to 0$ *and* $b_2 \to 0$.

Assumption A.1.7.

(1.1) $\sup_{v,\eta}|L^{(r)}(v,\eta)| + \int |v^{p_2}\eta^{p_2}L^{(r)}(v,\eta)|dvd\eta < \infty$ *and* $\int v^{p_2-i}\eta^{p_2-i}L^{(r)}(v,\eta)dvd\eta = 0;$

(1.2) $\sup_z|K_z(z)| + \int |z^{p_1}K_z(z)|dz < \infty$ *and* $\int z^{p_1-i}K(z)dz = 0,$
where $i = 1, 2, \ldots, p_l - 1$, l *is* 1 *or* 2, $p_2 = \max(\lambda_2 + \mu, \lambda_2 + \nu_2)$, *and* $p_1 = (\lambda_1 + \nu_1)$.

Assumption A.1.8.

(1) $E(\epsilon|x, v) \neq 0$, $E(\epsilon|x, \eta) \neq 0$, $E(\epsilon|x, z) \neq 0$, *and* $E(\epsilon|\mathcal{Z}, z) = 0$ *and* $E(e^2) = \sigma^2(\mathcal{Z}, v, \eta) < \infty;$

(2) $m_{\mathcal{Z}}(v, \eta) \in \mathcal{G}_\upsilon^8$ *for some* $\upsilon > 0$.

Assumption A.1.9. *As* $n \to \infty$,

(1) $n^5h_v^{6q}h_\eta^{6q+4}h_z^{q_z}b_1^4b_2^8 \to \infty;$

(2) $n^3h_v^{2q}h_\eta^{2q+4}h_z^{3q_z}b_1^4b_2^4 \to \infty;$

(3) $n^{1/2}h_\eta^{-2}h_z^{2\min(\lambda_1,\nu_1)}b_1^{-2}b_2^{-2} \to 0;$

(4) $n^{-1}h_v^{-3q}h_\eta^{-3q-2}h_z^{2\min(\lambda_1,\nu_1)}b_1^{-2}b_2^{-4} \to 0;$

(5) $n^{-3}h_v^{4\min(\lambda_2,\upsilon,\nu_2)}h_\eta^{4\min(\lambda_2,\upsilon,\nu_2)-4}h_z^{-3q_z}b_1^{-4}b_2^{-8} \to 0;$

(6) $n^{-3}h_v^{4\min(\lambda_2,\upsilon,\mu)}h_\eta^{4\min(\lambda_2,\upsilon,\mu)-4}h_z^{-3q_z}b_1^{-4}b_2^{-8} \to 0;$

(7) $h_v^{\min(\lambda_2,\mu,\nu_2,\upsilon)}h_\eta^{\min(\lambda_2,\mu,\nu_2,\upsilon)}h_z^{\min(\lambda_1,\nu_1)}b_1^{-1}b_2^{-2} \to 0;$

(8) $b_1 \to 0$ *and* $b_2 \to 0$.

Assumption A.1.2 indicates the presence of nonparametric endogeneity in the model. Furthermore, note that the moment conditions on Y and X are more restrictive than those in Robinson (1988) since the estimation procedure involves a two-step nonparametric estimation procedure in order to address nonparametric endogeneity, i.e. compare Assumption A.1.3 with Assumption A.0.4 in Section 6.A.0. Assumptions A.1.4 and A.1.5 state the smoothness and moment properties of the density and regression functions, and these are also more restrictive than Robinson (1988) ones, i.e. compare these with Assumptions A.0.5 and A.0.6 in Section 6.A.0. Given the higher-order kernel function in Assumption A.1.7, the bias is sufficiently decreased with Assumptions A.1.4 and A.1.5. Assumption A.1.7 states that the kernel functions used in this paper are bounded, integrable and high-order.

Note that Assumptions A.1.6 and A.1.7 should be satisfied simultaneously. For example, if the order of $L^{(r)}$ and K_z is greater than 3 (i.e., $p_l \geq 3$ where $l = 1$ or 2) then the lower bounds on the rates of decay of h_v, h_η and h_z are no better than $nh_z^{12} \to 0$, $nh_v^6h_\eta^6 \to 0$, and $h_z^{12}h_v^{12-3q}h_\eta^{12-3q}b_1^{-4}b_2^{-8} \to 0$, no matter which degree of smoothness prevails. A necessary condition for reconciling the components of Assumption A.1.6 is the following:

$$2/16q_z < \lambda_1, 2/16 < \nu_1, 6/16q < \lambda_2, 6/8q < (\lambda_2 + \nu_2), 6/8q < (\lambda_2 + \mu) \text{ and } 6/8q < (\nu_2 + \mu).$$

Assumptions A.1.8 and A.1.9 are for the case of the presence of both parametric endogeneity and nonparametric endogeneity in the model. In particular, Assumption A.1.8 (1) states that the model suffers from parametric endogeneity as well. Assumption A.1.8 (2) states the smoothness of the function $m_{\mathcal{Z}}(v,\eta)$ that sufficiently reduces the bias with Assumption A.1.6.

6.A.2. Proof of Theorem 2.1

By using the notation in Robinson (1988), we rewrite the linear reduced form including the bias term, as follows:

$$Y_i - \hat{Y}_{2,i} = (X_i - \hat{X}_{2,i})'\beta + (m_i - \hat{m}_{2,i}) + (e_i - \hat{e}_{2,i}), \tag{A.2.1}$$

where $m_i = g(V_i) + \iota(\eta_i)$ and $e_i = \epsilon_i - \iota(\eta_i)$. By incorporating the fact that the endogeneity control regressor is generated, (A.2.1) is rewritten below:

$$\begin{aligned} Y_i - \hat{Y}_{1,i} - (\hat{Y}_{2,i} - \hat{Y}_{1,i}) &= \{X_i - \hat{X}_{1,i} - (\hat{X}_{2,i} - \hat{X}_{1,i})\}'\beta \\ &+ \{m_i - \hat{m}_{1,i} - (\hat{m}_{2,i} - \hat{m}_{1,i})\} + \{e_i - \hat{e}_{1,i} - (\hat{e}_{2,i} - \hat{e}_{1,i})\} \\ Y_i - \hat{Y}_{1,i} - \delta_{y,i} &= (X_i - \hat{X}_{1,i} - \delta_{x,i})'\beta + (m_i - \hat{m}_{1,i} - \delta_{m,i}) + (e_i - \hat{e}_{1,i} - \delta_{e,i}), \end{aligned} \tag{A.2.2}$$

where $\hat{\delta}_i = \hat{\delta}_{2,i} - \hat{\delta}_{1,i}$, $\hat{\delta}_{1,i} = \frac{\sum_{j=1}^n \delta_j K_v\left(\frac{V_i - V_j}{h_v}\right) K_\eta\left(\frac{\eta_i - \eta_j}{h_\eta}\right)}{\sum_{l=1}^n K_v\left(\frac{V_i - V_l}{h_v}\right) K_\eta\left(\frac{\eta_i - \eta_l}{h_\eta}\right)}$ and $\hat{\delta}_{2,i} = \frac{\sum_{j=1}^n \delta_j K_v\left(\frac{V_i - V_j}{h_v}\right) K_\eta\left(\frac{\hat{\eta}_i - \hat{\eta}_j}{h_\eta}\right)}{\sum_{l=1}^n K_v\left(\frac{V_i - V_l}{h_v}\right) K_\eta\left(\frac{\hat{\eta}_i - \hat{\eta}_l}{h_\eta}\right)}$, and $\hat{\delta}_i$ is used to denote for $\delta_{y,i}$, $\delta_{x,i}$, $\delta_{m,i}$ and $\delta_{e,i}$, here. Using (A.2.1) and (A.2.2), we have:

$$\begin{aligned} \hat{\beta} - \beta &= S^{-1}_{X-\hat{X}_2}\left(S_{X-\hat{X}_2, m-\hat{m}_2} + S_{X-\hat{X}_2, e-\hat{e}_2}\right) \\ \hat{\sigma}^2 - \sigma^2 &= \left(S_{e-\hat{e}_2} - \sigma^2\right) + S_{m-\hat{m}_2} + \left(\hat{\beta} - \beta\right)' S_{X-\hat{X}_2}\left(\hat{\beta} - \beta\right) \\ &+ 2S_{m-\hat{m}_2, e-\hat{e}_2} - 2\left(\hat{\beta} - \beta\right) S_{X-\hat{X}_2, e-\hat{e}_2} - 2\left(\hat{\beta} - \beta\right) S_{X-\hat{X}_2, m-\hat{m}_2}, \end{aligned}$$

where

$$\begin{aligned} S_{X-\hat{X}_2} &= S_{m_x-\hat{m}_x} + S_{m_x-\hat{m}_x,U} - S_{m_x-\hat{m}_x,\hat{U}} - S_{m_x-\hat{m}_x,\delta_x} + S_{U,m_x-\hat{m}_x} + S_U - S_{U\hat{U}} - S_{U\delta_x} \\ &- S_{\hat{U},m_x-\hat{m}_x} - S_{\hat{U}U} + S_{\hat{U}} + S_{\hat{U}\delta_x} - S_{\delta_x,m_x-\hat{m}_x} - S_{\delta_x U} + S_{\delta_x\hat{U}} + S_{\delta_x} \\ S_{X-\hat{X}_2,m-\hat{m}_2} &= S_{m_x-\hat{m}_x,m-\hat{m}} - S_{m_x-\hat{m}_x,\delta_m} + S_{U,m-\hat{m}} - S_{U\delta_m} - S_{\hat{U},m-\hat{m}} + S_{\hat{U}\delta_m} - S_{\delta_x,m-\hat{m}} + S_{\delta_x\delta_m} \\ S_{X-\hat{X}_2,e-\hat{e}_2} &= S_{m_x-\hat{m}_x,e} - S_{m_x-\hat{m}_x,\hat{e}} - S_{m_x-\hat{m}_x,\delta_e} + S_{Ue} - S_{U\hat{e}} - S_{U\delta_e} - S_{\hat{U}e} + S_{\hat{U}\hat{e}} + S_{\hat{U}\delta_e} - S_{\delta_x e} \\ &+ S_{\delta_x\hat{e}} + S_{\delta_x\delta_e} \\ S_{m-\hat{m}_2,e-\hat{e}_2} &= S_{m-\hat{m},e} - S_{m-\hat{m},\hat{e}} - S_{m-\hat{m},\delta_e} - S_{\delta_m e} + S_{\delta_m\hat{e}} + S_{\delta_m\delta_e} \\ S_{e-\hat{e}_2} &= S_e - S_{e\hat{e}} - S_{e\delta_e} - S_{\hat{e}e} + S_{\hat{e}} + S_{\hat{e}\delta_e} - S_{\delta_e e} + S_{\delta_e\hat{e}} + S_{\delta_e} \\ S_{m-\hat{m}_2} &= S_{m-\hat{m}} - S_{m-\hat{m},\delta_m} - S_{\delta_m,m-\hat{m}} + S_{\delta_m}. \end{aligned}$$

These decompositions enable us to see the bias from the first step of the estimation procedure to generate the endogeneity control regressors. We show that $\hat{\beta}$ is still $\sqrt{n}$-consistent with these additional bias terms especially in Propositions A.2.2 to A.2.6. The proof is completed by applying Propositions A.2.1 to A.2.6 below, which imply, via the Cauchy inequality, that $S_{m_x-\hat{m}_x,U}$, $S_{m_x-\hat{m}_x,\hat{U}}$, $S_{U\hat{U}}$, $S_{m_x-\hat{m}_x,\delta_x}$, $S_{U\delta_x}$, $S_{\hat{U}\delta_x}$, $S_{m-\hat{m},e}$, $S_{m-\hat{m},\hat{e}}$, $S_{\delta_m e}$, $S_{\delta_m\hat{e}}$, $S_{\delta_m\delta_e}$, $S_{e\hat{e}}$, $S_{e\delta_e}$, $S_{\hat{e}\delta_e}$, and $S_{m-\hat{m},\delta_m}$ all $\xrightarrow{p} 0$. We use the notation in Robinson (1988), where $E_i(\cdot) = E(\cdot|V_i, Z_i)$, $\varsigma = (\lambda_2, \mu)$, $\xi_1 = \min(\lambda_1, \nu_1)$, $\xi_2 = \min(\lambda_2, \nu_2)$ and $\mathcal{C}$ denotes a generic constant.

Proposition A.2.1.

(1) $E|S_{m_x-\hat{m}_x}| = O(n^{-1}h_v^{-q}h_\eta^{-q}b_2^{-2} + h_v^{2\varsigma}h_\eta^{2\varsigma}b_2^{-2})$;

(2) $E|S_{m-\hat{m}}| = O(n^{-1}h_v^{-q}h_\eta^{-q}b_2^{-2} + h_v^{2\xi_2}h_\eta^{2\xi_2}b_2^{-2})$;

(3) $\sqrt{n}S_{m_x-\hat{m}_x,m-\hat{m}} = O_p(n^{-1/2}h_v^{-q}h_\eta^{-q}b_2^{-2} + n^{1/2}h_v^{\varsigma+\xi_2}h_\eta^{\varsigma+\xi_2}b_2^{-2})$;

(4) $S_U = \Phi_U + O_p(n^{-1/2}h_v^{-q/2}h_\eta^{-q/2}b_2^{-1} + h_v^{\lambda_2}h_\eta^{\lambda_2}b_2^{-1}) + o_p(1)$;

(5) $S_{\hat{U}} = O_p(n^{-1}h_v^{-q}h_\eta^{-q}b_2^{-2})$;

(6) $\sqrt{n}S_{U,m-\hat{m}} = O_p(n^{-1/2}h_v^{-q/2}h_\eta^{-q/2}b_2^{-1} + h_v^{\xi_2}h_\eta^{\xi_2}b_2^{-1})$;

(7) $\sqrt{n}S_{\hat{U},m-\hat{m}} = O_p(n^{-1/2}h_v^{-q/2}h_\eta^{-q/2}b_2^{-2} + h_v^{\xi_2}h_\eta^{\xi_2}b_2^{-2})$;

(8) $\sqrt{n}S_{m_x-\hat{m}_x,e} = O_p(n^{-1/2}h_v^{-q/2}h_\eta^{-q/2}b_2^{-1} + h_v^{\varsigma}h_\eta^{\varsigma}b_2^{-1})$;

(9) $\sqrt{n}S_{m_x-\hat{m}_x,\hat{e}} = O_p(n^{-1/2}h_v^{-q/2}h_\eta^{-q/2}b_2^{-2} + h_v^{\varsigma}h_\eta^{\varsigma}b_2^{-2})$;

(10) $\sqrt{n}S_{\hat{U}e} = O_p(n^{-1/2}h_v^{-q/2}h_\eta^{-q/2}b_2^{-1})$;

(11) $\sqrt{n}S_{U\hat{e}} = O_p(n^{-1/2}h_v^{-q/2}h_\eta^{-q/2}b_2^{-1})$;

(12) $\sqrt{n}S_{\hat{U}\hat{e}} = O_p(n^{-1/2}h_v^{-q/2}h_\eta^{-q/2}b_2^{-2})$;

(13) $S_{\hat{e}} = O_p(n^{-1}h_v^{-q}h_\eta^{-q}b_2^{-2})$;

(14) $S_e = \sigma^2 + o_p(1)$;

(15) $S_{eU} \to_D N(0, \sigma^2\Phi_U)$.

Proof: The proofs of Proposition A.2.1 (1) - (15) can be easily obtained by a simple extension of Robinson (1988). ■

Proposition A.2.2.

(1) $E|S_{\delta_x}| = O\left(n^{-2}h_v^{-q}h_\eta^{-q-2}h_z^{-3q_z/2}b_1^{-2}b_2^{-2} + h_\eta^{-2}h_z^{2\xi_1}b_1^{-2}b_2^{-2}\right)$;

(2) $E|S_{\delta_m}| = O\left(n^{-2}h_v^{-q}h_\eta^{-q-2}h_z^{-3q_z/2}b_1^{-2}b_2^{-2} + h_\eta^{-2}h_z^{2\xi_1}b_1^{-2}b_2^{-2}\right)$;

(3) $E|S_{\delta_e}| = O\left(n^{-2}h_v^{-q}h_\eta^{-q-2}h_z^{-3q_z/2}b_1^{-2}b_2^{-2} + h_\eta^{-2}h_z^{2\xi_1}b_1^{-2}b_2^{-2}\right)$.

Proof: Let us denote $\hat{\delta}$ as δ_x, δ_m and δ_e in the rest of the paper. Now, we have:

$$\hat{\delta}_i = \hat{\delta}_{2,i} - \hat{\delta}_{1,i} = \frac{1}{nh_v^q h_\eta^q}\sum_{j=1}^n \delta_j(\hat{w}_{ij} - w_{ij}), \tag{A.2.3}$$

where $w_{ij} = \frac{K_v\left(\frac{V_i-V_j}{h_v}\right)K_\eta\left(\frac{\eta_i-\eta_j}{h_\eta}\right)}{\frac{1}{nh_v^q h_\eta^q}\sum_{l=1}^n K_v\left(\frac{V_i-V_l}{h_v}\right)K_\eta\left(\frac{\eta_i-\eta_l}{h_\eta}\right)}$ and $\hat{w}_{ij} = \frac{K_v\left(\frac{V_i-V_j}{h_v}\right)K_\eta\left(\frac{\hat{\eta}_i-\hat{\eta}_j}{h_\eta}\right)}{\frac{1}{nh_v^q h_\eta^q}\sum_{l=1}^n K_v\left(\frac{V_i-V_l}{h_v}\right)K_\eta\left(\frac{\hat{\eta}_i-\hat{\eta}_l}{h_\eta}\right)}$. By the Taylor series expansion of the kernel function,

$$K_\eta\left(\frac{\hat{\eta}_i-\hat{\eta}_j}{h_\eta}\right) = K_\eta\left(\frac{\eta_i-\eta_j}{h_\eta}\right) + \sum_{r=1}^{\omega_2-1}\frac{1}{r!}K_\eta^{(r)}\left(\frac{\eta_i-\eta_j}{h_\eta}\right)\left(\frac{\triangle_{ij}}{h_\eta}\right)^{\omega_2} + R_{ij},$$

where $\triangle_{ij} = \{m_v(Z_i) - \hat{m}_v(Z_i)\} - \{m_v(Z_j) - \hat{m}_v(Z_j)\}$, $R_{ij} = \frac{1}{\omega_2!} K_\eta^{(\omega_2)} \left(\frac{\tilde{\eta}_i - \tilde{\eta}_j}{h_\eta}\right) \left(\frac{\triangle_{ij}}{h_\eta}\right)^{\omega_2}$ which is a remainder term, and $\tilde{\eta}_i - \tilde{\eta}_j$ is between the segment line of $\eta_i - \eta_j$ and $\hat{\eta}_i - \hat{\eta}_j$. Hence, $\hat{w}_{ij}$ is:

$$\frac{1}{nh_v^q h_\eta^q} \sum_{l=1}^n K_v \left(\frac{V_i - V_l}{h_v}\right) K_\eta \left(\frac{\hat{\eta}_i - \hat{\eta}_l}{h_\eta}\right) = A_{0,i} + \sum_{r=1}^{\omega_2 - 1} \frac{1}{r!} A_{r,i} \triangle_{il}^r + R_{il}, \tag{A.2.4}$$

where:

$$\begin{aligned} A_{0,i} &= \frac{1}{nh_v^q h_\eta^q} \sum_{l=1}^n K_v \left(\frac{V_i - V_l}{h_v}\right) K_\eta \left(\frac{\eta_i - \eta_l}{h_\eta}\right) = \hat{f}(V_i, \eta_i) \\ A_{r,i} &= \frac{1}{nh_v^q h_\eta^{q+r}} \sum_{l=1}^n K_v \left(\frac{V_i - V_l}{h_v}\right) K_\eta^{(r)} \left(\frac{\eta_i - \eta_l}{h_\eta}\right) = \hat{f}_\eta^{(r)}(V_i, \eta_i), \end{aligned}$$

where $\hat{f}_\eta^{(r)}(v, \eta)$ is the rth partial derivative of the joint density function of (v, η) with respect to η. The main dominating terms in (A.2.4) are:

$$A_{1,1,i} = \{m_v(Z_i) - \hat{m}_v(Z_i)\} \hat{f}_\eta^{(1)}(V_i, \eta_i) \tag{A.2.5}$$

$$A_{1,2,i} = \frac{1}{nh_v^q h_\eta^{q+1}} \sum_{l=1}^n K_v \left(\frac{V_i - V_l}{h_v}\right) K_\eta^{(1)} \left(\frac{\eta_i - \eta_l}{h_\eta}\right) \{m_v(Z_l) - \hat{m}_v(Z_l)\}. \tag{A.2.6}$$

We firstly consider (A.2.5) by the bound condition on the kernel function and the smoothness on the function $m_v(z)$:

$$E\left(\{m_v(Z_i) - \hat{m}_v(Z_i)\}\right)^2 \le (nh_z^{q_z} b_1)^{-2} E(T_z)^2 = O(n^{-1} h_z^{-q_z} b_1^{-2} + h_z^{2\xi_1} b_1^{-2}), \tag{A.2.7}$$

where $T_z = \sum_i t_i$ with $t_i = (m_{v,1} - m_{v,i}) K_{z,1i}$ and $E(T_z)^2 \le \mathcal{C}\left(E\left(\sum_{i=1} t_i - t\right)^2 + n^2 E(t^2)\right) = O(nh_z^{q_z} + n^2 h_z^{2\xi_1})$ with $t = E_1(t_i)$ and $t_i - t$ are independent with a mean of 0, and by the bound condition on the kernel function:

$$E\left(\hat{f}_\eta^{(1)}(V_i, \eta_i)\right)^2 = (nh_v^q h_\eta^{q+1})^{-2} E\left(\sum_i L_{1i}^{(1)}\right)^2 = O(n^{-1} h_v^{-q} h_\eta^{-q-2}).$$

By the Cauchy inequality:

$$A_{1,1,i} = O_p\left(n^{-1} h_v^{-q/2} h_\eta^{-q/2-1} h_z^{-q_z/2} b_1^{-1} + n^{-1/2} h_v^{-q/2} h_\eta^{-q/2-1} h_z^{\xi_1} b_1^{-1}\right). \tag{A.2.8}$$

By the *i.i.d.* assumption, we have the second moment bound of (A.2.6) as follows:

$$E(A_{1,2,i})^2 \le \left(n^2 h_v^q h_\eta^{q+1} h_z^{q_z} b_1^1\right)^{-2} \left[E\left\{\sum_{l=1}^n \left(L_{1l}^{(1)}\right)^2 T_z^2\right\} + E\left|\sum_{l=1}^n \sum_{j \ne l}^n \left(L_{1l}^{(1)}\right)\left(L_{1j}^{(1)}\right) T_z^2\right|\right]. \tag{A.2.9}$$

The first term in (A.2.9) is:

$$E\left\{\sum_{l=1}^n \left(L_{1l}^{(1)}\right)^2 T_z^2\right\} \le \mathcal{C} E\left\{\left(L^{(1)}(0)\right)^2 + n\left(L_{1l}^{(1)}\right)^2 t_{z,2}^{(2)} + n\left(L_{1l}^{(1)}\right)^2 T_{z,2}^2\right\},$$

where $T_{z,2} = T_z - t_{z,2}$ and $t_{z,2} = (m_{v,1} - m_{v,2}) K_{z,12}$, and by bound condition on the kernel function and the bounded moment condition on the $m_v(z)$ function:

$$E\left\{\left(L_{1l}^{(1)}\right)^2 t_{z,2}^2\right\} \leq \left[E\left\{E_l\left(L_{1l}^{(1)}\right)^4\right\}E(t_{z,2}^4)\right]^{1/2} = O(h_v^q h_\eta^q h_z^{q_z})^{1/2} \tag{A.2.10}$$

$$\begin{aligned} E\left\{\left(L_{1l}^{(1)}\right)^2 T_{z,2}^2\right\} &\leq \left[E\left\{E_l\left(L_{1l}^{(1)}\right)^2\right\}E\left\{E_l\left(L_{1l}^{(1)}\right)^2 T_{z,2}^4\right\}\right]^{1/2} \\ &= O\left(n^{1/2}h_v^q h_\eta^q h_z^{q_z/2} + n^2 h_v^q h_\eta^q h_z^{2(q_z+\xi_1)}\right), \end{aligned} \tag{A.2.11}$$

where $E(T_z)^4 = O\left(nh_z^{q_z} + n^4 h_z^{4(q_z+\xi_1)}\right)$ by the similar argument as in (A.2.7). Hence the first term on the right-hand side in (A.2.9) is $O(n^{-5/2}h_v^{-q}h_\eta^{-q-2}h_z^{-3q_z/2}b_1^{-2} + n^{-1}h_v^{-q}h_\eta^{-q-2}h_z^{2\xi_1}b_1^{-2})$. The second term on the right-hand side in (A.2.9) is bounded by:

$$\leq \mathcal{C}(n^{-3}h_v^{-2q}h_\eta^{-2(q+1)}h_z^{-2q_z}b_1^{-2})E\left\{\left(L_{1l}^{(1)}\right)^2 T_{z,2}^2 + n\left(L_{1l}^{(1)}L_{1j}^{(1)}\right)\left(t_{z,2}^2 + t_{z,3}^2 + T_{z,3}^2\right)\right\},$$

where $T_{z,3} = T_{z,2} - t_{z,3}$, and

$$E\left(\left(L_{1l}^{(1)}\right)^2 T_{z,2}^2\right) = O\left(nh_v^q h_\eta^q h_z^{q_z} + n^2 h_v^q h_\eta^q h_z^{2(q_z+\xi_1)}\right)$$

$$E\left(\left|L_{1l}^{(1)}L_{1j}^{(1)}\right| t_{z,\imath}^2\right) \leq \left[E\left\{E_l\left(L_{1l}^{(1)}\right)^2\right\}E\left\{E_j\left(L_{1j}^{(1)}\right)^2 t_{z,\imath}^4\right\}\right]^{1/2} = O(h_v^q h_\eta^q h_z^{q_z/2}) \tag{A.2.12}$$

$$\begin{aligned} E\left(\left|L_{1l}^{(1)}L_{1j}^{(1)}\right| T_{z,2}^2\right) &\leq \left[E\left\{E_l\left(L_{1l}^{(1)}\right)E_j\left(L_{1j}^{(1)}\right)\right\}E\left\{E_l\left(L_{1l}^{(1)}\right)E_j\left(L_{1j}^{(1)}\right)T_{z,2}^4\right\}\right]^{1/2} \\ &= O\left(n^{1/2}h_v^{2q}h_\eta^{2q}h_z^{q_z/2} + n^2 h_v^{2q}h_\eta^{2q}h_z^{2(q_z+\xi_1)}\right), \end{aligned} \tag{A.2.13}$$

where $\imath = 2$ or 3. The second term in (A.2.9) is $O\left(n^{-2}h_v^{-q}h_\eta^{-q-2}h_z^{-3q_z/2}b_1^{-2} + h_\eta^{-2}h_z^{2\xi_1}b_1^{-2}\right)$. Hence, we have:

$$A_{1,2,i} = O_p\left(n^{-1}h_v^{-q/2}h_\eta^{-q/2-1}h_z^{-q_z/4}b_1^{-1} + h_\eta^{-1}h_z^{\xi_1}b_1^{-1}\right). \tag{A.2.14}$$

We obtain the following results:

$$\begin{aligned} \hat{w}_{ij} - w_{ij} &= \left(\hat{f}(v,\eta) + o_p(1)\right)^{-1} K_v\left(\frac{V_i - V_j}{h_v}\right)\left\{K_\eta\left(\frac{\hat{\eta}_i - \hat{\eta}_j}{h_\eta}\right) - K_\eta\left(\frac{\eta_i - \eta_j}{h_\eta}\right)\right\} \\ &= \left(\hat{f}(v,\eta) + o_p(1)\right)^{-1} K_v\left(\frac{V_i - V_j}{h_v}\right)\left\{\frac{1}{h_\eta}K_\eta^{(1)}\left(\frac{\eta_i - \eta_j}{h_\eta}\right)\triangle_{ij} + \sum_{r'=2}^{\omega_2 - 1}\frac{1}{r'!}K_\eta^{(r')}\left(\frac{\eta_i - \eta_j}{h_\eta}\right)\left(\frac{\triangle_{ij}}{h_\eta}\right)^{r'} + R_{ij}\right\} \end{aligned}$$

and:

$$\hat{\delta}_i = \frac{\left\{B_{1,i}\triangle_{ij} + \sum_{r'=2}^{\omega_2-1}\frac{1}{r'!}B_{r',i}\triangle_{ij}^{r'} + R_{ij}\right\}}{\hat{f}(V_i,\eta_i)}, \tag{A.2.15}$$

where

$$B_{1,i} = \frac{1}{nh_v^q h_\eta^{q+1}}\sum_{j=1}^n \delta_j K_v\left(\frac{V_i - V_j}{h_v}\right)K_\eta^{(1)}\left(\frac{\eta_i - \eta_j}{h_\eta}\right) \text{ and } B_{r',i} = \frac{1}{nh_v^q h_\eta^{q+r'}}\sum_{j=1}^n \delta_j K_v\left(\frac{V_i - V_j}{h_v}\right)K_\eta^{(r')}\left(\frac{\eta_i - \eta_j}{h_\eta}\right)$$

by (A.2.8) and (A.2.14), and the Taylor series expansion of the kernel function. Hence, we have:

$$\begin{aligned} \hat{\delta}_i &= \frac{\frac{1}{nh_v^q h_\eta^{q+1}}\sum_{j=1}^n \delta_j K_v\left(\frac{V_i - V_j}{h_v}\right)K_\eta^{(1)}\left(\frac{\eta_i - \eta_j}{h_\eta}\right)\{m_v(Z_j) - \hat{m}_v(Z_j)\}}{\hat{f}(V_i,\eta_i)} \\ &+ O_p\left(n^{-1}h_v^{-q/2}h_\eta^{-q/2-1}h_z^{-q_z/2}b_1^{-1} + n^{-1/2}h_v^{-q/2}h_\eta^{-q/2-1}h_z^{\xi_1}b_1^{-1}\right) + o_p(1), \end{aligned} \tag{A.2.16}$$

since $E\left(m_v(z) - \hat{m}_v(z)\right)^2 = O(n^{-1}h_z^{-q_z}b_1^{-2} + h_z^{2\xi_1}b_1^{-2})$ and $E\left(B_{1,i}\right)^2 = O(n^{-1}h_v^{-q}h_\eta^{-q-2})$ by the bound condition on the kernel function.

Using (A.2.16),

$$E|S_{\hat\delta}| \;\le\; (nh_z^{q_z})^{-2}E\left\{\frac{1}{n}\sum_{i=1}^{n}|\breve\delta_i|^2 I_{2,i}T_z^2 I_1\right\} \tag{A.2.17}$$

$$+\;(nh_z^{q_z})^{-2}\left|E\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\breve\delta_i\breve\delta_j' I_{2,i}I_{2,j}'T_z^2 I_1\right\}\right|, \tag{A.2.18}$$

where $\breve\delta_i = \hat f(V_i,\eta_i)^{-1}\frac{1}{nh_v^q h_\eta^{q+1}}\sum_{j=1}^{n}\delta_j L_{ij}^{(1)}$. Because $E(|\breve\delta_1|^2 I_{2,1}|\mathcal{L}_n) \le (nh_v^q h_\eta^{q+1}b_2)^{-2}E\left(\sum_{i=1}^{n}|\delta_i|^2\left(L_{1i}^{(1)}\right)^2|\mathcal{L}_n\right)$, a.s., the right hand side of (A.2.17) is bounded by $(n^2h_v^q h_\eta^{q+1}h_z^{q_z}b_1b_2)^{-2}$ multiplies by:

$$E\left(\sum_{i=1}^{n}|\delta_i|^2\left(L_{1i}^{(1)}\right)^2T_z^2\right) \le \mathcal{C}E\left(|\delta_1|^2T_z^2 + n|\delta_2|^2\left(L_{12}^{(1)}\right)^2t_{z,2}^2 + n|\delta_2|^2\left(L_{12}^{(1)}\right)^2T_{z,1}^2\right), \tag{A.2.19}$$

where $\mathcal{L}_n = (V_1\times\eta_1,\ldots,V_n\times\eta_n)$. Consider the first term on the right-hand side of (A.2.19). By the Cauchy inequality and the similar argument as in (A.2.7):

$$E(|\delta_1|^2T_z^2) \le \{E|\delta_1|^4E(T_z^4)\}^{1/2} = O\left(n^{1/2}h_z^{q_z/2} + n^2h_z^{2(q_z+\xi_1)}\right).$$

Similarly, as in (A.2.10) and (A.2.11) with the moment restrictions on δ_i , the other two terms in (A.2.19) are:

$$E\left(|\delta_2|^2\left(L_{12}^{(1)}\right)^2t_{z,2}^2\right) \le \left[E\left\{|\delta_2|^4E_2\left(L_{12}^{(1)}\right)^4\right\}E(t_{z,2}^4)\right]^{1/2} = O(h_v^q h_\eta^q h_z^{q_z})^{1/2},$$

and:

$$\begin{aligned} E\left(|\delta_2|^2\left(L_{12}^{(1)}\right)^2T_{z,1}^2\right) &\le \left[E\left\{|\delta_2|^4E_2\left(L_{12}^{(1)}\right)^2\right\}E\left\{E_1\left(L_{12}^{(1)}\right)^2T_{z,1}^4\right\}\right]^{1/2} \\ &= O\left(n^{1/2}h_v^q h_\eta^q h_z^{q_z/2} + n^2h_v^q h_\eta^q h_z^{2(q_z+\xi_1)}\right). \end{aligned}$$

Thus (A.2.17) equals $O(n^{-3}h_v^{-3q/2}h_\eta^{-3q/2-2}h_z^{-3q_z/2}b_1^{-2}b_2^{-2} + n^{-1}h_v^{-q}h_\eta^{-q-2}h_z^{2\xi_1}b_1^{-2}b_2^{-2})$.

Next, we consider (A.2.18):

$$E\left(\breve\delta_1\breve\delta_2' I_{2,1}I_{2,2}'|\mathcal{L}\right) = (nh_v^q h_\eta^{q+1})^{-2}\hat f(V_1,\eta_1)^{-1}\hat f(V_2,\eta_2)^{-1}E\left(\sum_{i=1}^{n}|\delta_i|^2L_{1i}^{(1)}L_{2i}^{(1)}|\mathcal{L}\right).$$

Therefore, (A.2.18) is bounded by:

$$(n^{-3}h_v^{-2q}h_\eta^{-2q-2}h_z^{-2q_z}b_1^{-2}b_2^{-2})\,\mathcal{C}E\left\{(|\delta_1|^2+|\delta_2|^2)\left(t_{z,2}^2 + \left|L_{12}^{(1)}\right|^2T_{z,1}^2\right) + n|\delta_3|^2\left|L_{13}^{(1)}L_{23}^{(1)}\right|(t_{z,2}^2+t_{z,3}^2+T_{z,2}^2)\right\},$$

where $T_{z,2} = T_{z,1} - t_{z,3}$. Similar to the procedure in (A.2.10) and (A.2.11), for $\imath = 1$ or 2, we have:

$$E(|\delta_l|^2t_{z,2}^2) = O(h_z^{q_z/2}),$$

and:

$$E\left(|\delta_\imath|^2\left|L_{12}^{(1)}\right|^2T_{z,1}^2\right) = O\left(nh_v^q h_\eta^q h_z^{q_z} + n^2h_v^q h_\eta^q h_z^{2(q_z+\xi_1)}\right).$$

As in (A.2.12) and (A.2.13) with the bounded moment restriction on δ_i, for $\imath = 2$ or 3, we have:

$$E\left(|\delta_3|^2\left|L_{13}^{(1)}L_{23}^{(1)}\right|t_{z,\imath}^2\right) \le \left[E\left\{|\delta_3|^4E_3\left(L_{13}^{(1)}\right)^2\right\}E\left\{E_3\left(L_{23}^{(1)}\right)^2t_{z,\imath}^4\right\}\right]^{1/2} = O(h_v^q h_\eta^q h_z^{q_z/2}),$$

and:

$$
\begin{aligned}
E\left(|\delta_3|^2 \left|L_{13}^{(1)} L_{23}^{(1)}\right| T_{z,2}^2\right) &\leq \left[E\left\{|\delta_3|^4 E_1 \left|L_{13}^{(1)}\right| E_3 \left|L_{23}^{(1)}\right|\right\} E\left\{E_1 \left|L_{13}^{(1)}\right| E_3 \left|L_{23}^{(1)}\right| T_{z,2}^4\right\}\right]^{1/2}, \\
&= O\left(n^{1/2} h_v^{2q} h_\eta^{2q} h_z^{q_z/2} + n^2 h_v^{2q} h_\eta^{2q} h_z^{2(q_z+\xi_1)}\right).
\end{aligned}
$$

Thus (A.2.18) $= O\left(n^{-2} h_v^{-q} h_\eta^{-q-2} h_z^{-3q_z/2} b_1^{-2} b_2^{-2} + h_\eta^{-2} h_z^{2\xi_1} b_1^{-2} b_2^{-2}\right)$. ■

Proposition A.2.3.

(1) $\sqrt{n} S_{\delta_x \delta_e} = O_p\left(n^{-3/2} h_v^{-q} h_\eta^{-q-2} h_z^{-3q_z/2} b_1^{-2} b_2^{-2} + n^{1/2} h_\eta^{-2} h_z^{2\xi_1} b_1^{-2} b_2^{-2}\right)$;

(2) $\sqrt{n} S_{\delta_x \delta_m} = O_p\left(n^{-3/2} h_v^{-q} h_\eta^{-q-2} h_z^{-3q_z/2} b_1^{-2} b_2^{-2} + n^{1/2} h_\eta^{-2} h_z^{2\xi_1} b_1^{-2} b_2^{-2}\right)$.

Proof: The Cauchy inequality and Proposition A.2.2 (1) to (3) provide the proof. ■

Proposition A.2.4.

(1) $\sqrt{n} S_{U\delta_m} = O_p\left(n^{-1} h_v^{-q/2} h_\eta^{-q/2-1} h_z^{-7q_z/8} b_1^{-1} b_2^{-1} + h_\eta^{-1} h_z^{\xi_1} b_1^{-1} b_2^{-1}\right)$;

(2) $\sqrt{n} S_{U\delta_e} = O_p\left(n^{-1} h_v^{-q/2} h_\eta^{-q/2-1} h_z^{-7q_z/8} b_1^{-1} b_2^{-1} + h_\eta^{-1} h_z^{\xi_1} b_1^{-1} b_2^{-1}\right)$;

(3) $\sqrt{n} S_{e\delta_x} = O_p\left(n^{-1} h_v^{-q/2} h_\eta^{-q/2-1} h_z^{-7q_z/8} b_1^{-1} b_2^{-1} + h_\eta^{-1} h_z^{\xi_1} b_1^{-1} b_2^{-1}\right)$.

Proof: Let us denote ε_i as U_i and e_i in the rest of the paper. Then, by identity distribution:

$$
\begin{aligned}
E(\sqrt{n} S_{\varepsilon\hat{\delta}})^2 &\leq E\left\{\frac{1}{n}\sum_{i=1}^n \breve{\delta}_i^2 I_{2,i} I_1 |\varepsilon_1|^2\right\} && \text{(A.2.20)} \\
&+ \left|E\left\{\frac{1}{n}\sum_{i=1}^n \sum_{j\neq i}^n \breve{\delta}_i \breve{\delta}_j I_{2,i} I_{2,j} I_1 |\varepsilon_1|^2\right\}\right|. && \text{(A.2.21)}
\end{aligned}
$$

Because $E(|\breve{\delta}_1|^2 I_{2,1} | \mathcal{L}_n) \leq (n h_v^q h_\eta^{q+1} b_2)^{-2} E\left(\sum_{i=1}^n |\delta_i|^2 \left(L_{1i}^{(1)}\right)^2 |\mathcal{L}\right)$ *a.s.*, the right-hand side of (A.2.20) is bounded by $(n^2 h_v^q h_\eta^{q+1} h_z^{q_z} b_1 b_2)^{-2}$ multiplies by:

$$
E(\delta_i^2 |\varepsilon|^2) \leq \left[E|\varepsilon|^4 E\{\delta_i\}^4\right]^{1/2} \leq \left[E|\varepsilon|^4 E\left\{E\left(\sum_{i=1}^n |\delta_i|^8 \left(L_{1i}^{(1)}\right)^8\right) E(T_z^8)\right\}^{1/2}\right]^{1/2},
$$

by the Cauchy inequality. By the bound condition on the kernel function, the sum in the above bracket is:

$$
\left(L_{11}^{(1)}\right)^8 E|\delta|^8 + (n-1) E\left(|\delta_2|^8 \left|L_{12}^{(1)}\right|^8\right) \leq \mathcal{C} E|\delta|^8 + n E\left\{|\delta|^8 E_2 \left(L_{12}^{(1)}\right)^8\right\} \leq \mathcal{C}(1 + n h_v^q h_\eta^q) E|\delta|^8.
$$

By the similar argument as in (A.2.7), we have:

$$
E(T_z^8) = O\left(n h_z^{q_z} + n^8 h_z^{8(q_z+\xi_1)}\right).
$$

Hence (A.2.20) equals $O(n^{-7/2} h_v^{-7q/4} h_\eta^{-7q/4-2} h_z^{-7q_z/4} b_1^{-2} b_2^{-2} + n^{-7/4} h_v^{-7q/4} h_\eta^{-7q/4-2} h_z^{2\xi_1} b_1^{-2} b_2^{-2})$.

Next, we consider (A.2.21):

$$
E\left(\breve{\delta}_1 \breve{\delta}_2 I_{2,1} I_{2,2} | \mathcal{L}_n\right) = (n h_v^q h_\eta^{q+1})^{-2} \hat{f}^{-1}(V_1, \eta_1) \hat{f}^{-1}(V_2, \eta_2) E\left(\sum_{i=1}^n |\delta_i|^2 L_{1i}^{(1)} L_{2i}^{(1)} | \mathcal{L}_n\right).
$$

(A.2.21) is therefore bounded by:

$$\left(n^{-3}h_v^{-2q}h_\eta^{-2q-2}h_z^{-2q_z}b_1^{-2}b_2^{-2}\right)\mathcal{C}E\left\{(|\delta_1|^2+|\delta_2|^2)\left(t_{z,2}^2+\left|L_{12}^{(1)}\right|^2T_{z,1}^2\right)+n|\varepsilon|^2|\delta_3|^2\left|L_{13}^{(1)}L_{23}^{(1)}\right|(t_{z,2}^2+t_{z,3}^2+T_{z,2}^2)\right\}.$$

Similar to (A.2.10) and (A.2.11), and with the bounded moment conditions on δ_i, we have:

$$E\left(|\delta_\imath|^2\left|L_{12}^{(1)}\right|^2T_{z,1}^2\right)=O\left(nh_v^qh_\eta^qh_z^{q_z}+n^2h_v^qh_\eta^qh_z^{2(q_z+\xi_1)}\right),$$

and, for $\imath=1$ or 2, we have:

$$E(|\delta_\imath|^2t_{z,2}^2)\le\left\{E|\delta_l^4|E(t_{z,2}^4)\right\}^{1/2}=O(h_z^{q_z/2}).$$

By the Cauchy inequality, and the bound condition on the kernel function and the bounded moment condition on the function $m_v(z)$, for $\imath=2$ or 3, we have:

$$\begin{aligned}E\left(|\varepsilon|^2|\delta_3|^2\left|L_{13}^{(1)}L_{23}^{(1)}\right|t_{z,\imath}^2\right) &\le \left[E|\varepsilon|^4E\left\{|\delta_3|^4\left(L_{13}^{(1)}L_{23}^{(1)}\right)^2t_{z,\imath}^4\right\}\right]^{1/2}\\ &\le \left[E|\varepsilon|^4\left(E\left\{|\delta_3|^8E_3\left(L_{13}^{(1)}\right)^2E_3\left(L_{23}^{(1)}\right)^2\right\}E\left\{\left(L_{13}^{(1)}\right)^2E_3\left(L_{23}^{(1)}\right)^2t_{z,\imath}^8\right\}\right)^{1/2}\right]^{1/2}\\ &= O\left(h_v^qh_\eta^qh_z^{q_z/4}\right).\end{aligned}$$

By the Cauchy inequality, the bound condition on the kernel function and the similar argument as in (A.2.7):

$$\begin{aligned}E\left(|\varepsilon|^2|\delta_3|^2\left|L_{13}^{(1)}L_{23}^{(1)}\right|T_{z,2}^2\right) &\le \left[E|\varepsilon|^4E\left\{|\delta_3|^4\left(L_{13}^{(1)}\right)^2\left(L_{23}^{(1)}\right)^2T_{z,2}^4\right\}\right]^{1/2}\\ &\le \left[E\left(|\varepsilon|^4\left\{E_1\left(L_{13}^{(1)}\right)E_3\left(L_{23}^{(1)}\right)\right\}\right)\left(E\left\{|\delta_3|^8E_1\left(L_{13}^{(1)}\right)E_3\left(L_{23}^{(1)}\right)\right\}\right)^{1/2}\right]^{1/2}\\ &\times \left[E\left(|\varepsilon|^4\left\{E_1\left(L_{13}^{(1)}\right)E_3\left(L_{23}^{(1)}\right)\right\}\right)\left(E\left\{E_1\left(L_{13}^{(1)}\right)E_3\left(L_{23}^{(1)}\right)T_{z,2}^8\right\}\right)^{1/2}\right]^{1/2}\\ &= O\left(n^{1/4}h_v^{2q}h_\eta^{2q}h_z^{q_z/4}+n^2h_v^{2q}h_\eta^{2q}h_z^{2(q_z+\xi_1)}\right).\end{aligned}$$

Thus (A.2.21) equals $O\left(n^{-2}h_v^{-q}h_\eta^{-q-2}h_z^{-7q_z/4}b_1^{-2}b_2^{-2}+h_\eta^{-2}h_z^{2\xi_1}b_1^{-2}b_2^{-2}\right)$. ■

Proposition A.2.5.

(1) $\sqrt{n}S_{\hat{U}\delta_m}=O_p\left(n^{-5/4}h_v^{-3q/2}h_\eta^{-3q/2-1}h_z^{-3q_z/4}b_1^{-1}b_2^{-2}+n^{-1/2}h_v^{-3q/2}h_\eta^{-3q/2-1}h_z^{\xi_1}b_1^{-1}b_2^{-2}\right)$;

(2) $\sqrt{n}S_{\hat{U}\delta_e}=O_p\left(n^{-5/4}h_v^{-3q/2}h_\eta^{-3q/2-1}h_z^{-3q_z/4}b_1^{-1}b_2^{-2}+n^{-1/2}h_v^{-3q/2}h_\eta^{-3q/2-1}h_z^{\xi_1}b_1^{-1}b_2^{-2}\right)$;

(3) $\sqrt{n}S_{\hat{e}\delta_x}=O_p\left(n^{-5/4}h_v^{-3q/2}h_\eta^{-3q/2-1}h_z^{-3q_z/4}b_1^{-1}b_2^{-2}+n^{-1/2}h_v^{-3q/2}h_\eta^{-3q/2-1}h_z^{\xi_1}b_1^{-1}b_2^{-2}\right)$.

Proof: Let us denote $\hat{\varepsilon}_i$ as $\hat{U}_i$ and $\hat{e}_i$ in the rest of the paper.

$$\begin{aligned}E(\sqrt{n}S_{\hat{\varepsilon}\hat{\delta}})^2 &\le E\left\{\frac{1}{n}\sum_{i=1}^n|\hat{\varepsilon}_i|^2I_{2,i}\breve{\delta}_i^2I_{2,i}I_{1,i}\right\} &\text{(A.2.22)}\\ &+ \left|E\left\{\frac{1}{n}\sum_{i=1}^n\sum_{j\neq i}^n\hat{\varepsilon}_i\hat{\varepsilon}_j'\breve{\delta}_i\breve{\delta}_jI_{1,i}I_{1,j}I_{2,i}I_{2,j}'I_{2,j}'I_{2,i}\right\}\right|, &\text{(A.2.23)}\end{aligned}$$

where $\hat{\varepsilon}_i = \hat{f}^{-1}(V_i,\eta_i)\frac{1}{nh_v^q h_\eta^q}\sum_{j=1}^n \varepsilon_j K_v\left(\frac{V_i-V_j}{h_v}\right)K_\eta\left(\frac{\eta_i-\eta_j}{h_\eta}\right)$. Because we have:

$$E(|\hat{\varepsilon}_1|^2 I_{2,1}|\mathcal{L}_n) \le (nh_v^q h_\eta^q b_2)^{-2} E\left(\sum_{i=1}^n |\varepsilon|^2 L_{1i}^2 |\mathcal{L}_n\right) a.s.,$$

the right-hand side of (A.2.22) is bounded by $(n^3 h_v^{2q} h_\eta^{2q+1} h_z^{q_z} b_1 b_2^2)^{-2}$ multiplies by

$$E\left(\sum_{i=1}^n |\varepsilon_i|^2 L_{1i}^2 \sum_{j=1}^n |\delta_j|^2 \left(L_{1j}^{(1)}\right)^2 T_z^2\right) \le \left\{E\left(\sum_{i=1}^n |\varepsilon_i|^4 L_{1i}^4 \sum_{j=1}^n |\delta_j|^4 \left(L_{1j}^{(1)}\right)^4\right) E(T_z^4)\right\}^{1/2}, \qquad \text{(A.2.24)}$$

where $L_{1i} = K_v\left(\frac{V_1-V_i}{h_v}\right)K_\eta\left(\frac{\eta_1-\eta_i}{h_\eta}\right)$. By the bound condition on the kernel function:

$$\begin{aligned} E\left(\sum_{i=1}^n |\varepsilon_i|^4 L_{1i}^4 \sum_{j=1}^n |\delta_j|^4 \left(L_{1j}^{(1)}\right)^4\right) &\le \mathcal{C}\left[E|\varepsilon|^4 + n\left\{|\delta_1|^4 E_1(L_{12}^4)\right\}\right] \\ &+ \mathcal{C}\left[n^2 E\left\{|\varepsilon_3|^4 E_1(L_{13}^4)|\delta_2|^4 E_1\left(L_{12}^{(1)}\right)^4\right\}\right] \\ &= O(n^2 h_v^{2q} h_\eta^{2q}). \end{aligned}$$

Hence (A.2.24) is $O\left(n^{3/2} h_v^q h_\eta^q h_z^{q_z/2} + n^3 h_v^q h_\eta^q h_z^{2(q_z+\xi_1)}\right)$. The right-hand side of (A.2.22) therefore equals $O\left(n^{-9/2} h_v^{-3q} h_\eta^{-3q-2} h_z^{-3q_z/2} b_1^{-2} b_2^{-4} + n^{-3} h_v^{-3q} h_\eta^{-3q-2} h_z^{2\xi_1} b_1^{-2} b_2^{-4}\right)$.

Next, consider (A.2.23). $E(\hat{\varepsilon}_1 \hat{\varepsilon}_2' I_{2,1} I_{2,2}' |\mathcal{L}_n) = (nh_v^q h_\eta^q)^{-2} \hat{f}^{-1}(V_1,\eta_1)\hat{f}^{-1}(V_2,\eta_2) E\left(\sum_{i=1}^n |\varepsilon_i|^2 L_{1i} L_{2i} |\mathcal{L}_n\right)$, so (A.2.23) is bounded by:

$$(n^5 h_v^{4q} h_\eta^{4q+2} h_z^{2q_z} b_1^2 b_2^4)^{-1}\left[E\left\{\left(\sum_{i=1}^n |\varepsilon_i|^2 L_{1i} L_{2i}\right)\left(\sum_{j=1}^n |\delta_j|^2 L_{1j}^{(1)} L_{2j}^{(1)}\right)\right\}^2 E(T_z^4)\right]^{1/2},$$

and the sum in the bracket above is, by the bound condition on the kernel function:

$$\begin{aligned} &E\left\{\left(\sum_{i=1}^n |\varepsilon_i|^2 L_{1i} L_{2i}\right)\left(\sum_{j=1}^n |\delta_j|^2 L_{1j}^{(1)} L_{2j}^{(1)}\right)\right\}^2 \\ &\le E\left\{\left(|\varepsilon_1|^4 |L_{12}| + n^2 |\varepsilon_3|^4 E_3 |L_{13}^2 L_{23}^2|\right) \times \left(|\delta_1|^4 \left|L_{12}^{(1)}\right| + |\delta_3|^4 \left|L_{13}^{(1)} L_{23}^{(1)}\right|^2 + n^2 |\delta_4|^4 E_4 \left|L_{14}^{(1)} L_{24}^{(1)}\right|^2\right)\right\} \\ &= O(n^4 h_v^{2q} h_\eta^{2q}). \end{aligned}$$

Hence (A.2.23) equals $O\left(n^{-5/2} h_v^{-3q} h_\eta^{-3q-2} h_z^{-3q_z/2} b_1^{-2} b_2^{-4} + n^{-1} h_v^{-3q} h_\eta^{-3q-2} h_z^{2\xi_1} b_1^{-2} b_2^{-4}\right)$. ■

Proposition A.2.6.

(1)

$$\begin{aligned}\sqrt{n}S_{m_x-\hat{m}_x,\delta_m} &= O_p\left(n^{-3/2}h_v^{-3q/4}h_\eta^{-3q/4-1}h_z^{-q_z/4}b_1^{-1}b_2^{-2}+n^{-3/4}h_v^{-3q/4}h_\eta^{-3q/4-1}h_z^{\xi_1}b_1^{-1}b_2^{-2}\right)\\ &+ O_p\left(n^{-3/4}h_v^{\varsigma}h_\eta^{\varsigma-1}h_z^{-3q_z/4}b_1^{-1}b_2^{-2}+h_v^{\varsigma}h_\eta^{\varsigma-1}h_z^{\xi_1}b_1^{-1}b_2^{-2}\right);\end{aligned}$$

(2)

$$\begin{aligned}\sqrt{n}S_{m_x-\hat{m}_x,\delta_e} &= O_p\left(n^{-3/2}h_v^{-3q/4}h_\eta^{-3q/4-1}h_z^{-q_z/4}b_1^{-1}b_2^{-2}+n^{-3/4}h_v^{-3q/4}h_\eta^{-3q/4-1}h_z^{\xi_1}b_1^{-1}b_2^{-2}\right)\\ &+ O_p\left(n^{-3/4}h_v^{\varsigma}h_\eta^{\varsigma-1}h_z^{-3q_z/4}b_1^{-1}b_2^{-2}+h_v^{\varsigma}h_\eta^{\varsigma-1}h_z^{\xi_1}b_1^{-1}b_2^{-2}\right);\end{aligned}$$

(3)

$$\begin{aligned}\sqrt{n}S_{m-\hat{m},\delta_x} &= O_p\left(n^{-3/2}h_v^{-3q/4}h_\eta^{-3q/4-1}h_z^{-q_z/4}b_1^{-1}b_2^{-2}+n^{-3/4}h_v^{-3q/4}h_\eta^{-3q/4-1}h_z^{\xi_1}b_1^{-1}b_2^{-2}\right)\\ &+ O_p\left(n^{-3/4}h_v^{\xi_2}h_\eta^{\xi_2-1}h_z^{-3q_z/4}b_1^{-1}b_2^{-2}+h_v^{\xi_2}h_\eta^{\xi_2-1}h_z^{\xi_1}b_1^{-1}b_2^{-2}\right).\end{aligned}$$

Proof: Let us denote φ_i as $m_{x,i}$ and m_i, and $\hat{\varphi}_i$ as $\hat{m}_{x,i}$ and $\hat{m}_i$.

$$E\left(\sqrt{n}S_{\varphi-\hat{\varphi},\hat{\delta}}\right)^2 \le E\left\{\frac{1}{n}\sum_{i=1}^{n}(\varphi_i-\hat{\varphi}_i)^2\breve{\delta}_i^2 I_{2,i}I_{2,i}'I_{1,i}\right\} \tag{A.2.25}$$

$$+\left|E\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}^{n}(\varphi_i-\hat{\varphi}_i)(\varphi_j-\hat{\varphi}_j)'I_{2,i}I_{2,j}'I_{2,j}'I_{2,i}\breve{\delta}_i\breve{\delta}_j I_{1,i}I_{1,j}\right\}\right|. \tag{A.2.26}$$

The right-hand side of (A.2.25) is bounded by $(n^3h_v^{2q}h_\eta^{2q+1}h_z^{q_z}b_1b_2^2)^{-2}$ multiplies by

$$\begin{aligned}& E\left\{\sum_{i=1}^{n}(\varphi_1-\varphi_i)^2L_{1i}^2\sum_{j=1}^{n}(m_{v,1}-m_{v,j})^2K_{z,1j}^2\sum_{l=1}^{n}|\delta_l|^2\left(L_{1l}^{(1)}\right)^2\right\}\\ \le & \left[E\left\{\sum_{i=1}^{n}(\varphi_1-\varphi_i)^2L_{1i}^2\sum_{j=1}^{n}(m_{v,1}-m_{v,j})^2K_{z,1j}^2\right\}^2 E\left\{\sum_{l=1}^{n}|\delta_l|^2\left(L_{1l}^{(1)}\right)^2\right\}^2\right]^{1/2}\\ \le & \left[\left\{(E(T_z^8)E(T_\varphi^8))\right\}^{1/2}E\left\{\sum_{l=1}^{n}|\delta_l|^2\left(L_{1l}^{(1)}\right)^2\right\}^2\right]^{1/2},\end{aligned} \tag{A.2.27}$$

by the Cauchy inequality. By the similar argument as in (A.2.7), we have:

$$E(T_\varphi^8)=O\left(nh_v^qh_\eta^q+n^8h_v^{8(q+\xi)}h_\eta^{8(q+\xi)}\right),$$

where $\xi=\xi_2$ when $\varphi_i=m_{x,i}$ and $\xi=\varsigma$ when $\varphi_i=m_i$. By the bound condition on the kernel function, the last term in (A.2.27) is:

$$
\begin{aligned}
E\left\{\sum_{l=1}^{n}|\delta_l|^2\left(L_{1l}^{(1)}\right)^2\right\}^2 &\leq \left(L_{11}^{(1)}\right)^4 E|\delta|^4+(n-1)E\left(|\delta_2|^4\left(L_{12}^{(1)}\right)^4\right)\\
&+ (n^2-1)E\left\{|\delta_2|^2|\delta_3|^2\left(L_{12}^{(1)}L_{13}^{(1)}\right)^2\right\}\\
&\leq \mathcal{C}E|\delta|^4+nE\left\{|\delta_2|^4E_2\left(L_{12}^{(1)}\right)^4\right\}+n^2E\left\{|\delta_2|^2|\delta_3|^2E_1\left(L_{12}^{(1)}\right)^2E_3\left(L_{13}^{(1)}\right)^2\right\}\\
&\leq \mathcal{C}\left(1+nh_v^qh_\eta^q+n^2h_v^{2q}h_\eta^{2q}\right)E|\delta|^4.
\end{aligned}
$$

Hence (A.2.27) is:

$$
O\left(n^{3/2}h_v^{5q/4}h_\eta^{5q/4}h_z^{q_z/4}+n^{13/4}h_v^{3q+2\xi}h_\eta^{3q+2\xi}h_z^{q_z/4}+n^{13/4}h_v^{5q/4}h_\eta^{5q/4}h_z^{2(q_z+\xi_1)}+n^5h_v^{3q+2\xi}h_\eta^{3q+2\xi}h_z^{2(q_z+\xi_1)}\right).
$$

The right-hand side of (A.2.25) equals:

$$
\begin{aligned}
&O\left(n^{-9/2}h_v^{-11q/4}h_\eta^{-11q/4-2}h_z^{-7q_z/4}b_1^{-2}b_2^{-4}+n^{-11/4}h_v^{-q+2\xi}h_\eta^{-q-2+2\xi}h_z^{-7q_z/4}b_1^{-2}b_2^{-4}\right)\\
&+O\left(n^{-11/4}h_v^{-11q/4}h_\eta^{-11q/4-2}h_z^{2\xi_1}b_1^{-2}b_2^{-4}+n^{-1}h_v^{-q+2\xi}h_\eta^{-q-2+2\xi}h_z^{2\xi_1}b_1^{-2}b_2^{-4}\right).
\end{aligned}
$$

Next, we consider (A.2.26). Since we already know that:

$$
E(\breve{\delta}_1\breve{\delta}_2I_{2,1}I_{2,2}|\mathcal{L})\leq\mathcal{C}(nh_v^qh_\eta^{q+1}b_2)^{-2}E\left(\sum_{i=1}^{n}|\delta_i|^2\mathrm{L}_{1i}^{(1)}L_{2i}^{(1)}|\mathcal{L}\right)a.s.,
$$

(A.2.26) is bounded by $\left(n^{-5}h_v^{-4q}h_\eta^{-4q-2}h_z^{-2q_z}b_1^{-2}b_2^{-4}\right)$ multiplies by

$$
E\left\{(|\delta_1|^2+|\delta_2|^2)\left(t_{z,2}^2t_{\varphi,2}^2+\left|L_{12}^{(1)}\right|T_{z,1}^2T_{\varphi,1}^2\right)+n|\delta_3|^2\left|L_{12}^{(1)}L_{23}^{(1)}\right|(t_{z,2}^2t_{\varphi,2}^2+t_{z,3}^2t_{\varphi,3}^2+T_{z,2}^2T_{\varphi,2}^2)\right\},
$$

where $T_{\varphi,2}=T_{\varphi,1}-t_{\varphi,3}$. By the bound condition on the kernel function and the bound moment condition on the functions $m_v(z)$, $m(v,\eta)$ and $m_x(v,\eta)$:

$$
E(|\delta_l|^2t_{z,2}^2t_{\varphi,2}^2)=O(h_v^qh_\eta^qh_z^{q_z})
$$

and, by the bound condition on the kernel function and the similar argument as in (A.2.7), for $\imath=2$ or 3:

$$
\begin{aligned}
&E\left(|\delta_\imath|^2\left|L_{12}^{(1)}\right|T_{z,1}^2T_{\varphi,1}^2\right)=\\
&O\left(n^2h_v^{2q}h_\eta^{2q}h_z^{q_z}+n^3h_v^{3q+2\xi}h_\eta^{3q+2\xi}h_z^{q_z}+n^3h_v^{2q}h_\eta^{2q}h_z^{2(q_z+\xi_1)}+n^4h_v^{(3q+2\xi)}h_\eta^{(3q+2\xi)}h_z^{2(q_z+\xi_1)}\right).
\end{aligned}
$$

By the bound condition on the kernel function and the bound moment condition on the functions $m_v(z)$, $m(v,\eta)$ and $m_x(v,\eta)$, for $\imath=2$ or 3:

$$
\begin{aligned}
E\left(|\delta_3|^2\left|L_{12}^{(1)}L_{23}^{(1)}\right|t_{z,\imath}^2t_{\varphi,\imath}^2\right) &\leq \left[E\left\{|\delta_3|^4E_3\left(L_{13}^{(1)}\right)^2E_3\left(L_{23}^{(1)}\right)^2\right\}E\left\{t_{z,\imath}^4t_{\varphi,\imath}^4\right\}\right]^{1/2}\\
&= O\left(h_v^{2q}h_\eta^{2q}h_z^{q_z}\right)^{1/2}.
\end{aligned}
$$

By the bound condition on the kernel function and the similar argument as in (A.2.7),

$$
E\left(|\delta_3|^2\left|L_{13}^{(1)}L_{23}^{(1)}\right|T_{z,2}^2T_{\varphi,2}^2\right)\leq E\left[\left\{|\delta_3|^4E_1\left|L_{13}^{(1)}\right|E_3\left|L_{23}^{(1)}\right|\right\}E\left\{T_{z,2}^4T_{\varphi,2}^4E_1\left(\left|L_{13}^{(1)}\right|E_3\left|L_{23}^{(1)}\right|\right)\right\}\right]^{1/2}
$$

which is $O\left(nh_v^{5q/2}h_\eta^{5q/2}h_z^{q_z/2}+n^{5/2}h_v^{5q/2}h_\eta^{5q/2}h_z^{2(q_z+\xi_1)}+n^{5/2}h_v^{4q+2\xi}h_\eta^{4q+2\xi}h_z^{q_z/2}+n^4h_v^{(4q+2\xi)}h_\eta^{(4q+2\xi)}h_z^{2(q_z+\xi_1)}\right)$. Hence (A.2.26) is:

$$O\left(n^{-3}h_v^{-3q/2}h_\eta^{-3q/2-2}h_z^{-3q_z/2}b_1^{-2}b_2^{-4}+n^{-3/2}h_v^{-3q/2}h_\eta^{-3q/2-2}h_z^{2\xi_1}b_1^{-2}b_2^{-4}+n^{-3/2}h_v^{2\xi}h_\eta^{2\xi-2}h_z^{-3q_z/2}b_1^{-2}b_2^{-4}\right)$$
$$+O\left(h_v^{2\xi}h_\eta^{-2+2\xi}h_z^{2\xi_1}b_1^{-2}b_2^{-4}\right).$$

■

6.A.3. Proof of Theorem 2.2

Let us define $\check{m}(v)=\frac{1}{n}\sum_{i=1}^n m(v,\eta_i)$. We omit τ in $\hat{m}_\tau(v)$ and $\hat{\beta}_\tau$ throughout the proof, since it is a trivial indicator for the proof of the consistency of the unknown structural function. The condition of boundness on the function $m(v,\eta)$ and the *i.i.d.* assumption on η_i allow us to apply the Chebyshev's law of large numbers as carried out by Gao et al. (2006):

$$\begin{aligned}\hat{m}(v)-m(v) &= \hat{m}(v)-\check{m}(v)+\check{m}(v)-m(v)\\ &= \hat{m}(v)-\check{m}(v)+O_p(n^{-1/2}),\end{aligned}$$

where:

$$\hat{m}(v)-\check{m}(v)=\frac{1}{n}\sum_{i=1}^n\{\hat{m}(v,\hat{\eta}_i)-m(v,\eta_i)\}. \tag{A.3.1}$$

Given $\hat{\beta}$ and by using the definition of $m(v,\eta_i)$, we can rewrite the term in the bracket of (A.3.1) as:

$$\begin{aligned}\hat{m}(v,\hat{\eta}_i)-m(v,\eta_i) &= \{\hat{m}_y(v,\eta_i)-m_y(v,\eta_i)+\delta_{m_y,i}\}-\{\hat{m}_x(v,\eta_i)-m_x(v,\eta_i)+\delta_{m_x,i}\}'\beta\\ &- \{\hat{m}_x(v,\eta_i)-m_x(v,\eta_i)+\delta_{m_x,i}\}'\{\hat{\beta}-\beta\}\\ &= \{\hat{m}_{y^{**}}(v,\eta_i)-m_{y^{**}}(v,\eta_i)+\delta_{y^{**},i}\}\\ &- \{\hat{m}_x(v,\eta_i)-m_x(v,\eta_i)+\delta_{m_x,i}\}'\{\hat{\beta}-\beta\},\end{aligned} \tag{A.3.2}$$

where $\delta_{m_y,i}=\hat{m}_y(v,\hat{\eta}_i)-\hat{m}_y(v,\eta_i)$, $\delta_{m_x,i}=\hat{m}_x(v,\hat{\eta}_i)-\hat{m}_x(v,\eta_i)$, $Y_i^{**}=Y_i-X_i'\beta$, and $\delta_{y^{**},i}=\delta_{y,i}-\delta_{x,i}'\beta=\hat{m}(y^{**}|v,\hat{\eta}_i)-\hat{m}(y^{**}|v,\eta_i)$. We use a similar set of arguments as in Propositions A.2.1 (1) - (2) and A.2.2 (1) - (3), and uniform boundness in Härdle et al. (1993). Let ψ_i be a possibly quantity for which we show that, for all integers $l\geq 1$:

$$\sup_i|\psi_i|=o_p(n^a)\ \text{ since }\ \sup_i E\left(\psi_i/n^{a^*}\right)^{2l}=O(1),$$

where $a^*<a$ (see Step (ii) in Section 4 of Härdle et al. (1993), for details about this). Hence we have, uniformly in i:

$$\hat{m}_x(v,\eta_i)-m_x(v,\eta_i)=O_p\left((nh_v^qh_\eta^q)^{-1/2}+h_v^\varsigma h_\eta^\varsigma b_2^{-1}\right),$$

and:

$$\delta_{x,i}=O_p\left(n^{-1}h_v^{-q/2}h_\eta^{-q/2-1}h_z^{-q_z/4}b_1^{-1}b_2^{-1}+h_\eta^{-1}h_z^{\xi_1}b_1^{-1}b_2^{-1}\right).$$

Hence (A.3.2) is:

$$\hat{m}(v,\hat{\eta}_i)-m(v,\eta_i)=\{\hat{m}_{y^{**}}(v,\eta_i)-m_{y^{**}}(v,\eta_i)+\delta_{y^{**},i}\}+o_p(1), \tag{A.3.3}$$

where $\delta_{y^{**}}=O_p\left(n^{-1}h_v^{-q/2}h_\eta^{-1}h_z^{-q_z/4}b_1^{-1}b_2^{-1}+h_\eta^{-1}h_z^{\xi_1}b_1^{-1}b_2^{-1}\right)$ uniformly in i.

Take the sample mean version of the marginal integration of equation (A.3.3),

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\{\hat{m}(v,\hat{\eta}_i)-m(v,\eta_i)\} &= \frac{1}{n}\sum_{i=1}^{n}\{\hat{m}_{y^{**}}(v,\eta_i)-m_{y^{**}}(v,\eta_i)\}+o_p(1)\\ &\equiv \frac{1}{n}\sum_{i=1}^{n}\{\hat{m}(v,\eta_i)-m(v,\eta_i)\}+o_p(1). \qquad (A.3.4)\end{aligned}$$

Define $\tilde{m}(v,\eta_i)=\hat{m}(v,\eta_i)\hat{f}(v,\eta_i)$. We then rewrite the last term in the bracket of (A.3.4) as:

$$\begin{aligned}\hat{m}(v,\eta_i)-m(v,\eta_i) &= \frac{\tilde{m}(v,\eta_i)-m(v,\eta_i)\hat{f}(v,\eta_i)}{\hat{f}(v,\eta_i)}\\ &= \frac{\tilde{m}(v,\eta_i)-m(v,\eta_i)\hat{f}(v,\eta_i)}{f(v,\eta_i)}\left[1-\frac{\hat{f}(v,\eta_i)-f(v,\eta_i)}{\hat{f}(v,\eta_i)}\right]. \qquad (A.3.5)\end{aligned}$$

Note that the term $\left(\hat{f}(v,\eta_i)-f(v,\eta_i)\right)/\hat{f}(v,\eta_i)$ is $O_p\left(h_v^{p_2}h_\eta^{p_2}+(nh_v^q h_\eta^q)^{-1/2}\right)$ uniformly in i and hence it can be dropped. We now consider the bias term:

$$E(\hat{m}(v,\eta_i)-m(v,\eta_i))=f^{-1}(v,\eta_i)\left(E\tilde{m}(v,\eta_i)-m(v,\eta_i)E(\hat{f}(v,\eta_i))\right),$$

where:

$$\begin{aligned}E\tilde{m}(v,\eta_i) &= E\left[\frac{1}{nh_v^q h_\eta^q}\sum_{j=1}^{n}K_v\left(\frac{V_j-v}{h_v}\right)K_\eta\left(\frac{\eta_j-\eta_i}{h_\eta}\right)Y_j^{**}\right]\\ &= E\left[E_{v,\eta_i}\left\{\frac{1}{nh_v^q h_\eta^q}\sum_{j=1}^{n}K_v\left(\frac{V_j-v}{h_v}\right)K_\eta\left(\frac{\eta_j-\eta_i}{h_\eta}\right)Y_j^{**}\right\}\right]\\ &= E\left[\frac{1}{nh_v^q h_\eta^q}\sum_{j=1}^{n}K_v\left(\frac{V_j-v}{h_v}\right)K_\eta\left(\frac{\eta_j-\eta_i}{h_\eta}\right)m(V_j,\eta_j)\right]\\ &= f(v,\eta_i)m(v,\eta_i)+\mathcal{K}_{v,p_2}h_v^{p_2}\sum_{r=1}^{p_2}f_v^{(r)}(v,\eta_i)m^{(p_2-r)}(v)+\mathcal{K}_{\eta,p_2}h_\eta^{p_2}\sum_{r=1}^{p_2}f_\eta^{(r)}(v,\eta_i)m^{(p_2-r)}(\eta_i)\\ &+ O(h_v^{p_2+1})+O(h_\eta^{p_2+1}).\end{aligned}$$

E_{v,η_i} denotes as the expectation conditional on v and η_i, and $\mathcal{K}_{v,p_2}=\int v^{p_2}K_v(v)dv$ and $\mathcal{K}_{\eta,p_2}=\int \eta^{p_2}K_\eta(\eta)d\eta$. Hence we have:

$$E(\hat{m}(v,\eta_i)-m(v,\eta_i))=\{h_v^{p_2}B_v(v,\eta_i)+h_\eta^{p_2}B_\eta(v,\eta_i)\}+o(1). \qquad (A.3.6)$$

The single sum of (A.3.6) converges to its population mean by Chebyshev's law of large numbers; see Linton and Härdle (1996), for example. Now we consider the variance term. Note that $f(v,\eta_i)=f(v,\eta)+O_p(n^{1/2})$ and $m(v,\eta_i)=m(v,\eta)+O_p(n^{-1/2})$ by the law of large numbers since both functions satisfy the bounded moment conditions. Therefore, we have:

$$\begin{aligned}V\left(\frac{1}{n}\sum_{i=1}^{n}\hat{m}(v,\eta_i)\right) &= f(v,\eta)^{-2}V\left(\frac{1}{n}\sum_{i=1}^{n}\left\{\tilde{m}(v,\eta_i)-m(v,\eta_i)\hat{f}(v,\eta_i)\right\}\right)\\ &= f(v,\eta)^{-2}V\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{m}(v,\eta_i)\right)+f(v,\eta)^{-2}m(v,\eta)^2V\left(\frac{1}{n}\sum_{i=1}^{n}\hat{f}(v,\eta_i)\right)\\ &- f(v,\eta)^{-2}2m(v,\eta)Cov\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{m}(v,\eta_i),\frac{1}{n}\sum_{i=1}^{n}\hat{f}(v,\eta)\right),\end{aligned}$$

where $V(\cdot)$ and $Cov(\cdot)$ denote variance and covariance, respectively, and:

$$\begin{aligned}
V\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{m}(v,\eta_i)\right) &= E\left(V_{v,\eta_i}\left\{\frac{1}{n}\sum_{i=1}^{n}\tilde{m}(v,\eta_i)\right\}\right)+V\left(E_{v,\eta_i}\left\{\frac{1}{n}\sum_{i=1}^{n}\tilde{m}(v,\eta_i)\right\}\right)\\
&= \sigma^2 f(\eta)^2 E\left[\frac{1}{nh_v^q}\sum_{j=1}^{n}K_v\left(\frac{V_j-v}{h_v}\right)\right]^2+f(\eta)^2 V\left[\frac{1}{nh_v^q}\sum_{j=1}^{n}K_v\left(\frac{V_j-v}{h_v}\right)m(V_j,\eta_j)\right]\\
&= \frac{\sigma^2 f(\eta)^2}{nh_v^q}\mathcal{K}_v+\frac{m(v,\eta)^2 f(\eta)^2 f(v)}{nh_v^q}\mathcal{K}_v+O(n^{-1})\\
V\left(\frac{1}{n}\sum_{i=1}^{n}\hat{f}(v,\eta_i)\right) &= \frac{f(\eta)^2 f(v)\mathcal{K}_v}{nh_v^q}+O(n^{-1})\\
Cov\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{m}(v,\eta_i),\frac{1}{n}\sum_{i=1}^{n}\hat{f}(v,\eta_i)\right) &= E\left\{\frac{1}{n}\sum_{i=1}^{n}\tilde{m}(v,\eta_i)\frac{1}{n}\sum_{i=1}^{n}\hat{f}(v,\eta_i)\right\}-E\left\{\frac{1}{n}\sum_{i=1}^{n}\tilde{m}(v,\eta_i)\right\}E\left\{\frac{1}{n}\sum_{i=1}^{n}\hat{f}(v,\eta_i)\right\}\\
&= \frac{m(v,\eta)f(\eta)^2 f(v)\mathcal{K}_v}{nh_v^q}+O(n^{-1}).
\end{aligned}$$

V_{v,η_i} denotes the variance conditional on v and η_i. Hence we have:

$$\sqrt{nh_v^q}(\hat{m}(v)-m(v)-bias)\to_D N(0,var).$$

The consistency of $\hat{g}_\tau(v)$ and its asymptotic normality is argued in the same way as above, since $m(v)=g(v)+\mathcal{C}$. ■

6.A.4. Proof of Proposition 3.1

For the sake of convenience and clarity in introducing the main idea of the proof, let us first point out that the adaptive data-driven estimation problem at hand can be simplified to that of the semiparametric simultaneous equations such that, for the example where $p=q=q_z=1$, we have:

$$\begin{aligned}
Y_i &= \beta X_i+g(V_i,\eta_i)+\epsilon_i\\
\eta_i &= V_i-m_v(Z_i) \qquad \text{(A.4.1)}\\
E(\epsilon|z,\eta) &= E(\epsilon|\eta) \;\text{ and }\; E(\eta|z)=0.
\end{aligned}$$

For the case where η_i is an observable regressor, to check (3.4), it is enough to show that:

$$\sup_{h_v,h_\eta,h'_v,h'_\eta\in H_n}\frac{\left|D(h_v,h_\eta)-D(h'_v,h'_\eta)-\left\{CV(h_v,h_\eta)-CV(h'_v,h'_\eta)\right\}\right|}{D(h_v,h_\eta)}=o_P\{1\}, \qquad \text{(A.4.2)}$$

which is well-known in the literature (see Härdle et al. (2000), for example). The second equation in (A.4.1) suggests, however, that for the current case, we also have to show the following:

$$\sup_{h_v,h_{\hat\eta}\in H_n}\frac{\left|\hat{D}(h_v,h_{\hat\eta})-D(h_v,h_\eta)-\left\{\hat{CV}(h_v,h_{\hat\eta})-CV(h_v,h_\eta)\right\}\right|}{D(h_v,h_\eta)}=O_P\{1\}. \qquad \text{(A.4.3)}$$

A detailed decomposition of $\hat{D}(h_v,h_{\hat\eta})-D(h_v,h_\eta)$ and $\hat{CV}(h_v,h_{\hat\eta})-CV(h_v,h_\eta)$, as done for the proof of Theorem 2.2 in Saart et al. (2013), suggests that a number of results are required in order to check (A.4.3). By letting $\widehat{g}_1(v_i,\widehat{\eta}_i)\equiv\hat{E}(y|v,\hat\eta)$ and $\widehat{g}_2(v_i,\widehat{\eta}_i)\equiv\hat{E}(x|v,\hat\eta)$, these results are, for $j=1,2$, and uniformly over $h_v,h_{\widehat\eta}\in H_n$:

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\widehat{g}_j(v_i,\widehat{\eta}_i)-g_j(v_i,\eta_i)\right\}b_i=O_P\left\{D(h_v,h_\eta)\right\}, \qquad \text{(A.4.4)}$$

where $b_i = x_i - g_2(v_i, \eta_i)$, and:

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\widehat{g}_j(v_i,\hat{\eta}_i)-g_j(v_i,\eta_i)\right\}^2 = O_P\left\{D(h_v,h_\eta)\right\}. \tag{A.4.5}$$

According to Härdle et al. (2000), it is a straightforward task to show, using

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\widehat{g}_j(v_i,\eta_i)-g_j(v_i,\eta_i)\right\}^2 = O\left\{(nh_vh_\eta)^{-1}\right\}, \tag{A.4.6}$$

that, in fact:

$$D(h_v,h_\eta) = O\left\{(nh_vh_\eta)^{-1}\right\}. \tag{A.4.7}$$

In this case, if $\widehat{\eta}_i$ only converges to η_i pointwise in Z, but not uniformly, then we do not know whether the difference $\eta_i - \widehat{\eta}_i$ is converging to zero. Hence, one solution is to apply a uniform convergence result of the kernel regression in the first stage. Furthermore, the rate of such a uniform convergence is also useful in the proof of (A.4.4) and (A.4.5). A standard nonparametric kernel study (see Li and Racine (2007), for example) reports that the NW estimator can be shown to have a uniform convergence rate of:

$$O_P\left\{\left(\frac{\ln N}{N}\right)^{2/(q_z+4)}\right\}. \tag{A.4.8}$$

For example, when $q_z = 1$, we have:

$$O_P\left\{\left(\frac{\ln N}{N}\right)^{2/5}\right\} = o_P\left\{N^{-1/4}\right\} \tag{A.4.9}$$

for the model in (A.4.1); (A.4.9) is a weak version of Theorem 3.3.6 of Györfi et al. (1990).

The proof of (A.4.4) and (A.4.5), as done in Saart et al. (2013) for example, relies heavily on the use of the Taylor series expansion of the kernel function and (A.4.9). For a semiparametric simultaneous equation model obtained by replacing $g(V_i, \eta_i)$ in (A.4.1) by $g(\eta_i)$ and with some additional restrictions on the smoothing parameter, we are able to show the results for the case where $q_z \geq q = 1$. In the current case, these results are obtained much more conveniently especially given that the slowing down in the uniform convergence rate in (A.4.8) (as a function of q) is being compensated for by the slowing down in $D(h_v, h_\eta)$ (as the function of $2q$). By putting these together with the condition we imposed earlier in Section 2.4, we conclude that (A.4.4) and (A.4.5) hold for $q \leq q_z \leq 2q$. ■

7. References

References

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

Blundell, R., Powell, J., 2003. Endogeneity in nonparametric and semiparametric regression models. In: Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress, Volume II. Cambridge University Press, 312–357.

Blundell, R., Powell, J., 2004. Endogeneity in semiparametric binary response models. Review of Economics Studies 71, 655–679.

Darolles, S., Fan, Y., Florens, J.P., Renault, E., 2011. Nonparametric instrumental regression. Econometrica 79, 1541–1565.

Fan, J., Gijbels, I., 1996. Local Polynomial Modeling and its Applications. Chapman & Hall, London.

Fan, Y., Li, Q., 1999. Root-n-consistent estimation of partially linear time series models. Journal of Nonparametric Statistics 11, 251–269.

Fan, Y., Li, Q., Stengos, T., 1995. Root-n-consistent semiparametric regression with conditional heteroscedastic disturbances. Journal of Quantitative Economics 11, 229–140.

Gao, J., 2007. Nonlinear Time Series: Semiparametric and Nonparametric Methods. Chapman & Hall/CRC, London.

Gao, J., Lu, J., Tjøstheim, D., 2006. Estimation in semiparametric spatial regression. The Annals of Statistics 34, 1395–1435.

Györfi, L., Härdle, W., Sarda, P., Vieu, P., 1990. Nonparametric Curve Estimation from Time Series. volume 60 of *Lecture Notes in Statistics*. Springer-Verlag, Berlin.

Hall, P., Horowitz, J., 2005. Non-parametric method for inference in the presence of instrumental variables. The Annals of Statistics 33, 2904–2929.

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

Härdle, W., Liang, H., Gao, J., 2000. Partially Linear Models. Springer-Verlag, Berlin.

Härdle, W., Vieu, P., 1992. Kernel regression smoothing of time series. Journal of Time Series Analysis 13, 209–232.

Hart, J., Vieu, P., 1990. Data-driven bandwidth choice for density estimation based on dependent data. The Annals of Statistics 18, 873–890.

Hastie, T., Tibishirani, R., 1991. Generalized Additive Models. Chapman & Hall/CRC, Boca Raton, FL.

Hausman, J., 1978. Specification tests in econometrics. Econometrica 46, 1251–1271.

Li, Q., 1999. Consistent model specification tests for time series econometric models. Journal of Econometrics 92, 101–147.

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

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

Linton, A., Härdle, W., 1996. Estimation of additive regression models with known links. Biometrika 83, 529–540.

Linton, A., Nielsen, J., 1995. A kernel method of estimating structural nonparametric regression based on marginal integration. Biometrika 82, 93–100.

Newey, W., Powell, J., 2003. Instrumental variable estimation of nonparametric models. Econometrica 71, 1565–1578.

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

O'Sullivan, F., 1986. A statistical perspective on ill-posed inverse problems. Statistical Science 1, 502–518.

Pinkse, J., 2000. Nonparametric two-step regression estimation when regressors and error are dependent. The Canadian Journal of Statistics 28, 289–300.

Robinson, P., 1988. Root-N-consistent semiparametric regression. Econometrica 56, 931–954.

Saart, P.W., Kim, N.H., Reale, M., 2013. Optimal smoothing in a semiparametric simultaneous equations model of time series. Available at: https://sites.google.com/site/patrickwsaart/home .

Sargan, D., 1998. Lectures on Advanced Econometric Theory. Basil Blackwell.

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

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

Tjøstheim, D., Austad, R., 1996. Nonparametric identification of nonlinear time series: projection. Journal of American Statistical Association 89, 1398–1409.