The University of Adelaide

Semiparametric Models with Endogeneity and their Application to an Empirical Demand Analysis

by

Nam-Hyun Kim

This thesis is presented for the degree of

Doctor of Philosophy

at
the School of Economics

April 2013

For a thesis that does not contain work already in the public domain

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Signature

Date 23/08/2013

Acknowledgements

I would like to take this opportunity to express my sincere gratitude to a number of people and institutions, who have provided me with useful assistance and support at various stages of my PhD study.

Firstly, I thank my supervisors: Professor Jiti Gao, Dr Duygu Yengin and Dr Nicholas Sim. In specific, I offer my sincerest gratitude to Professor Jiti Gao, who has supported me throughout my thesis with his patience and knowledge. I also thank a number of faculty members at the School of Economics, The University of Adelaide, for their assistance, both while I was there and when I was a visiting PhD student at other institutions. Special thanks go to Professor Christopher Findlay, Dr Mandar Oak, Dr Ralph Bayer and Dr Jacob Wong for their assistance with administrative issues on various occasions. In addition, I am sincerely grateful to Dr Seungmoon Choi for his kind assistance and support on academic and various other issues, which helped me to overcome the difficult obstacles. I also thank Professor Fabrice Collard at the Department of Economics, The University of Bern, for his kind assistance with my presentation skills and his support.

I also thank the Department of Econometrics and Business Statistics, Monash University, especially Professor Brett Inder, for the opportunity to visit the department.

Similarly, many thanks to all the faculty members at the Department of Mathematics and Statistics, The University of Canterbury, for giving me such a warm welcome and hospitality, especially Professor Jennifer Brown, Dr Marco Reale, Dr Carl Scarrott, Dr Qui Bui and Dr Dominic Lee. Furthermore, Dr Marco Reale has broadened my research scope and perspective, and Dr Dominic Lee has provided me with such invaluable suggestions on the application of the work in this thesis.

This thesis is dedicated to my partner Patrick, my parents and my brother, who have provided me with absolute and unconditional support and encouragement throughout my PhD study. Furthermore, I offer my sincerest gratitude to Professor Bonggeun Kim at the Department of Economics, Seoul National University, for his unconditional support and invaluable suggestions during the difficult times. Without these very important people in my life, this thesis would not have been completed. I am also thankful for all their valuable support and dearly appreciate the friendship I obtained from Dr Duygu Yengin, Dr Prue Kerr and Dr Erkan Yalcin. In particular, Dr Duygu Yengin has given me absolute and unconditional friendship which has helped me to get through the most difficult times at Adelaide.

Abstract

During the past few decades, nonparametric models have been extensively applied to empirical studies in various fields of economics due to its flexibility for depicting any type of relationship among key economic variables. However, one of the most well-known shortfalls of the model is the *curse of dimensionality.* It can be conveniently overcome with semiparametric modelling such as partially linear (PL) models and/or single-index (SI) models. Nonetheless, the practicality of these models in the empirical studies has been hampered by the lack of appropriate estimation procedures and a method to address endogeneity. Hence the ultimate goal of this thesis is to establish a novel econometric method for estimating semiparametrics, specifically a PL model and an extended generalised partially linear single-index (EGPLSI) model, with the presence of endogeneity. Furthermore, semiparametric analysis is an important tool for analysing empirical Engel curves, which often involve endogeneity in total expenditure. We show that, our newly developed estimation procedures and methods are able to address the endogeneity problem in the semiparametric analysis of empirical Engel curves. These goals can be broken down into a few research objectives.

(1) Firstly, this thesis aims to construct a comprehensive and systematic treatment of endogeneity in semiparametrics, given the complexity of the models containing both parametric and nonparametric components.

(2) Secondly, it aims to develop novel estimation procedures and methods to address endogeneity in a PL model and an EGPLSI model.

(3) Lastly, it aims to analyse the empirical demand function semiparametrically by applying the estimation procedures and methods in this thesis.

Publications arising from the thesis

(1) Kim, N. and Saart, P. W. (2013). Estimation in Partially Linear Semiparametric Models with parametric and/or nonparametric endogeneity. *Under review at the Econometrics Journal.*

(2) Kim, N., Saart, P. W. and Gao, J. (2013). Semi-parametric Analysis of Shape-Invariant Engel Curves with Control Function Approach. *Under review at the Journal of Econometrics.*

Contents

List of Figures

List of Tables

Chapter 1

Background and Motivation

> *The analysis of data with endogenous regressors - that is, observable explanatory variables that are correlated with unobservable error terms - is arguably the main contribution of econometrics to statistical science. ... extensions to nonparametric and semiparametric models have only recently considered.*
>
> Richard Blundell and James L. Powell (2003)

1.1 Introduction

During the past two decades, nonparametric models have been extensively applied to a large number of empirical studies in various areas of economics, since it provides great flexibility in depicting any type of relationship among key economic variables. However, nonparametric techniques suffer from the well-known shortfall called the *"curse of dimensionality"* when the number of regressors is greater than 3; see Saart & Gao (2013) for details. The curse of dimensionality can be conveniently overcome with semiparametric modelling, such as partially linear (PL) models and single-index (SI) models.

After the study of Engle et al. (1986), a PL model became the most popular semiparametric techniques for the economic applications. The apparent advantage of employing the model in an empirical study is its capability to incorporate benefits of parametric and nonparametric analysis. It is flexible enough to depict any type of nonlinear relationship among key economic variables without prespecifying an appropriate functional form. Since it also contains a parametric component in the specification, it minimises the curse of dimensionality, allowing efficient use of the data-set at hand. Furthermore, we can easily interpret the effect of each variable and impose an economic hypothesis in the parametric component; see Blundell & Powell (2003) and Härdle et al. (2000) for examples. The PL type of semiparametric models, specifically the PL model of Robinson (1988), the generalised partially linear single-index (GPLSI) model of Carroll et al. (1997), and an extended generalised partially linear single-index (EGPLSI) model of Xia et al. (1999) have been theoretically established in recent years.

Although the SI model is well-known in the statistical literature for minimising the dimensionality of the regressors, the study of Ichimura (1993) motivates the use of the model in economic applications. Since the model does not require the pre-specification of the functional form, it provides more flexibility than the conventional nonlinear parametric analysis. Furthermore, it also allows for the choice modelling with great flexibility of no need to specify the distribution of disturbance term with some minor identification conditions; see Ichimura (1993), Blundell & Powell (2004) and chapter 2 of Horowitz (2009) for details. As it is mentioned above, the SI type of semiparametric models, specifically the SI model of Härdle et al. (1993), the GPLSI model of Carroll et al. (1997) and the EGPLSI model of Xia et al. (1999), have been theoretically established.

Nonetheless, the practicality of these above models in empirical studies has been hampered by the lack of appropriate estimation procedures and methods to address endogeneity. Endogeneity issues have only recently been considered in the nonparametric and semiparametric models; see Blundell & Powell (2003) for an excellent review of the issue. The so-called "endogeneity problem" is a technical name given by econometricians to a problem that is well-known in developmental studies and empirical economics; see Deaton & Muellbauer (1980*a*) and Nakamura & Nakamura (1998) for some excellent surveys. For example, the endogeneity of total expenditure is a well-known issue in the empirical demand literature (see Blundell et al. (1998) and Blundell et al. (2007) for details). Hence we aim to provide a comprehensive treatment and a systematic way to address endogeneity in semiparametric models, specifically the PL model and the EGPLSI model.

Furthermore, semiparametric modelling is an important tool to analyse empirical Engel curves since severe nonlinear relationships between particular budget shares and total expenditure are apparent (see Blundell et al. (1998) and Blundell et al. (2007) for example) and also the semiparametric models allow for the analysis of the effects of demographic variables on demand. More importantly, the EGPLSI model allows the shape-invariant specification of an empirical model which is coherent with the consumer optimisation theory (see Blundell et al. (1998), Blundell et al. (2003) and Blundell et al. (2007) for details). However, the endogeneity of total expenditure is a well-known problem in the literature. A number of studies, such as Blundell et al. (1998), and Blundell et al. (2007), address the problem with two common alternatives: the Control Function (CF) approach and nonparametric instrumental variables (NpIV) estimation, respectively. We conduct an empirical demand study in Australia with the methodologies proposed in this thesis. This empirical study shows the usefulness and practicality of our methodologies.

1.2 Review of Nonparametrics and Semiparametrics in the Presence of Endogeneity

Endogeneity was originally recognised and studied in the setting of a simultaneous system of equations, where explanatory variables were determined inside the system simultaneously with dependent variables. Hence the correlations between error terms and explanatory variables cause the inconsistency of the conventional estimators. A good review of the nature of endogeneity is presented in Nakamura & Nakamura (1998). Recently, endogeneity has been often observed in other settings. Measurement error, and mis-specifications in a model, such as an omitted variable or/and an omitted interaction term, are common causes of endogeneity in a cross-sectional case; the correlated random effect is main cause in the case of panel data. In this thesis, we focus on the formal setting.

As the partially linear semiparametric models possess complex features containing both parametric and nonparametric components, there are two sources of endogeneity. Those are *parametric endogeneity* and *nonparametric endogeneity*; the details of these are discussed in Chapter 2. Although the former causes the inconsistency of both unknowns estimators of a parametric coefficient and an unknown structural function the latter does not have an effect on the consistency of the parametric estimator when we conduct the two-step estimation procedure of Robinson (1988) and Speckman (1988) due to the *partialling out* process. We partial the nonparametric component out from the structural relation to obtain the linear reduced form and thus to estimate the parametric coefficient. During the partialling out process, we partial the nonparametric endogeneity out along with the nonparametric component from the structural relation. Although parametric endogeneity can conveniently be addressed with the conventional parametric approaches such as parametric instrumental variables (PIV) estimation (see chapter 16 of Li & Racine (2011) for details) and parametric two-stage least squares (P2SLS) estimation, it is not a trivial issue to construct a consistent parametric estimator due to the dominance of the parametric part in the models. Hence we systematically discuss endogeneity issues in a PL model in Chapter 2.

In the single-index semiparametrics, the presence of endogeneity causes more severe consequences. These are the inconsistency of the estimator of the index coefficient and a failure to identify the unknown structural function. The inconsistency of the index coefficient's estimator is caused by similar reasoning to that in the classical linear regression model. The minimising objective function with respect to the index coefficient does not produce the function which provides a consistent estimator of the index coefficient in the presence of endogeneity; see chapter 8 of Amemiya (1985) for details. We comprehensively discuss the endogeneity issues in SI semiparametrics in Chapter 3.

Let us consider a simple SI model which contains a nonparametric model as a special case to conveniently illustrate the effects of the presence of endogeneity as follows:

$$Y_i = g(V_i) + \epsilon_i, \tag{1.2.1}$$

where $V_i = X_i'\alpha$, (X, Y) is a $\mathbb{R}^q \times \mathbb{R}$-valued vector and $E(\epsilon|x) \neq 0$ implies that $E(\epsilon|v) \neq 0$. Then it is the case that $\hat{\alpha} \stackrel{p}{\nrightarrow} \alpha$ and $\hat{g}(v) \stackrel{p}{\nrightarrow} g(v)$, where $\stackrel{p}{\nrightarrow}$ denotes no convergence in probability, since $E(y|v) = g(v) + E(\epsilon|v)$.

Two commonly used alternatives to address endogeneity are a CF approach and NpIV estimation. The NpIV approach is based on a series approximation by Newey & Powell (2003) and a local constant kernel estimation by Härdle & Horowitz (2005) in a pure nonparametric model, with a series approximation by Ai & Chen (2003) in a semiparametric model containing a PL model, a GPLSI model and an EGPLSI model as special cases. However, the difficulty of conducting NpIV estimation is to overcome the ill-posed inverse problem (see O'Sullivan (1986) for details) in order to identify the structural relation from the reduced one obtained from the NpIV estimation. Härdle & Horowitz (2005) overcome the problem by a ridge-type regulation on the linear operator when estimating the reduced relation; meanwhile, Ai & Chen (2003) and Newey & Powell (2003) address it by regulating the inversion matrix and a constraining the space of the reduced relation to be compact. On the other hand, Newey et al. (1999), Pinkse (2000) and Su & Ullah (2008) consider the CF approach in a pure nonparametric model, while Blundell & Powell (2004) do so for a special case of a single-index model, i.e. only a case of discrete dependent variable is considered. With regard to the nonparametric estimation employed, Newey et al. (1999) and Pinkse (2000) rely on the series approximation, while Su & Ullah (2008) use the local polynomial estimation of Fan & Gijbels (1996). Blundell & Powell (2004), on the other hand, rely on local constant kernel estimation.

The CF approach is based on the widely-used nonparametric simultaneous equation framework, specifically a nonparametric triangular structure. Although full details are presented in Chapters 2 and 3, let us provide a brief discussion here. Consider the model in (1.2.1) again. In addition, there is a nonparametric reduced form equation as shown below:

$$X_i = m_x(Z_i) + \eta_i, \tag{1.2.2}$$

where Z is a $\mathbb{R}^{q_z}$-valued IV vector for X with $q_z \geq q$, $E(\eta|z) = 0$ and $E(\epsilon|z, \eta) = E(\epsilon|\eta) \neq 0$. In order to address endogeneity, we take the standard CF approach of Newey et al. (1999). That is, $E(y|v, \eta) = g(v) + E(\epsilon|\eta) = g(v) + \iota(\eta)$, where the endogeneity is controlled by introducing an additional unknown function $\iota(\eta)$. As this is a simple nonparametric additive structure, we can apply the marginal integration technique of Linton & Nielsen (1995) and Tjøstheim & Austad (1996) to identify each individual function. In addition, a consistent estimator of the index coefficient is also obtained. However, the difficulty of conducting the CF approach lies on the generated regressor issue, which arises because the endogeneity control variable is not observable but is instead generated from the reduced form (1.2.2) for the flexibility of the functional form (as in Newey et al. (1999)) in this thesis. Note that Li & Woodridge (2002) study the case of a parametrically generated regressor in a PL model, and they show that the parametric estimator is still $\sqrt{n}$-consistent and asymptotically normal.

More importantly, because of the importance of this topic, even though an effective tool is lacking for testing endogeneity in semiparametrics, an additional advantage of the CF approach is that it enables a rather simple procedure to be established for the purpose. This is brought about mainly by its ability to identify and disentangle the effect of endogeneity in the model. This simple tool relies on the variability bands being constructed over the estimates of the endogeneity measures (to be defined in Chapter 4) as the means of testing their statistical significance.

1.3 Review of the Empirical Engel Curves Literature

The study of Ernst Engel (1895) showed how the expenditure of a household on food varies with income level, which became known as an empirical Engel curve. It depicts an expansion path of a commodity demand as a household's expenditure increases. Hence it allows us to predict demand for the commodity and also to analyse the effects of an economic policy on consumption behaviour, such as welfare comparisons such as cost-of-living indices and equivalence scales, or effects of implementing a tax on consumption (see Banks et al. (1997), and Deaton & Muellbauer (1980*b*) for details). Hence, it is of great interest to all economic subjects to attain accurate empirical Engel functions.

Now let us begin the review of the literature of an empirical demand study with the most well-accepted and widely-used parametric specification, the Working-Leser specification as follows:

$$w_l = \alpha_l + y\beta_l = \epsilon_l, \tag{1.3.1}$$

where w_l is the budget share of good l and y is log of total expenditure. The Working-Leser specification, (1.3.1), is also the baseline model for the popular "Almost Ideal Demand System" and "Trans Log" models of consumer behaviour developed by Deaton & Muellbauer (1980*a*) and Jorgenson & Slesnick (1987), respectively. However, the study of Banks et al. (1997) suggests that empirical Engel curves at the micro-level rather than aggregate ones provide plausible results for the applications mentioned above. Banks et al. (1997) also provide the evidence of the severe nonlinear relationships of alcohol and adult clothing budget shares with total expenditure using the nonparametric approach, and this nonlinearity leads to a nonparametric approach that has been frequently employed in the more recent studies; see Blundell et al. (1998) and Blundell et al. (2007) for example. The nonparametric model of the empirical demand function of good l is:

$$w_{li} = g_l(y_i) + \epsilon_{li}, \tag{1.3.2}$$

where w_{li} is the budget share of good l for an individual household i, y_i is the log of household i's total expenditure, and $g_l(\cdot)$ is an unknown function. As has been observed from (1.3.2), the shape of an empirical Engel function is flexible enough to take any type of nonlinear relationship, unlike the parametric counterpart, (1.3.1).

Furthermore, demographic variables, for example, family size and regional differences, theoretically play a major role in the analysis of demand function. Empirical support for this can be found in Blundell et al. (1998) and Blundell et al. (2007), use the number of children in a household to differentiate household sizes in order to study the effect of family size on demand, and Gong et al. (2005), who studied regional effects on demand. Traditionally, the consumption behaviour of a household with a different demographic profile is specified by introducing *demographic translating* into the demand system; see Pollack & Wales (1995) for details. Following the traditional approach, the specification including demographic components is:

$$w_{li} = X_i'\beta_l + g_l(y_i) + \epsilon_{li}, \tag{1.3.3}$$

where X_i is a vector of the demographic components of household i. However, Blundell et al. (1998) show empirically that (1.3.3) restricts all budget shares to have a similar functional form. More recently, Blundell et al. (2003), theoretically demonstrated that the PL model is not coherent with economic theory, specifically the consumption optimisation theory. Demographic components cannot enter additively into each Engel curve equation while remaining consistency with consumer optimisation theory; they must also enter so as to scale the total expenditure variable inside nonparametric Engel curves for each commodity. Hence the shape-invariant type of specification (shown below as in (1.3.4)) provides the flexibility for individual functional forms not to be restricted within a similar functional form and is coherent with the economics theory.

$$w_{li} = X_i'\beta_l + g_l(y_i - \phi(X_i'\gamma)) + \epsilon_{li}, \tag{1.3.4}$$

where γ represents the equivalence scale coefficient.

However, it is well-known in the literature that the total expenditure is endogenous ($E(\epsilon_l|y) \neq 0$; e.g. Blundell et al. (1998), and Blundell et al. (2007)). The theoretical reason behind this is two-stage budgeting. The separability theorem of Gorman (1959) ultimately implies that budget decisions can be divided into two stages. In the first stage, total expenditure is divided into broader groups of spending such as saving vs expenditure and durable vs nondurable. Then in the second stage, individual budget shares are determined. This implies that error terms systematically contain the first budgeting decision stage. In other words, endogeneity in the empirical demand literature is caused by omitting information about the first budgeting stage. In Chapter 4, we present the details of how the methodologies developed in Chapters 2 and 3 are applied to a semiparametric analysis of Engel curves taking the effects of the endogeneity of the total expenditure into account in particular. Furthermore, we demonstrate how the EGPLSI model allows for the shape-invariant specification of (1.3.4) with minor modifications.

1.4 Research Objectives and Thesis Structure

In this thesis, we intend to provide two main contributions to the econometric literature. Firstly, we aim to introduce methods to address endogeneity in the estimation of the semiparametrics, particularly the PL model and the EGPLSI model. In particular, we aim to do this by establishing a CF approach based on (i) the Robinson (1988) and Speckman (1988) type of two-stage estimation procedure and (ii) the widely-used triangular structure of Newey et al. (1999), Pinkse (2000), Blundell & Powell (2004) and Su & Ullah (2008); see Chapters 2 and 3 for details. Secondly, we also intend to provide a further contribution to the economic literature, particularly on the cross-sectional relationships between expenditure on specific goods and the level of total expenditure. To achieve this objective, we employ our newly established methods to conduct a semiparametric analysis of Engel curves in Australia.

In Chapter 2, we comprehensively discuss endogeneity issues in a PL model. In particular, we discuss a systematic approach for addressing endogeneity in the PL model, given the complexity of the model containing both parametric and nonparametric components. We also address nonparametric endogeneity in the model using the CF approach and show that the estimator of the parametric coefficient is still $\sqrt{n}$-consistent and asymptotically normal. We also show that the unknown structural function can conveniently be recovered by using the marginal integration technique, and the estimator of the function is consistent and asymptotically normal. We further provide the results of Monte Carlo simulation exercises to provide the finite sample properties of those estimators.

In Chapter 3, we address endogeneity in an EGPLSI model containing the parametric coefficient and also the index coefficient based on the two-stage estimation procedure and the CF approach. We show that the estimators of both the parametric and the index coefficients are still $\sqrt{n}$-consistent and asymptotically normal. We also show that the unknown structural function is recovered by the marginal integration technique, and it is consistent and asymptotically normal. The finite sample properties of the estimators of all unknowns are also presented using Monte Carlo simulation exercises.

In Chapter 4, we conduct an empirical demand analysis in Australia with the methods developed in this thesis in order to overcome the endogeneity of total expenditure. Although both the PL and the EGPLSI models allow us to undertake micro-level analysis with demographic components (we use the number of children in a household to see the effects of household size on demand), our empirical results show that the PL model restricts the overall shapes of budgets shares to be a quadratic type of function, unlike the EGPLSI counterpart. This empirical study demonstrates the usefulness and practicality of the methods developed in this thesis and provides an accurate empirical demand analysis which is coherent with the economic theory.

We conclude this thesis with a discussion of the results as well as outlining future research directions in Chapter 5.

Chapter 2

Endogeneity in a PL Model

> *Statistical inference on a multidimensional random variable commonly focuses on functionals of its distribution that are either purely parametric or purely nonparametric. A reasonable parametric model affords precise inferences, a badly misspecified one, possibly seriously misleading ones, while nonparametric modeling is associated both greater robustness and lesser precision. An intermediate strategy employs a semiparametric form ...*
>
> Peter M. Robinson (1998)

2.1 Introduction

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

$$Y_i = X_i'\beta + g(V_i) + \epsilon_i \quad \text{for} \quad i = 1, \ldots, n$$
$$E(\epsilon|x, v) = 0, \tag{2.1.1}$$

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 (2.1.1), 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 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 (2.1.1) 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 & Racine (2011), for examples).

It should be noted that the exogeneity condition of the regressors in the model (as stated mathematically in the second line in (2.1.1)) 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 modelling 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 (2.1.1) is mistakenly replaced by the following:

$$Y_i = X_i'\beta + g(V_{1,i}) + \epsilon_i. \tag{2.1.2}$$

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:

$$X_i = m_x(V_i) + U_i, \tag{2.1.3}$$

where $E(u|v) = 0$. The corresponding *"parametric endogeneity"* may occur if the equation below is employed instead of (2.1.3):

$$X_i = m_x(V_{1,i}) + U_i, \tag{2.1.4}$$

where $E(u|v_1) = 0$. Given that the PL model in (2.1.1) is correctly specified, then the use of (2.1.2) and (2.1.4) 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:

$$\log(wage_i) = age_i\beta + g(edu_i) + H(agedu_i) + e_i, \tag{2.1.5}$$

where $agedu_i = age_i \times edu_i$ and $E(e|age, edu) = 0$. Although the identification of each function in (2.1.5) can be dealt with as shown by Linton & Nielsen (1995) and Tjøstheim & Austad (1996), 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 2.2.2.

The goal of the current chapter 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 & Li (1999), and Härdle et al. (2000) among others), the endogeneity issues have only recently been considered in pure nonparametric and semiparametric models (see Blundell & Powell (2003) for details). 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 chapter, 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 the key contributions of the current chapter 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 chapter.

In principle, the methods considered in this chapter 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 variables (PIV) estimation can be used. Although the PIV estimation has been considered in the literature (see chapter 16 of Li & Racine (2011), 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. Ai & Chen (2003) developed the NpIV estimation 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 & 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.

This chapter 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 & Nielsen (1995), and Tjøstheim & Austad (1996) to identify the unknown structural function.

To summarise, in this chapter 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 & 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 chapter is organised as follows. In Section 2.2, 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. In Section 2.3, we conduct an experimental study to investigate the finite sample properties of the estimators introduced in the current chapter. Finally, Section 2.4 concludes the chapter, while mathematical proofs of the main results are presented in Appendix 2.5.

2.2 Endogeneity in a PL 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 2.2.3 and 2.2.4.

2.2.1 The PL Model

To estimate the model in (2.1.1), 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 (2.1.1), since it satisfies the exogeneity condition:

$$g(v) = E(y|v) - E(x|v)'\beta. \tag{2.2.1}$$

Subtracting the conditional expectation relation (2.2.1) from the structural one (2.1.1) produces a simple linear reduced form:

$$W_i^* = U_i^{*\prime}\beta + \epsilon_i, \tag{2.2.2}$$

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 (2.2.2) immediately suggests an infeasible estimator for β by a LS estimation of W_i^* on U_i^*:

$$\overline{\beta}_{LS}^* = \left(\bar{S}_{U^*}\right)^{-1} \bar{S}_{U^*W^*}, \tag{2.2.3}$$

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 (2.2.1):

$$\overline{g}^*_{LS}(v) = m^*_y(v) - m^*_x(v)'\overline{\beta}^*_{LS}. \tag{2.2.4}$$

Nonetheless, the estimators in (2.2.3) and therefore in (2.2.4) 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 & 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 (2.1.1) can be consistently estimated at $\sqrt{n}$-rate; see Robinson (1988) and Fan et al. (1995), for example.

Remark 2.2.1. *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 2.2.4 in the context of the estimation method introduced in this chapter.* ■

2.2.2 Endogeneity in the PL Model

Prior to presenting a more detailed discussion of the method studied in this chapter, 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 (3.2.18):

$$W_i^* = U_i^{*\prime}\beta + e_i^*, \tag{2.2.5}$$

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 (2.2.5), 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 no 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 2.2.2. *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 (2.2.5), since we discuss the endogeneity issues in terms of infeasible estimators such as (2.2.3) and (2.2.4). Note that the cost of having the former rather than the latter is 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: 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)*

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

2.2.3 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:

$$W_i^* = U_i^{*\prime}\beta + e_i^*, \tag{2.2.6}$$

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:

$$\mathcal{Z}_i = m^*_{\mathcal{Z}}(V_i) + \varrho_i^*, \tag{2.2.7}$$

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 2.2.2; $E(\varrho^*\epsilon) = 0$ implies that $E(\varrho^* e^*) = 0$. In this case, we replace $E(y)^2 < \infty$ with $E(y)^4 < \infty$ in Assumption 2.0.4 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 & Racine (2011), 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 least squares (2SLS) estimation is the most optimal, as in a conventional parametric case; see chapter 5 of Sargan & Desai (1988), 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 2.2.3. *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 2.5. 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 2.5.1 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

$$H_0 : \hat{d}^* \xrightarrow{p} 0 \text{ i.e. } \textit{parametric exogeneity}; \; H_1 : \hat{d}^* \not\xrightarrow{p} 0 \text{ i.e. } \textit{parametric endogeneity}. \tag{2.2.8}$$

Under the null hypothesis in (2.2.8), 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 (2.2.8), 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.

2.2.4 Nonparametric Endogeneity

In Section 2.2.2, 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:

$$\begin{aligned} Y_i &= X_i'\beta + g(V_i) + \epsilon_i && (2.2.9) \\ V_i &= m_v(Z_i) + \eta_i, && (2.2.10) \\ E(\epsilon|z,\eta) &= E(\epsilon|\eta) \text{ a.s.} && (2.2.11) \\ E(\eta|z) &= 0 \text{ a.s.} && (2.2.12) \end{aligned}$$

where a.s. denotes for almost surely, (2.2.9) is as defined in (2.1.1), $m_v(z) \equiv E(v|z)$ 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 (2.2.11) and (2.2.12), 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 (2.2.9) to (2.2.12), 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:

$$g(v) + \iota(\eta) = E(y|v,\eta) - E(x|v,\eta)'\beta, \qquad (2.2.13)$$

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 chapter is to estimate the endogeneity control regressors from a structural relation between the endogenous regressors and their instrumental variables, (i.e. expression (2.2.10) above). Such a structural relation is referred to as "a reduced form" in Blundell & Powell (2004). The next step is to control the endogeneity in the structural relation (2.2.9) by introducing an endogeneity control function, $\iota(\eta)$. Finally, the nonparametric additive structure derived in (2.2.13) suggests that the unknown structural function can be identified by using the marginal integration technique of Linton & Nielsen (1995) and Tjøstheim & 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 2.2.1: Estimate the endogeneity control regressor, η_i, from (2.2.10).

Step 2.2.2: Obtain consistent estimators $\hat{\beta}_\tau$ of the unknown parameters as in Section 2.2.3.

Step 2.2.3: Given the consistent parametric estimators in Step 2.2.2, estimate the conditional expectation relation.

Step 2.2.4: Perform the marginal integration technique on the resulting estimated conditional expectation relation in Step 2.2.3 to estimate the structural $g(\cdot)$-function.

In the remainder of this section, let us discuss each of these steps in more detail. *Step 2.2.1* estimates the endogeneity control regressors, η, from the reduced form in (2.2.10) 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:

$$\hat{m}_{v,l}(Z_i) = \frac{\sum_{j=1}^{n} V_j K_z\left(\frac{Z_i - Z_j}{h_z}\right)}{\sum_{l=1}^{n} K_z\left(\frac{Z_i - Z_l}{h_z}\right)}, \tag{2.2.14}$$

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

$$\hat{\eta}_i = V_i - \hat{m}_{v,l}(Z_i). \tag{2.2.15}$$

For a constant $b_1 > 0$, let $I_{1,i} = I(|\hat{f}(Z_i)| > b_1)$, where $\hat{f}(z)$ is the estimate of the probability density function $f(z)$ with a random argument Z_i.

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:

$$m_y(v,\eta) = m_x(v,\eta)'\beta + g(v) + \iota(\eta) \tag{2.2.16}$$

such that $\iota(\eta) \neq 0$ controls the endogeneity. Finally, if we subtract the conditional expectation relation (3.2.10) from the structural one, the transformed simple linear reduced form is then:

$$W_i = U_i'\beta + e_i, \tag{2.2.17}$$

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 2.2.3) is applied to (2.2.17). 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:

$$\hat{\beta}_{LS} = (S_{\hat{U}_2})^{-1} S_{\hat{U}_2 \hat{W}_2},\ \hat{\beta}_{IV} = (S_{\hat{\varrho}_2 \hat{U}_2})^{-1} S_{\hat{\varrho}_2 \hat{W}_2} \ \text{ and } \ \hat{\beta}_{2SLS} = (S_{\hat{\hat{\varrho}}_2 \hat{U}_2})^{-1} S_{\hat{\hat{\varrho}}_2 \hat{W}_2}, \tag{2.2.18}$$

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.:

$$\hat{E}(x|v,\hat{\eta}) = \frac{\sum_{i=1}^n X_i K_v\left(\frac{v-V_i}{h_v}\right) K_\eta\left(\frac{\hat{\eta}-\hat{\eta}_i}{h_\eta}\right)}{\sum_{j=1}^n K_v\left(\frac{v-V_j}{h_v}\right) K_\eta\left(\frac{\hat{\eta}-\hat{\eta}_j}{h_\eta}\right)}, \tag{2.2.19}$$

$$\hat{E}(y|v,\hat{\eta}) = \frac{\sum_{i=1}^n Y_i K_v\left(\frac{v-V_i}{h_v}\right) K_\eta\left(\frac{\hat{\eta}-\hat{\eta}_i}{h_\eta}\right)}{\sum_{j=1}^n K_v\left(\frac{v-V_j}{h_v}\right) K_\eta\left(\frac{\hat{\eta}-\hat{\eta}_j}{h_\eta}\right)} \tag{2.2.20}$$

and:

$$\hat{E}(\mathcal{Z}|v,\hat{\eta}) = \frac{\sum_{i=1}^n \mathcal{Z}_i K_v\left(\frac{v-V_i}{h_v}\right) K_\eta\left(\frac{\hat{\eta}-\hat{\eta}_i}{h_\eta}\right)}{\sum_{j=1}^n K_v\left(\frac{v-V_j}{h_v}\right) K_\eta\left(\frac{\hat{\eta}-\hat{\eta}_j}{h_\eta}\right)}. \tag{2.2.21}$$

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 2.2.1 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 rth 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* $l - 1 \leq \mu \leq l$; *for some* $\rho > 0$, $\sup_{y \in \phi_{z\rho}} |g(y) - g(z) - Q_g(y,z)|/||y - z||^\mu \leq G_g(z)$ *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 2.2.1. *Under Assumptions 2.1.1-2.1.6, the condition that* Φ_U *is positive and definite is necessary and sufficient for:*

$$\begin{aligned} \sqrt{n}(\hat{\beta}_{LS} - \beta) &\overset{D}{\to} N[0, \Phi_U^{-1}\sigma^2] \\ S_{\hat{U}_2}^{-1}\hat{\sigma}^2 &\overset{p}{\to} \Phi_U^{-1}\sigma^2, \end{aligned} \tag{2.2.22}$$

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

Corollary 2.2.1. *Under Assumptions 2.1.1, 2.1.3 - 2.1.7, 2.1.8 and 2.1.9, the conditions that* $\Phi_{\varrho U}, \Phi_\varrho$ *and* $\Phi_{U\varrho}$ *are positive and definite are necessary and sufficient for:*

$$\begin{aligned} \sqrt{n}(\hat{\beta}_{IV} - \beta) &\overset{D}{\to} N[0, (\Phi_{\varrho U})^{-1}\sigma^2\Phi_\varrho(\Phi_{U\varrho})^{-1}] \\ (S_{\hat{\varrho}_2\hat{U}_2})^{-1}\hat{\sigma}^2 S_{\hat{\varrho}_2}(S_{\hat{U}_2\hat{\varrho}_2})^{-1} &\overset{p}{\to} (\Phi_{\varrho U})^{-1}\sigma^2\Phi_\varrho(\Phi_{U\varrho})^{-1}. \end{aligned} \tag{2.2.23}$$

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

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

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 2.2.4. *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 $S_{\mathcal{Z}-\hat{\mathcal{Z}},X-\hat{X}} \xrightarrow{p} \Phi_{\varrho U}$ 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 2.5 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 mis-specification 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:

$$\sqrt{n}(\hat{d} - d) \xrightarrow{D} N[0, V(\hat{d})], \tag{2.2.24}$$

where $V(\hat{d}) = V(\hat{\beta}_{IV}) - V(\hat{\beta}_{LS})$. These asymptotic variances are given above as (2.2.22) and (2.2.23), respectively. As $n \to \infty$, the test statistic is:

$$t = \hat{d}' \left(\hat{V}(\hat{d})\right)^{-1} \hat{d} \to_D \chi^2_p, \tag{2.2.25}$$

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 2.2.3 and 2.2.4, 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 (3.2.10):

$$m(v,\eta) = m_y(v,\eta) - m_x(v,\eta)'\beta = g(v) + \iota(\eta). \tag{2.2.26}$$

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 & Tibishirani (1991), and Gao (2007) for example. Since (2.2.26) 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.:

$$m(v) = \int m(v,\eta)dQ(\eta) = g(v) + c_1 \text{ and } m(\eta) = \int m(v,\eta)dQ(v) = \iota(\eta) + c_2, \tag{2.2.27}$$

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 & 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 (2.2.27):

$$\hat{m}_\tau(v) = \frac{1}{n}\sum_{i=1}^{n} \hat{m}_\tau(v, \hat{\eta}_i) \tag{2.2.28}$$

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 (2.2.28) 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 2.2.2. *Under Assumptions 2.1.1 - 2.1.6 when the parametric regressors are exogenous, or else, under Assumptions 2.1.1, 2.1.3 - 2.1.7, 2.1.8 and 2.1.9, 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* $\mathcal{K}_v = \int K(v)^2 dv$. ■

The proofs of Theorems 2.2.1 and 2.2.2 are given in the Appendix 2.5.

2.3 Simulations

In this section, we discuss Monte Carlo simulation exercises 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 2.2.1) 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. 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.

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 2.2.3 with the LS and IV estimators of the unknown parameter, respectively. Furthermore, *2SCF* refers to the control function approach as explained in Section 2.2.4. 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 2.2 to 2.10.

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

Table 2.2: Exogenous model with 2SR

n	$\hat{\beta}$	Bias	Var	$\lvert\hat{\beta}-\beta\rvert$	$\text{ae}_{\hat{g}}$	$\hat{h}_v$
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

Nonparametric endogeneity

For the sake of comparisons, we will employ the PL model in (2.3.1) as a baseline model:

$$\begin{aligned} Y_i = 1.2X_i + 0.5\left(\frac{V_i}{1+V_i^2}\right) + \epsilon_i, \text{ where} \\ V_i = Z_i + \eta_i, \quad X_i = \sin(V_i - \eta_i) + U_i, \\ \epsilon_i = \iota(\eta_i) + e_i, \\ Z_i \sim \mathcal{U}(0,3), \ \eta_i \sim \mathcal{U}(-1,1), \ e_i, U_i \sim N(0,1). \end{aligned} \tag{2.3.1}$$

Defining the PL model as in (2.3.1) 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:

$$\iota(\eta) = 0 \times \eta, \ \iota(\eta) = 1 \times \eta \text{ and } \iota(\eta) = \frac{\eta}{1+\eta^2}, \tag{2.3.2}$$

respectively. While Tables 2.2, 2.3 and 2.4 present the estimation results of the exogenous model, the linear endogeneity model and the nonlinear endogeneity model, respectively, based on the 2SR procedure, Tables 2.5 and 2.6 summarise those obtained based on the 2SCF procedure.

Table 2.3: Linear Endogeneity model with 2SR

n	$\hat{\beta}$	Bias	Var	$\lvert\hat{\beta}-\beta\rvert$	$ae_{\hat{g}}$	$\hat{h}_v$
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 2.4: Nonlinear Endogeneity model with 2SR

n	$\hat{\beta}$	Bias	Var	$\lvert\hat{\beta}-\beta\rvert$	$ae_{\hat{g}}$	$\hat{h}_v$
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

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 2.3 and 2.4 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

Table 2.5: Linear Endogeneity model with 2SCF

n	$\hat{\beta}$	Bias	Var	$\lvert\hat{\beta} - \beta\rvert$	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 2.6: 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

Table 2.7: Linear parametric endogeneity model with 2SR

n	$\hat{\beta}$	Bias	Var	$\lvert\hat{\beta}-\beta\rvert$	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

2.3 and 2.4, those in Tables 2.5 and 2.6 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:

$$Y_i = 1.2X_i + 0.5\left(\frac{V_i}{1+V_i^2}\right) + \epsilon_i, \text{ where} \tag{2.3.3}$$
$$X_i = \mathcal{Z}_i + \eta_i,\ \mathcal{Z}_i = \sin(V_i) + \varrho_i,\ \epsilon_i = \iota(\eta)_i + e_i,$$
$$\eta_i \sim \mathcal{U}(-1,1,), V_i \sim \mathcal{U}(0,3) \text{ and } e_i \sim N(0,1).$$

Clearly, the model in (2.3.3) implies that $E(\epsilon|x) \neq 0$, $E(\epsilon|v) = 0$, $E(\mathcal{Z}x) \neq 0$ and $E(\varrho\epsilon) = 0$. Tables 2.7 and 2.8 present the estimation results of the linear parametric endogeneity model and the nonlinear parametric endogeneity model respectively, based on the *2SR* procedure; Tables 2.9 and 2.10 provide those based on the *2SR-PIV* procedure.

Let us now discuss some important findings as follows. When compared to the estimation results of an exogenous model in Table 2.2, those in Tables 2.7 and 2.8 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

Table 2.8: Nonlinear parametric endogeneity model with 2SR

n	$\hat{\beta}$	Bias	Var	$\|\hat{\beta}-\beta\|$	$ae_{\hat{g}}$	$\hat{h}_v$
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 2.9: Linear parametric endogeneity model with 2SR-PIV

n	$\hat{\beta}$	Bias	Var	$\|\hat{\beta}-\beta\|$	$ae_{\hat{g}}$	$\hat{h}_v$
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 2.10: Nonlinear parametric endogeneity model with 2SR-PIV

n	$\hat{\beta}$	Bias	Var	$\|\hat{\beta}-\beta\|$	$ae_{\hat{g}}$	$\hat{h}_v$
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

parametric unknown parameter and the nonparametric structural function using the *2SR* procedure. Compared to the results shown in Tables 2.7 and 2.8, those in Tables 2.5 and 2.6 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.

2.4 Conclusions

In this chapter, 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 2.2.2 and 2.2.3. 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 chapter. 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. 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.

2.5 Appendix

In this Appendix, we present detailed discussion of the theoretical analysis of the main results of the current chapter. We firstly list two sets of conditions which are required for the case without and with the presence of nonparametric endogeneity, respectively. The rest of this Appendix presents the mathematical proofs of Theorems 2.2.1 and 2.2.2.

2.5.1 Conditions for the PIV estimator

We state a set of conditions for the PIV and P2SLS estimators 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 2.0.1. *$(V_i, X_i, Y_i, \mathcal{Z}_i)$, $i = 1, 2, \ldots, n$ are i.i.d. observations.*

Assumption 2.0.2. *$E(\epsilon|x, v) \neq 0$ and $E(\epsilon|\mathcal{Z}, v) = 0$.*

Assumption 2.0.3. *$E(\epsilon^2|v, \mathcal{Z}) = \sigma^2(v, \mathcal{Z})$ is continuous in $(v, \mathcal{Z})$.*

Assumption 2.0.4. *All Y_i have a finite second moment, and all X_i and $\mathcal{Z}_i$ have finite fourth moments.*

Assumption 2.0.5. *V_i admits a density function $f \in \mathcal{G}_{\lambda}^{\infty}$ for some $\lambda > 0$.*

Assumption 2.0.6.

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

Assumption 2.0.7. *As* $n \to \infty$,

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

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

(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 2.0.8. $\sup_{v\in\mathbb{R}^q}|K(v)| + \int |v^p K(v)|dv < \infty$ *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 2.0.9. $E(\epsilon|v, x) = 0$ *and* $E(\epsilon^2|v, x)$ *is continuous in* (v, x).

Assumption 2.0.10. *As* $n \to \infty$,

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

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

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

(4) $b \to 0$.

Assumption 2.0.2 indicates that the model suffers from parametric endogeneity. If this is not the case then we consider Assumption 2.0.9 instead. Furthermore, if endogeneity is present, then we impose the condition in Assumption 2.0.7 on the bandwidth parameter; otherwise, Assumption 2.0.10 is used. Assumption 2.0.4 provides the moment conditions on the regressors. Assumptions 2.0.5 and 2.0.6 collectively provide the moment bounds and the smoothness of the density and regression functions. By Assumption 2.0.8, the kernel function is bounded, integrable and high-order. Assumptions 2.0.7 and 2.0.8 should be satisfied simultaneously (see Robinson (1988), for example) in the case of parametric endogeneity, while Assumptions 2.0.8 and 2.0.10 are used instead for the case where there is no parametric endogeneity.

2.5.2 Conditions for Theorems 2.2.1 and 2.2.2

Assumption 2.1.1.

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

Assumption 2.1.2.

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

Assumption 2.1.3.

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

Assumption 2.1.4.

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

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

Assumption 2.1.5.

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

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

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

Assumption 2.1.6. *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 2.1.7.

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

(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 2.1.8.

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

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

Assumption 2.1.9. *As* $n\to\infty$*,*

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

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

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

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

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

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

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

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

Assumption 2.1.2 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 2.1.3 with Assumption 2.0.4 in Section 2.5.1. Assumptions 2.1.4 and 2.1.5 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 2.0.5 and 2.0.6 in Section 2.5.1. Given the higher-order kernel function in Assumption 2.1.7, the bias is sufficiently decreased with Assumptions 2.1.4 and 2.1.5. Assumption 2.1.7 states that the kernel functions used in this paper are bounded, integrable and high-order.

Note that Assumptions 2.1.6 and 2.1.7 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 $nh_z^{12} \to 0$, $nh_v^6 h_\eta^6 \to 0$, 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 2.1.6 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 2.1.8 and 2.1.9 are for the case of the presence of both parametric endogeneity and nonparametric endogeneity in the model. In particular, Assumption 2.1.8 (1) states that the model suffers from parametric endogeneity as well. Assumption 2.1.8 (2) states the smoothness of the function $m_{\mathcal{Z}}(v, \eta)$ that sufficiently reduces the bias with Assumption 2.1.7.

2.5.3 Proof of Theorem 2.2.1

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

$$Y_i - \hat{Y}_{2,i} = (X_i - \hat{X}_{2,i})'\beta + (m_i - \hat{m}_{2,i}) + (e_i - \hat{e}_{2,i}), \tag{2.A.1}$$

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, (2.A.1) is rewritten below:

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

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 (2.A.1) and (2.A.2), we have:

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

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:

$$\hat{\delta}_i = \hat{\delta}_{2,i} - \hat{\delta}_{1,i} = \frac{1}{nh_v^q h_\eta^q}\sum_{j=1}^n \delta_j(\hat{w}_{ij} - w_{ij}), \tag{2.A.3}$$

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:

$$\frac{1}{nh_v^q h_\eta^q}\sum_{l=1}^n K_v\left(\frac{V_i-V_l}{h_v}\right)K_\eta\left(\frac{\hat{\eta}_i-\hat{\eta}_l}{h_\eta}\right) = A_{0,i} + \sum_{r=1}^{\omega_2-1}\frac{1}{r!}A_{r,i}\triangle_{il}^r + R_{il}, \tag{2.A.4}$$

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 (2.A.4) are:

$$A_{1,1,i} = \{m_v(Z_i)-\hat{m}_v(Z_i)\}\,\hat{f}_\eta^{(1)}(V_i,\eta_i) \tag{2.A.5}$$

$$\begin{aligned} A_{1,2,i} &= \frac{1}{nh_v^q h_\eta^{q+1}}\sum_{l=1}^n K_v\left(\frac{V_i-V_l}{h_v}\right)K_\eta^{(1)}\left(\frac{\eta_i-\eta_l}{h_\eta}\right) \\ &\times \{m_v(Z_l)-\hat{m}_v(Z_l)\}. \end{aligned} \tag{2.A.6}$$

We firstly consider (2.A.5) by the boundedness condition on the kernel function and the smoothness on the function $m_v(z)$:

$$E\left(\{m_v(Z_i)-\hat{m}_v(Z_i)\}\right)^2 \leq (nh_z^{q_z}b_1)^{-2}E(T_z)^2 = O(n^{-1}h_z^{-q_z}b_1^{-2}+h_z^{2\xi_1}b_1^{-2}), \quad (2.A.7)$$

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

$$A_{1,1,i} = O_p\left(n^{-1}h_v^{-q/2}h_\eta^{-q/2-1}h_z^{-q_z/2}b_1^{-1} + n^{-1/2}h_v^{-q/2}h_\eta^{-q/2-1}h_z^{\xi_1}b_1^{-1}\right). \quad (2.A.8)$$

By the $i.i.d.$ assumption, we have the second moment bound of (2.A.6) as follows:

$$E(A_{1,2,i})^2 \leq \left(n^2h_v^q h_\eta^{q+1}h_z^{q_z}b_1^1\right)^{-2}\left[E\left\{\sum_{l=1}^n\left(L_{1l}^{(1)}\right)^2T_z^2\right\} + E\left|\sum_{l=1}^n\sum_{j\neq l}^n\left(L_{1l}^{(1)}\right)\left(L_{1j}^{(1)}\right)T_z^2\right|\right]. \quad (2.A.9)$$

The first term in (2.A.9) is:

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

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:

$$\begin{aligned}
E\left\{\left(L_{1l}^{(1)}\right)^2t_{z,2}^2\right\} &\leq \left[E\left\{E_l\left(L_{1l}^{(1)}\right)^4\right\}E(t_{z,2}^4)\right]^{1/2} \\
&= O(h_v^q h_\eta^q h_z^{q_z})^{1/2} \quad (2.A.10)\\
E\left\{\left(L_{1l}^{(1)}\right)^2T_{z,2}^2\right\} &\leq \left[E\left\{E_l\left(L_{1l}^{(1)}\right)^2\right\}E\left\{E_l\left(L_{1l}^{(1)}\right)^2T_{z,2}^4\right\}\right]^{1/2} \\
&= O\left(n^{1/2}h_v^q h_\eta^q h_z^{q_z/2} + n^2h_v^q h_\eta^q h_z^{2(q_z+\xi_1)}\right), \quad (2.A.11)
\end{aligned}$$

where $E(T_z)^4 = O\left(nh_z^{q_z}+n^4h_z^{4(q_z+\xi_1)}\right)$ by the similar argument as in (2.A.7). Hence the first term on the right-hand side in (2.A.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 (2.A.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

$$\begin{aligned}
E\left(\left(L_{1l}^{(1)}\right)^2 T_{z,2}^2\right) &= O\left(nh_v^q h_\eta^q h_z^{q_z} + n^2 h_v^q h_\eta^q h_z^{2(q_z+\xi_1)}\right)\\
E\left(\left|L_{1l}^{(1)}L_{1j}^{(1)}\right| t_{z,\imath}^2\right) &\leq \left[E\left\{E_l\left(L_{1l}^{(1)}\right)^2\right\}E\left\{E_j\left(L_{1j}^{(1)}\right)^2 t_{z,\imath}^4\right\}\right]^{1/2}\\
&= O(h_v^q h_\eta^q h_z^{q_z/2}) \qquad (2.A.12)\\
E\left(\left|L_{1l}^{(1)}L_{1j}^{(1)}\right| T_{z,2}^2\right) &\leq \left[E\left\{E_l\left(L_{1l}^{(1)}\right)E_j\left(L_{1j}^{(1)}\right)\right\}E\left\{E_l\left(L_{1l}^{(1)}\right)E_j\left(L_{1j}^{(1)}\right)T_{z,2}^4\right\}\right]^{1/2}\\
&= O\left(n^{1/2}h_v^{2q}h_\eta^{2q}h_z^{q_z/2} + n^2 h_v^{2q}h_\eta^{2q}h_z^{2(q_z+\xi_1)}\right), \qquad (2.A.13)
\end{aligned}$$

where $\imath = 2$ or 3. The second term in (2.A.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:

$$A_{1,2,i} = O_p\left(n^{-1}h_v^{-q/2}h_\eta^{-q/2-1}h_z^{-q_z/4}b_1^{-1} + h_\eta^{-1}h_z^{\xi_1}b_1^{-1}\right). \qquad (2.A.14)$$

We obtain the following results:

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

and:

$$\hat{\delta}_i = \frac{\left\{B_{1,i}\triangle_{ij} + \sum_{r'=2}^{\omega_2-1}\frac{1}{r'!}B_{r',i}\triangle_{ij}^{r'} + R_{ij}\right\}}{\hat{f}(V_i,\eta_i)}, \qquad (2.A.15)$$

where

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

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 (2.A.8) and (2.A.14), and the Taylor series expansion of the kernel function. Hence we have:

$$\begin{aligned}\hat{\delta}_i &= \frac{\frac{1}{nh_v^q h_\eta^{q+1}}\sum_{j=1}^n \delta_j K_v\left(\frac{V_i-V_j}{h_v}\right) K_\eta^{(1)}\left(\frac{\eta_i-\eta_j}{h_\eta}\right)\{m_v(Z_j)-\hat{m}_v(Z_j)\}}{\hat{f}(V_i,\eta_i)} \\ &+ O_p\left(n^{-1}h_v^{-q/2}h_\eta^{-q/2-1}h_z^{-q_z/2}b_1^{-1}+n^{-1/2}h_v^{-q/2}h_\eta^{-q/2-1}h_z^{\xi_1}b_1^{-1}\right) \\ &+ o_p(1), \qquad (2.A.16)\end{aligned}$$

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 boundedness condition on the kernel function.

Using (2.A.16),

$$\begin{aligned}E|S_{\hat{\delta}}| &\leq (nh_z^{q_z})^{-2}E\left\{\frac{1}{n}\sum_{i=1}^n|\breve{\delta}_i|^2 I_{2,i}T_z^2 I_1\right\} \qquad (2.A.17)\\ &+ (nh_z^{q_z})^{-2}\left|E\left\{\frac{1}{n}\sum_{i=1}^n\sum_{j\neq i}^n\breve{\delta}_i\breve{\delta}_j' I_{2,i}I_{2,j}'T_z^2 I_1\right\}\right|, \qquad (2.A.18)\end{aligned}$$

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 $E(|\breve{\delta}_1|^2 I_{2,1}|\mathcal{L}_n) \leq (nh_v^q h_\eta^{q+1}b_2)^{-2}E\left(\sum_{i=1}^n|\delta_i|^2\left(L_{1i}^{(1)}\right)^2|\mathcal{L}_n\right)$, a.s., the right hand side of (2.A.17) is bounded by $(n^2h_v^q h_\eta^{q+1}h_z^{q_z}b_1b_2)^{-2}$ multiplies by:

$$E\left(\sum_{i=1}^n|\delta_i|^2\left(L_{1i}^{(1)}\right)^2T_z^2\right) \leq \mathcal{C}E\left(|\delta_1|^2T_z^2+n|\delta_2|^2\left(L_{12}^{(1)}\right)^2t_{z,2}^2+n|\delta_2|^2\left(L_{12}^{(1)}\right)^2T_{z,1}^2\right), \qquad (2.A.19)$$

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 (2.A.19). By the Cauchy inequality and the similar argument as in (2.A.7):

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

Similarly, as in (2.A.10) and (2.A.11) with the moment restrictions on δ_i , the other two terms in (2.A.19) are:

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

and:

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

Thus (2.A.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 (2.A.18):

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

Therefore, (2.A.18) is bounded by:

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

where $T_{z,2} = T_{z,1} - t_{z,3}$. Similar to the procedure in (2.A.10) and (2.A.11), for $\imath = 1$ or 2, we have:

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

and:

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

As in (2.A.12) and (2.A.13) with the bounded moment restriction on δ_i, for $\imath = 2$ or 3, we have:

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

and:

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

Thus (2.A.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:

$$E(\sqrt{n}S_{\varepsilon\hat{\delta}})^2 \leq E\left\{\frac{1}{n}\sum_{i=1}^{n}\breve{\delta}_i^2 I_{2,i}I_1|\varepsilon_1|^2\right\} \quad (2.A.20)$$

$$+ \left|E\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\breve{\delta}_i\breve{\delta}_j I_{2,i}I_{2,j}I_1|\varepsilon_1|^2\right\}\right|. \quad (2.A.21)$$

Because $E(|\breve{\delta}_1|^2 I_{2,1}|\mathcal{L}_n) \leq (nh_v^q h_\eta^{q+1}b_2)^{-2}E\left(\sum_{i=1}^n |\delta_i|^2\left(L_{1i}^{(1)}\right)^2|\mathcal{L}\right)$ *a.s.*, the right-hand side of (2.A.20) is bounded by $(n^2h_v^q h_\eta^{q+1}h_z^{q_z}b_1b_2)^{-2}$ multiplies by:

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

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

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

By the similar argument as in (2.A.7), we have:

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

Hence (2.A.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 (2.A.21):

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

(2.A.21) is therefore bounded by:

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

Similar to (2.A.10) and (2.A.11), and with the bounded moment conditions on δ_i, we have:

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

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

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

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:

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

By the Cauchy inequality, the bound condition on the kernel function and the similar argument as in (2.A.7):

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

Thus (2.A.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.

$$E(\sqrt{n}S_{\hat{\varepsilon}\hat{\delta}})^2 \leq E\left\{\frac{1}{n}\sum_{i=1}^{n}|\hat{\varepsilon}_i|^2 I_{2,i}\breve{\delta}_i^2 I_{2,i}I_{1,i}\right\} \quad (2.A.22)$$

$$+\left|E\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\hat{\varepsilon}_i\hat{\varepsilon}'_j\breve{\delta}_i\breve{\delta}_j I_{1,i}I_{1,j}I_{2,i}I'_{2,j}I'_{2,j}I_{2,i}\right\}\right|, \quad (2.A.23)$$

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

$$E(|\hat{\varepsilon}_1|^2I_{2,1}|\mathcal{L}_n)\leq(nh_v^qh_\eta^qb_2)^{-2}E\left(\sum_{i=1}^{n}|\varepsilon|^2L_{1i}^2|\mathcal{L}_n\right)\ a.s.,$$

the right-hand side of (2.A.22) is bounded by $(n^3h_v^{2q}h_\eta^{2q+1}h_z^{q_z}b_1b_2^2)^{-2}$ multiplies by

$$E\left(\sum_{i=1}^{n}|\varepsilon_i|^2L_{1i}^2\sum_{j=1}^{n}|\delta_j|^2\left(L_{1j}^{(1)}\right)^2T_z^2\right)\leq\left\{E\left(\sum_{i=1}^{n}|\varepsilon_i|^4L_{1i}^4\sum_{j=1}^{n}|\delta_j|^4\left(L_{1j}^{(1)}\right)^4\right)E(T_z^4)\right\}^{1/2}, \quad (2.A.24)$$

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:

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

Hence (2.A.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 (2.A.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 (2.A.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 (2.A.23) is bounded by:

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

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

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

Hence (2.A.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 $\hat{m}_i$.

$$E\left(\sqrt{n}S_{\varphi-\hat{\varphi},\hat{\delta}}\right)^2 \leq E\left\{\frac{1}{n}\sum_{i=1}^{n}(\varphi_i-\hat{\varphi}_i)^2\breve{\delta}_i^2 I_{2,i}I'_{2,i}I_{1,i}\right\} \quad (2.A.25)$$

$$+\left|E\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}^{n}(\varphi_i-\hat{\varphi}_i)(\varphi_j-\hat{\varphi}_j)' I_{2,i}I'_{2,j}I'_{2,j}I_{2,i}\breve{\delta}_i\breve{\delta}_j I_{1,i}I_{1,j}\right\}\right|. \quad (2.A.26)$$

The right-hand side of (2.A.25) is bounded by $(n^3 h_v^{2q} h_\eta^{2q+1} h_z^{q_z} b_1 b_2^2)^{-2}$ multiplies by

$$E\left\{\sum_{i=1}^{n}(\varphi_1-\varphi_i)^2 L_{1i}^2 \sum_{j=1}^{n}(m_{v,1}-m_{v,j})^2 K_{z,1j}^2 \sum_{l=1}^{n}|\delta_l|^2\left(L_{1l}^{(1)}\right)^2\right\}$$

$$\leq \left[E\left\{\sum_{i=1}^{n}(\varphi_1-\varphi_i)^2 L_{1i}^2 \sum_{j=1}^{n}(m_{v,1}-m_{v,j})^2 K_{z,1j}^2\right\}^2 E\left\{\sum_{l=1}^{n}|\delta_l|^2\left(L_{1l}^{(1)}\right)^2\right\}^2\right]^{1/2}$$

$$\leq \left[\left\{(E(T_z^8)E(T_\varphi^8))\right\}^{1/2} E\left\{\sum_{l=1}^{n}|\delta_l|^2\left(L_{1l}^{(1)}\right)^2\right\}^2\right]^{1/2}, \quad (2.A.27)$$

by the Cauchy inequality. By the similar argument as in (2.A.7), we have:

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

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 (2.A.27) is:

$$E\left\{\sum_{l=1}^{n}|\delta_l|^2\left(L_{1l}^{(1)}\right)^2\right\}^2 \leq \left(L_{11}^{(1)}\right)^4 E|\delta|^4 + (n-1)E\left(|\delta_2|^4\left(L_{12}^{(1)}\right)^4\right)$$

$$+(n^2-1)E\left\{|\delta_2|^2|\delta_3|^2\left(L_{12}^{(1)}L_{13}^{(1)}\right)^2\right\}$$

$$\leq \mathcal{C}E|\delta|^4 + nE\left\{|\delta_2|^4 E_2\left(L_{12}^{(1)}\right)^4\right\} + n^2 E\left\{|\delta_2|^2|\delta_3|^2 E_1\left(L_{12}^{(1)}\right)^2 E_3\left(L_{13}^{(1)}\right)^2\right\}$$

$$\leq \mathcal{C}\left(1+nh_v^q h_\eta^q + n^2 h_v^{2q} h_\eta^{2q}\right)E|\delta|^4.$$

Hence (2.A.27) is:

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

The right-hand side of (2.A.25) equals:

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

Next, we consider (2.A.26). Since we already know that:

$$E(\breve{\delta}_1\breve{\delta}_2 I_{2,1}I_{2,2}|\mathcal{L}) \leq \mathcal{C}(nh_v^q h_\eta^{q+1}b_2)^{-2}E\left(\sum_{i=1}^n |\delta_i|^2 Ł_{1i}^{(1)}L_{2i}^{(1)}|\mathcal{L}\right) a.s.,$$

(2.A.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 $m_x(v,\eta)$:

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

and, by the bound condition on the kernel function and the similar argument as in (2.A.7), for $\imath = 2$ or 3:

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

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 (2.A.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 $O\left(nh_v^{5q/2}h_\eta^{5q/2}h_z^{q_z/2}+n^{5/2}h_v^{5q/2}h_\eta^{5q/2}h_z^{2(q_z+\xi_1)}+n^{5/2}h_v^{4q+2\xi}h_\eta^{4q+2\xi}h_z^{q_z/2}\right)$ $+O\left(n^4h_v^{(4q+2\xi)}h_\eta^{(4q+2\xi)}h_z^{2(q_z+\xi_1)}\right)$. Hence (2.A.26) is:

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

■

2.5.4 Proof of Theorem 2.2.2

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:

$$\hat{m}(v)-\check{m}(v)=\frac{1}{n}\sum_{i=1}^n\left\{\hat{m}(v,\hat{\eta}_i)-m(v,\eta_i)\right\}. \quad (2.A.28)$$

Given $\hat{\beta}$ and by using the definition of $m(v,\eta_i)$, we can rewrite the term in the bracket of (2.A.28) as:

$$\begin{aligned}\hat{m}(v,\hat{\eta}_i)-m(v,\eta_i) &= \{\hat{m}_y(v,\eta_i)-m_y(v,\eta_i)+\delta_{m_y,i}\}\\ &- \{\hat{m}_x(v,\eta_i)-m_x(v,\eta_i)+\delta_{m_x,i}\}'\beta\\ &- \{\hat{m}_x(v,\eta_i)-m_x(v,\eta_i)+\delta_{m_x,i}\}'\{\hat{\beta}-\beta\}\\ &= \{\hat{m}_{y^{**}}(v,\eta_i)-m_{y^{**}}(v,\eta_i)+\delta_{y^{**},i}\}\\ &- \{\hat{m}_x(v,\eta_i)-m_x(v,\eta_i)+\delta_{m_x,i}\}'\{\hat{\beta}-\beta\}, \quad (2.A.29)\end{aligned}$$

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 $l \geq 1$:

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

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

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

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 (2.A.29) is:

$$\hat{m}(v, \hat{\eta}_i) - m(v, \eta_i) = \{\hat{m}_{y^{**}}(v, \eta_i) - m_{y^{**}}(v, \eta_i) + \delta_{y^{**},i}\} + o_p(1), \qquad (2.A.30)$$

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 (2.A.30),

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} \{\hat{m}(v, \hat{\eta}_i) - m(v, \eta_i)\} &= \frac{1}{n}\sum_{i=1}^{n} \{\hat{m}_{y^{**}}(v, \eta_i) - m_{y^{**}}(v, \eta_i)\} + o_p(1) \\ &\equiv \frac{1}{n}\sum_{i=1}^{n} \{\hat{m}(v, \eta_i) - m(v, \eta_i)\} + o_p(1). \quad (2.A.31) \end{aligned}$$

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 (2.A.31) as:

$$\begin{aligned} \hat{m}(v, \eta_i) - m(v, \eta_i) &= \frac{\tilde{m}(v, \eta_i) - m(v, \eta_i)\hat{f}(v, \eta_i)}{\hat{f}(v, \eta_i)} \\ &= \frac{\tilde{m}(v, \eta_i) - m(v, \eta_i)\hat{f}(v, \eta_i)}{f(v, \eta_i)} \\ &\times \left[1 - \frac{\hat{f}(v, \eta_i) - f(v, \eta_i)}{\hat{f}(v, \eta_i)}\right]. \qquad (2.A.32) \end{aligned}$$

Note that the term $\left(\hat{f}(v, \eta_i) - f(v, \eta_i)\right)/\hat{f}(v, \eta_i)$ is $O_p\left(h_v^{p_2} h_\eta^{p_2} b_2^{-1} + (nh_v^q h_\eta^q b_2^2)^{-1/2}\right)$ 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:

$$
E(\hat{m}(v,\eta_i) - m(v,\eta_i)) = \{h_v^{p_2}B_v(v,\eta_i) + h_\eta^{p_2}B_\eta(v,\eta_i)\} + o(1). \tag{2.A.33}
$$

The single sum of (2.A.33) converges to its population mean by Chebyshev's law of large numbers; see Linton & 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:

$$V\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{m}(v,\eta_i)\right) = E\left(V_{v,\eta_i}\left\{\frac{1}{n}\sum_{i=1}^{n}\tilde{m}(v,\eta_i)\right\}\right) + V\left(E_{v,\eta_i}\left\{\frac{1}{n}\sum_{i=1}^{n}\tilde{m}(v,\eta_i)\right\}\right)$$

$$= \sigma^2 f(\eta)^2 E\left[\frac{1}{nh_v^q}\sum_{j=1}^{n}K_v\left(\frac{V_j - v}{h_v}\right)\right]^2 + f(\eta)^2 V\left[\frac{1}{nh_v^q}\sum_{j=1}^{n}K_v\left(\frac{V_j - v}{h_v}\right)m(V_j,\eta_j)\right]$$

$$= \frac{\sigma^2 f(\eta)^2}{nh_v^q}\mathcal{K}_v + \frac{m(v,\eta)^2 f(\eta)^2 f(v)}{nh_v^q}\mathcal{K}_v + O(n^{-1})$$

$$V\left(\frac{1}{n}\sum_{i=1}^{n}\hat{f}(v,\eta_i)\right) = \frac{f(\eta)^2 f(v)\mathcal{K}_v}{nh_v^q} + O(n^{-1})$$

$$Cov\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{m}(v,\eta_i), \frac{1}{n}\sum_{i=1}^{n}\hat{f}(v,\eta_i)\right) = E\left\{\frac{1}{n}\sum_{i=1}^{n}\tilde{m}(v,\eta_i)\frac{1}{n}\sum_{i=1}^{n}\hat{f}(v,\eta_i)\right\}$$

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

$$= \frac{m(v,\eta)f(\eta)^2 f(v)\mathcal{K}_v}{nh_v^q} + O(n^{-1}).$$

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 $m(v) = g(v) + \mathcal{C}$. ■

Chapter 3

Extended Generalised Partially Linear Single-Index Model with Control Function Approach

An important case of regression analysis is the comparison of regression curves from related samples. ... The problem of comparison of the two curves could be modeled parametrically because, to a large extent, the difference between them seems to be quantified by two parameters, horizontal shift and vertical scale.

Wolfgang Härdle and J. Steve Marron (1990)

3.1 Introduction

Since its introduction in the study by Carroll et al. (1997), the Generalised Partially Linear Single-Index (GPLSI) model has received constant attention and been studied by many researchers; see Yatchew (2003) and Gao (2007), for example. Furthermore, Xia et al. (1999) provide a useful extension to the model; in this chapter, let us refer to it as the extended GPLSI (EGPLSI) model. The EGPLSI model allows for the well-known advantages of a Single-Index (SI) model and a Partially Linear (PL) model (see the discussion in Chapter 2 of Horowitz (2009) for details) and also enables the analysis of the so-called shape-invariant specification as will be illustrated in Chapter 4. Unlike its GPLSI counterpart, the EGPLSI model concedes instead a more extensible specification, which includes the shape-invariant one as a special case.

Recently, considerable effort has been made in studies of the shape-invariant specification in the literature. While some interesting theoretical studies can be found in Härdle & Marron (1990) and Pinkse & Robinson (1995), the best known application is in the empirical demand study literature such as Blundell et al. (1998), Blundell et al. (2003) and Blundell et al. (2007). In the context of empirical demand studies, this specification enables the analysis of both a scale coefficient and a shift coefficient of a household characteristic in the modelling specification, which is coherent with the consumer theory; see Blundell et al. (1998), Pendakur (1999), Blundell et al. (2003) and Blundell et al. (2007) for details.

With regard to nonparametric estimation techniques employed, the study by Carroll et al. (1997) propose the local constant kernel estimation method, while Xia & Härdle (2006) consider the local polynomial estimation method of Fan & Gijbels (1996) to estimate the GPLSI model. On the other hand, Xia et al. (1999) employ the local constant kernel estimation method to estimate the EGPLSI model and to examine its identification condition. However, these methods are not directly applicable to empirical studies in various economic areas, since they do not take endogeneity into account. For example, the endogeneity of total expenditure is a well-known issue in the empirical demand study literature; see Blundell et al. (1998) and Blundell et al. (2007) for detail. If present, it might cause an inconsistent estimation of the model's scale coefficient and lead to non-identification of structural Engel curves. Recently, various methods of addressing endogeneity in nonparametric and semiparametric models have been discussed in the literature. Among these, a couple of the most popular methods are the nonparametric instrumental variables (NpIV) estimation and the control function (CF) approaches; see Blundell & Powell (2003) for an excellent review of these methods.

In the current chapter, we intend to introduce a method to address endogeneity in the estimation of the above-mentioned EGPLSI model. In particular, we aim to do so by establishing a CF approach based on (i) the Robinson (1988) and Speckman (1988) type of the two-stage estimation procedure and (ii) the widely-used triangular structure of Newey et al. (1999), Pinkse (2000), Blundell & Powell (2004) and Su & Ullah (2008). The two-stage estimation procedure allows us to conveniently identify the source(s) of endogeneity and hence systematically address it in a partially linear type of semiparametric models via the partialling-out process; see Chapter 2 for details. Furthermore, we present in detail below how imposition of the triangular structure enables us to identify the unknown structural relationship (e.g. the structural Engel curves) in a simple nonparametric additive structure which can be conveniently estimated using the marginal integration technique of Linton & Nielsen (1995), and Tjøstheim & Austad (1996). In spite of the involvement of an endogeneity control variable which is not observable in practice and hence is nonparametrically estimated for the flexibility (as in Newey et al. (1999)), we derive the asymptotic normality and the $\sqrt{n}$-consistency of parameter estimators of both the parametric coefficients and the index coefficients. More importantly, we show that the practicality of the study in Xia et al. (1999), which allows the same smoothing parameter in the estimation of the index coefficients and the unknown structural function, is still applicable to the EGPLSI model with the endogeneity control variable generated.

The structure of the rest of the chapter is as follows. In Section 3.2, we discuss an alternative method for addressing endogeneity in the estimation of the EGPLSI model in details. In Section 3.3, we presents the finite sample properties of the proposed estimators from Monte Carlo simulation exercises. Finally, Section 3.4 concludes the chapter, while mathematical proofs of the main results are presented in Appendix 3.5.

3.2 EGPLSI Model with/without Endogeneity

Let us begin the current section with a brief review of the EGPLSI model and its estimation procedure as often discussed in the literature (see Xia et al. (1999) and Gao (2007), for example). We introduce endogeneity into the model and then discuss our alternative CF based estimation procedure in Section 3.2.2. We present the main theoretical results of this chapter, which focus on the asymptotic properties of estimators of the model in Section 3.2.3. All mathematical proofs are discussed in the Appendix 3.5.

3.2.1 EGPLSI Model without Endogeneity

Generally, without the presence of endogeneity, the EGPLSI model can be defined as:

$$Y_i = X_i'\beta_0 + g(X_i'\alpha_0) + \epsilon_i, \tag{3.2.1}$$

where (X, Y) is a $\mathbb{R}^q \times \mathbb{R}$-valued observable random vector, β_0 and α_0 are unknown vector parameters, and $g(\cdot)$ is an unknown link function such that $g : \mathbb{R} \to \mathbb{R}$. The exogeneity assumption suggests that $E(\epsilon|x) = 0$, which implies that $E(\epsilon|v_0) = 0$ for $v_0 = x'\alpha_0$. Throughout the rest of the paper, let us assume that the random sample $\{(X_i', Y_i); i = 1, \dots, n\}$ is independently and identically distributed (i.i.d.). Furthermore, let $f(x)$ and $f(v_0)$ denote the density functions of x and v_0, respectively, with the random argument of X_i. We also assume that $\mathcal{A}_x \subseteq \mathbb{R}^q$ is the union of a finite number of open convex sets such that $f(x) > M_x$ on $\mathcal{A}_x$ for some constant $M_x > 0$. Finally, note the identification condition of the EGPLSI model investigated in Xia et al. (1999), the orthogonality of the two coefficients so that $\beta_0 \perp \alpha_0$ with $||\alpha_0|| = 1$.

Given α and β, we smooth the nonparametric index component out from the structural relation (3.2.1) to obtain the minimising objective function for both unknown coefficients as shown below:

$$\min_{\alpha,\beta} J^*(\alpha, \beta) = \min_{\alpha,\beta} E\left(W_i^* - U_i^{*\prime}\beta\right)^2, \tag{3.2.2}$$

where $W_i^* = Y_i - E^*(Y_i|V_i)$ and $U_i^* = X_i - E^*(X_i|V_i)$ with $V_i = X_i'\alpha$. In order to estimate those unknown parameters and functions involved in (3.2.1), we need to obtain a feasible version of (3.2.2). Firstly, consider the nonparametric kernel estimators of $E^*(Y_i|V_i)$ and $E^*(X_i|V_i)$ of the form:

$$\hat{E}^*(y|v) = \frac{\sum_{X_i \in \mathcal{A}_x} k_h(V_i - v)Y_i}{\sum_{X_i \in \mathcal{A}_x} k_h(V_i - v)} \quad \text{and} \quad \hat{E}^*(x|v) = \frac{\sum_{X_i \in \mathcal{A}_x} k_h(V_i - v)X_i}{\sum_{X_i \in \mathcal{A}_x} k_h(V_i - v)}, \tag{3.2.3}$$

where $k_h(\cdot) = k(\cdot/h)$, $k(\cdot)$ is a kernel function satisfying Assumption 3.2.4 below and h is a bandwidth parameter. Next, we turn to the corresponding estimators based on the usual cross-validation criterion. Let the estimators in (3.2.3) be the leave-one-out estimators by omitting (X_i, Y_i, V_i):

$$\hat{E}_i^*(y|v) = \frac{\sum_{j\neq i} k_h(V_j - v)Y_j}{\sum_{j\neq i} k_h(V_j - v)} \quad \text{and} \quad \hat{E}_i^*(x|v) = \frac{\sum_{j\neq i} k_h(V_j - v)X_j}{\sum_{j\neq i} k_h(V_j - v)}. \tag{3.2.4}$$

Let A_n denote the set of all unit q-vectors. Given $C > 0$ and $0 < C_1 < C_2 < \infty$, $A_n = \{\alpha \in A_n : ||\alpha - \alpha_0|| \leq Cn^{-1/2}\}$ and $\mathcal{H}_n = \left\{h : C_1 n^{-1/5} \leq h \leq C_2 n^{-1/5}\right\}$. These definitions are motivated by the fact that, since we anticipate that $\hat{\alpha}^*$ is $\sqrt{n}$-consistent and we expect $\hat{h}$ to be close to $h_0 \sim$ const $n^{1/5}$, we should look for a minimum of the feasible objective function of (3.2.2), i.e. $\hat{J}(\alpha, h)$, defined in Step 3.2.1.3 of Procedure 3.2.1 below. The feasible objective function involves α to be distant from α_0 by the order of $n^{-1/2}$ and h to be approximately equal to a constant multiple of $n^{-1/5}$; see Härdle et al. (1993) and Xia et al. (1999), for example. The estimation procedure of (3.2.1) can be summarised as follows. Hereafter, let us collectively refer to these estimation steps as *"Procedure 3.2.1"*.

Procedure 3.2.1

Step 3.2.1.1: Given α, obtain the feasible objective function of (3.2.2) by estimating $E^*(y|v)$ and $E^*(x|v)$ by $\hat{E}_i^*(y|v)$ and $\hat{E}_i^*(x|v)$ in (3.2.4).

Step 3.2.1.2: Define the feasible objective function of (3.2.2) as:

$$\hat{J}^*(\beta) = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{W}_i^* - \hat{U}_i^{*\prime}\beta\right)^2, \tag{3.2.5}$$

where $\hat{W}_i^* = Y_i - \hat{E}_i^*(Y_i|V_i)$ and $\hat{U}_i^* = X_i - \hat{E}_i^*(X_i|V_i)$. Perform the least squares (LS) estimation on (3.2.5) to obtain $\hat{\beta}^* = \left(S_{\hat{U}^*}\right)^- S_{\hat{U}^*\hat{W}^*}$, where $S_{AB} = \frac{1}{n}\sum_{i=1}^{n} A_i B_i'$, $S_A = S_{AA}$, and $\left(S_{\hat{U}^*}\right)^-$ is a generalised inverse of $\left(S_{\hat{U}^*}\right)$.

Step 3.2.1.3: Given $\hat{\beta}^*$ from the previous step, obtain $\hat{\alpha}^*$ and $\hat{h}$ by minimising the feasible objective function:

$$\min_{\alpha\in A_n, h\in\mathcal{H}_n} \hat{J}^*(\alpha, h) = \min_{\alpha\in A_n, h\in\mathcal{H}_n} \frac{1}{n}\sum_{i=1}^{n}(\hat{W}_i^* - \hat{U}_i^{*\prime}\hat{\beta}^*)^2.$$

Step 3.2.1.4: Re-estimate β_0 using $\hat{\alpha}^*$ and $\hat{h}$ from Step 3.2.1.3 as in 3.2.1.2:

$$\hat{\beta}^*_{\hat{\alpha}} = \left(S_{\hat{U}^*_{\hat{\alpha}}}\right)^{-} S_{\hat{U}^*_{\hat{\alpha}}\hat{W}^*_{\hat{\alpha}}},$$

where $\hat{W}_{\hat{\alpha}^*,i} = Y_i - \hat{E}^*_i(Y_i|\hat{V}_i)$ and $\hat{U}_{\hat{\alpha}^*,i} = X_i - \hat{E}^*_i(X_i|\hat{V}_i)$ with $\hat{V}_i = X_i'\hat{\alpha}$, $\hat{E}^*_i(Y_i|\hat{V}_i)$ and $\hat{E}^*_i(X_i|\hat{V}_i)$ obtained by replacing α in (3.2.4) with $\hat{\alpha}^*$.

Step 3.2.1.5: Given $\hat{\alpha}^*$ and $\hat{\beta}^*_{\hat{\alpha}}$, estimate the unknown structural function $g(\cdot)$ by $\hat{g}^*(\hat{v}) = \hat{E}^*(y|\hat{v}) - \hat{E}^*(x|\hat{v})'\hat{\beta}^*_{\hat{\alpha}}$. ■

The benefits of Procedure 3.2.1 of Xia et al. (1999) relies on the Robinson (1988) and Speckman (1988) type of the two-stage estimation procedure and the direct extension of the study in Härdle et al. (1993) to the EGPLSI model. On the one hand, the former conveniently allows for the identification of the source(s) of endogeneity and hence a systematic way of addressing endogeneity in partially linear semiparametrics due to the *partialling out* process as discussed above. On the other hand, the latter provides an empirical and practical way of estimating single-index semiparametrics. The study of Härdle et al. (1993) allows for the same bandwidth for the optimal estimation of $\hat{\alpha}^*$ and $\hat{g}^*(\cdot)$, and the simultaneous estimation of index coefficients and a smoothing parameter. Procedure 3.2.1 accommodates this practicality of Härdle et al. (1993) in the EGPLSI model. In the next section, we show that these benefits of Xia et al. (1999) can be extended to the proposed estimation procedure in the current paper to address endogeneity in the EGPLSI model.

3.2.2 EGPLSI Model with Endogeneity

Let us now introduce endogeneity into the EGPLSI model, (3.2.1). There are two potential sources of endogeneity, namely endogeneity in the parametric and the nonparametric components. Hereafter, let us refer to these as parametric endogeneity and nonparametric endogeneity, respectively. Clearly, these two types of endogeneity may also occur simultaneously. To simplify the argument, we assume that the parametric regressors belong to a subset of X, i.e. $X_1 \subseteq \mathbb{R}^{q_1}$ for $q_1 < q$, such that the regressors are exogenous with $E(\epsilon|x_1) = 0$. Nonparametric endogeneity exists for the case where $E(\epsilon|x) \neq 0$, which implies that $E(\epsilon|v_0) \neq 0$. Unless the parametric regressors are endogenous, the LS estimation results in the consistent estimation of the parametric coefficients even with nonparametric endogeneity in the model due to the partialling out process in the two-stage estimation procedure of Robinson (1988) and Speckman (1988). Note also that, if present, parametric endogeneity can be conveniently dealt with using the parametric IV estimation; see the discussion in Chapter 2 for details. Nonetheless, Procedure 3.2.1 does not take the above mentioned nonparametric endogeneity into account and may therefore result in inconsistent estimators for the index coefficients and in a failure to identify the unknown structural function. The formal result is due to similar reasoning to that in the classical linear regression model; see also the discussion in chapter 8 of Amemiya (1985) for details. Given β_0, reconsider the objective function of (3.2.2), particularly the following:

$$\begin{aligned} J(\alpha) &= E(W_i^* - U_i^{*\prime}\beta_0)^2 \\ &= E\left[\{g(V_{0i}) - g(V_i)\} + \epsilon_i - E(\epsilon_i|V_i)\right]^2 \\ &= E\left\{g(V_{0i}) - g(V_i)\right\}^2 + E\left\{\epsilon_i - E(\epsilon_i|V_i)\right\}^2 + 2E\left[\{g(V_{0i}) - g(V_i)\}\{\epsilon_i - E(\epsilon_i|V_i)\}\right] \\ &\equiv A_{1,1,i} + A_{1,2,i} + A_{1,3,i}. \end{aligned}$$

The feasible objective function in Step 3.2.1.3 of Procedure 3.2.1 does not converge to the function which provides consistent estimators of the index coefficients, since $A_{1,3,i}$ may not converge to 0 in probability, due to endogeneity, i.e. $E(\epsilon|x) \neq 0$; see Amemiya (1974), for example. When there is no endogeneity, the estimator of $A_{1,3,i}$ converges to 0 and the estimator of $A_{1,1,i}$ converges to the unique function providing the minimum value of the objective function with respect to the index coefficients in probability. Note that $A_{1,2,i}$ is not relevant to the index coefficients. Here more importantly, the unknown structural function is not identified. This is mainly because $E(\epsilon|x) \neq 0$, the conditional expectation of ϵ on any function of x is not 0. This leads to the conditional expectation relation $E^*(y|v) - E^*(x|v)'\beta_0 = g(v) + E(\epsilon|v)$, and $E(\epsilon|v) \neq 0$. Hence it is the case that $\hat{E}^*(y|\hat{v}) - \hat{E}^*(x|\hat{v})'\hat{\beta}^* = \hat{g}^*(\hat{v}) + \hat{E}(\epsilon|\hat{v}) \overset{p}{\nrightarrow} g(v_0)$, where $\overset{p}{\nrightarrow}$ denotes no convergence in probability.

In order to obtain consistent estimators of the index coefficients and to recover the unknown structural function when nonparametric endogeneity is present, we propose in the current section an alternative estimation method which is based on the CF approach; see the discussions in Newey et al. (1999), Blundell & Powell (2004) and Su & Ullah (2008) for its application to the nonparametric and semiparametric models. Let Z_i denote a vector of valid instruments for X_i such that:

$$X_i = m_x(Z_i) + \eta_i, \tag{3.2.6}$$

where:

$$E(\eta|z) = 0 \text{ and } E(\epsilon|x,\eta) = E(\epsilon|z,\eta) = E(\epsilon|\eta) = \iota(\eta), \tag{3.2.7}$$

and Z is an $\mathbb{R}^{q_z}$-valued vector, $q_z \geq q_2$ with $q_2 \equiv q - q_1$, $m_x(z)$ is a vector of unknown real functions, $m_x \equiv (m_{x,l}(Z_i))'$, $\{(Z_i'); i = 1, \ldots, n\}$ is i.i.d. and $m_{x,l} : \mathbb{R}^{q_z} \to \mathbb{R}$ for $l = 1, \ldots, q_2$. Also, let $f(z)$ denote the density function of z with the random argument of Z_i. Assume that $\mathcal{A}_z \subseteq \mathbb{R}^{q_z}$ is the union of a finite number of open convex sets such that $f(z) > M_z$ on $\mathcal{A}_z$ for some constant $M_z > 0$. The conditional expectation of the disturbance term in the reduced relation of (3.2.6), i.e. (3.2.7), is the distributional exclusion restriction; see the discussion on page 658 of Blundell & Powell (2004), which leads to the following argument. Hereafter, let us define the following:

$$m_y(v_0, \eta) = E(y|v_0, \eta) \quad \text{and} \quad m_x(v_0, \eta) = E(x|v_0, \eta), \tag{3.2.8}$$

by which:

$$Y_i = m_y(V_{0i}, \eta_i) + W_{0i} \quad \text{and} \quad X_i = m_x(V_{0i}, \eta_i) + U_{0i}, \tag{3.2.9}$$

where $E(w_0|x,\eta) = 0$ and $E(u_0|x,\eta) = 0$. We are now able to derive the conditional expectation relation which controls endogeneity by using (3.2.6) to (3.2.9):

$$m(v_0, \eta) \equiv m_y(v_0, \eta) - m_x(v_0, \eta)'\beta_0 = g(v_0) + \iota(\eta), \tag{3.2.10}$$

where $\iota(\eta) \neq 0$ is the endogeneity control function which controls the endogeneity in the structural relation.

By imposing the above mentioned distributional exclusion restriction (3.2.7), we have gained control over the endogeneity in the nonparametric regressors. As the results show, it provides the consistent estimators of the index coefficients and also a way to identify the unknown structural function. Given β_0, reconsider (3.2.2) so that we have:

$$\begin{aligned} J(\alpha) &= E\left(W_i - U_i'\beta_0\right)^2 \\ &= E\left[\{g(V_{0i}) - g(V_i)\} + \epsilon_i - \iota(\eta_i)\right]^2 \\ &\equiv E\left\{g(V_{0i}) - g(V_i)\right\}^2 + E(e_i)^2 - 2E\left[\{g(V_{0i}) - g(V_i)\} e_i\right] \\ &\equiv A_{2,1,i} + A_{2,2,i} + A_{2,3,i}, \end{aligned}$$

where $e_i \equiv \epsilon_i - \iota(\eta_i)$, $W_i = Y_i - E(Y_i|V_i, \eta_i)$ and $U_i = X_i - E(X_i|V_i, \eta_i)$. Note that the estimator of $A_{2,3,i}$ converges to 0 in probability, since $E(e|x, \eta) = 0$. Hence, the feasible objective function (3.2.17) defined in Step 3.2.2.3 of Procedure 3.2.2 below converges to the function which provides the local minimum value with respect to the index coefficients in probability; see chapters 4 and 8 of Amemiya (1985) for details. Furthermore, we may now identify the unknown structural function using the marginal integration technique, since (3.2.10) is a simple nonparametric additive structure. The details for implementing technique are given in Step 3.2.2.5. of Procedure 3.2.2 below.

Given β and α, the minimising objective function is:

$$\min_{\alpha,\beta} J(\beta, \alpha) = \min_{\alpha,\beta} E\left(W_i - U_i'\beta\right)^2. \tag{3.2.11}$$

Furthermore, let:

$$\hat{E}(y|v, \eta) = \frac{\sum_{X_i \in \mathcal{A}_x, Z_i \in \mathcal{A}_z} L_{h_v,h_\eta}(V_i - v, \eta_i - \eta) Y_i}{\sum_{X_i \in \mathcal{A}_x, Z_i \in \mathcal{A}_z} L_{h_v,h_\eta}(V_i - v, \eta_i - \eta)}, \tag{3.2.12}$$

and:

$$\hat{E}(x|v, \eta) = \frac{\sum_{X_i \in \mathcal{A}_x, Z_i \in \mathcal{A}_z} L_{h_v,h_\eta}(V_i - v, \eta_i - \eta) X_i}{\sum_{X_i \in \mathcal{A}_x, Z_i \in \mathcal{A}_z} L_{h_v,h_\eta}(V_i - v, \eta_i - \eta)}, \tag{3.2.13}$$

where $L_{h_v,h_\eta}(\cdot)$ is the product kernel function constructed from the product of the univariate kernel functions of $k_{h_{\eta_1}}(\cdot) \times \cdots \times k_{h_{\eta_{q_2}}}(\cdot) \times k_{h_v}(\cdot)$, and h_v and h_{η_j}

with $j = 1, \ldots, q_2$ are the relevant bandwidth parameters and are nonparametric kernel estimators of $E(y|v, \eta)$ and $E(x|v, \eta)$, respectively. Next, we turn to the corresponding leave-one-out estimators of (3.2.12) and (3.2.13) by omitting (X_i, Y_i, V_i, η_i):

$$\hat{E}_i(y|v, \eta) = \frac{\sum_{j \neq i} L_{h_v, h_\eta}(V_j - v, \eta_j - \eta) Y_j}{\sum_{j \neq i} L_{h_v, h_\eta}(V_j - v, \eta_j - \eta)} \tag{3.2.14}$$

and:

$$\hat{E}_i(x|v, \eta) = \frac{\sum_{j \neq i} L_{h_v, h_\eta}(V_j - v, \eta_j - \eta) X_j}{\sum_{j \neq i} L_{h_v, h_\eta}(V_j - v, \eta_j - \eta)}. \tag{3.2.15}$$

We redefine $\mathcal{H}_n$ in the previous section as
$\mathcal{H}_n = \left\{ h_v, h_\eta, h_z : C_1 n^{-1/5} \leq h_v, h_\eta, h_z \leq C_2 n^{-1/5} \right\}$. We propose the following estimation procedure. Hereafter, let us collectively refer to these estimation steps as *"Procedure 3.2.2"*.

Procedure 3.2.2

Step 3.2.2.0: Estimate the endogeneity control regressors from (3.2.6) as:

$$\hat{\eta}_i = X_i - \hat{m}_x(Z_i), \tag{3.2.16}$$

where $\hat{m}_x(z) = \frac{\sum_{Z_i \in \mathcal{A}_z} K_{h_z}(Z_i - z) X_i}{\sum_{Z_i \in \mathcal{A}_z} K_{h_z}(Z_i - z)}$, in which $K_{h_z}(\cdot)$ is the product kernel function constructed from the product of the univariate kernel functions of $k_{h_{z_1}}(\cdot) \times \cdots \times k_{h_{z q_z}}(\cdot)$ and h_{z_j} with $j = 1, \ldots, q_z$ is the relevant bandwidth parameter. By omitting the pair (X_i, Z_i), the corresponding leave-one-out estimator is $\hat{m}_{x,i}(z) = \frac{\sum_{j \neq i} K_{h_z}(Z_j - z) X_j}{\sum_{j \neq i} K_{h_z}(Z_j - z)}$.

Step 3.2.2.1: Given α and the nonparametrically generated endogeneity control regressors $\hat{\eta}_i$, obtain the feasible objective function of (3.2.11) by the estimates of $\hat{E}_i(y|v, \hat{\eta})$ and $\hat{E}_i(x|v, \hat{\eta})$, which are the corresponding estimates of those in (3.2.14) and (3.2.15) obtained by replacing η_i with $\hat{\eta}_i$.

Step 3.2.2.2: Define the feasible objective function of (3.2.11) as given below:

$$\hat{J}(\beta) = \frac{1}{n} \sum_{i=1}^{n} \left(\hat{W}_{2i} - \hat{U}_{2i}' \beta \right)^2,$$

where $\hat{W}_{2i} = Y_i - \hat{E}_i(Y_i|V_i, \hat{\eta}_i)$ and $\hat{U}_{2i} = X_i - \hat{E}_i(X_i|V_i, \hat{\eta}_i)$. We may compute the LS estimate of the unknown parametric coefficients as:

$$\hat{\beta}_\alpha = \left(S_{\hat{U}_2} \right)^- S_{\hat{U}_2 \hat{W}_2}.$$

Step 3.2.2.3: Given $\hat{\beta}$ from the previous step, compute $\hat{\alpha}$, $\hat{h}_v$ and $\hat{h}_{\hat{\eta}}$ by minimising the feasible objective function as follows:

$$\min_{\alpha \in A_n, h_v, h_{\hat{\eta}} \in \mathcal{H}_n} \hat{J}(\alpha, h_v, h_{\hat{\eta}}) = \min_{\alpha \in A_n, h_v, h_{\hat{\eta}} \in \mathcal{H}_n} \frac{1}{n} \sum_{i=1}^{n} (\hat{W}_{2i} - \hat{U}_{2i}' \hat{\beta})^2. \tag{3.2.17}$$

Step 3.2.2.4: Re-estimate β_0 using $\hat{\alpha}$, $\hat{h}_v$ and $\hat{h}_{\hat{\eta}}$ from the previous step as follows:

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

where $\hat{W}_{3i} = Y_i - \hat{E}_i(Y_i|\hat{V}_i, \hat{\eta}_i)$ and $\hat{U}_{3i} = X_i - \hat{E}_i(X_i|\hat{V}_i, \hat{\eta}_i)$ with $\hat{V}_i = X_i'\hat{\alpha}$.

Step 3.2.2.5 below is mainly due to the involvement of the marginal integration technique in an attempt to identify the unknown structural relation in question.

Step 3.2.2.5: Perform the marginal integration technique of Linton & Nielsen (1995) or Tjøstheim & Austad (1996) to identify the unknown structural function. ■

In the following paragraphs, we discuss an application of the marginal integration technique in Step 3.2.2.5 of Procedure 3.2.2 in greater detail. Let us first recall from (3.2.10) that $m(v_0, \eta) = g(v_0) + \iota(\eta)$, which is clearly a nonparametric additive specification. Hence a standard identification condition as discussed extensively in the literature (see Hastie & Tibishirani (1991), Gao et al. (2006) and Gao (2007), for example) assumes that $E(g(v_0)) = E(\iota(\eta)) = 0$. The implementation of the marginal integration technique identifies $g(\cdot)$ and $\iota(\cdot)$ up to some constant values as follows:

$$m(v_0) = \int m(v_0, \eta) dQ(\eta) = g(v_0) + c_1,$$

and:

$$m(\eta) = \int m(v_0, \eta) dQ(v_0) = \iota(\eta) + c_2,$$

where $c_1 = \int \iota(\eta) dQ(\eta)$, $c_2 = \int g(v_0) dQ(v_0)$ and Q is a probability measure with $\int dQ(\eta) = \int dQ(v_0) = 1$. Here, the estimate of the structural relation can therefore be obtained by the following sample version of the integration:

$$\hat{m}(v) = \frac{1}{n} \sum_{i=1}^{n} \hat{m}(v, \hat{\eta}_i), \tag{3.2.18}$$

and:

$$\hat{g}(v) = \hat{m}(v) - \hat{c}_1, \tag{3.2.19}$$

where $\hat{m}(v, \hat{\eta}_i) = \hat{E}(y|v, \hat{\eta}_i) - \hat{E}(x|v, \hat{\eta}_i)'\hat{\beta}_{\hat{\alpha}}$, and $\hat{c}_1 = \frac{1}{n}\sum_{i=1}^{n}\hat{m}(\hat{V}_i)$. Note that (3.2.18) is estimated by keeping $\hat{V}_i$ at v, while taking an average over the remaining regressors, $\hat{\eta}_i$. In (3.2.19), in order to ensure that the identification condition of a nonparametric additive model is satisfied, the constant value is estimated as $\hat{c}_1$.

An attractive feature of Procedure 3.2.2 is that the practicality of Xia et al. (1999), which provides a way of selecting the same smooth parameter(s) for optimal estimation of both α_0 and $g(\cdot)$ is still applicable, despite the regressors generated in order to control endogeneity in the model. The feasible objective function (3.2.17) can be expanded in the form of $\hat{J}(\alpha, h_v, h_{\hat{\eta}}) = \tilde{J}(\alpha) + T(h_v, h_\eta) + R_1(\alpha, h_v, h_\eta, h_z)$, where $\tilde{J}(\alpha)$ is an accurate approximation to $E(W_i - U_i'\beta_0)^2$ and does not depend on the smoothing parameters, $T(h_v, h_\eta)$ is the usual cross-validation criterion for choosing optimal bandwidths to estimate $m(x'\alpha_0, \eta)$ for known values of α_0 and true values of η, and R_1 is shown to be $o_p(n^{-1/2})$ in Theorem 3.2.1 below. Hence, minimising $\hat{J}(\alpha, h_v, h_{\hat{\eta}})$ simultaneously with respect to α, h_v and $h_{\hat{\eta}}$ is very much like separately minimising $\tilde{J}(\alpha)$ with respect to α and $T(h_v, h_\eta)$ with respect to h_v and h_η.

3.2.3 Asymptotic Properties

In this section, we present the main theoretical results of the current chapter. First, we present the necessary conditions and then the main theoretical results in Theorems 3.2.1 and 3.2.2. Within the results of Theorem 3.2.1, the asymptotic properties of both estimators of parametric and index coefficients are presented in Corollary 3.2.2, particularly the fact that they are $\sqrt{n}$-consistent. The asymptotic properties of the estimator of the unknown structural function are presented in Theorem 3.2.2. The formal proofs of these results are presented in the Appendix 3.5.

We impose the following regularity conditions. Assume that $\mathcal{A} = \mathcal{A}_x \times \mathcal{A}_\eta \subseteq \mathcal{R}^{2q_2}$ and $\mathcal{A}_z \subseteq \mathcal{R}^{q_z}$ are the unions of a finite number of open convex sets, respectively. Given ε_x, ε_η and ε_z, let $\mathcal{A}_x^{\varepsilon_x}$, $\mathcal{A}_\eta^{\varepsilon_\eta}$ and $\mathcal{A}_z^{\varepsilon_z}$ denote the sets of all points in $\mathcal{R}^{q_2}$ and $\mathcal{R}^{q_z}$ that are no more distant than ε_x, ε_η and ε_z, respectively. Put $\mathcal{U} = \left\{(v_0 = x'\alpha_0, \eta) : x \in \mathcal{A}_x^\varepsilon \text{ and } \eta \in \mathcal{A}_\eta^\varepsilon\right\}$, where ε is the smaller value of ε_x and ε_η, and $\mathcal{U}_z = \{z : z \in \mathcal{A}_z^{\varepsilon_z}\}$. Let $f(v_0, \eta)$ denote the joint density function of $(x'\alpha_0, \eta)$ with random arguments of X_i and η_i. Assume that for some ε and ε_z, we have the assumptions below.

Assumption 3.2.1. *$f(x, \eta)$ and $f(z)$ are bounded away from 0 on $\mathcal{U}$ and $\mathcal{U}_z$, respectively.*

Assumption 3.2.2. *$f(z)$ and $m_x(z)$ have bounded and continuous second derivatives on $\mathcal{U}_z$.*

Assumption 3.2.3. *$m(v, \eta)$, $m_y(v, \eta)$, $m_x(v, \eta)$ and $f(v, \eta)$ have bounded and continuous second derivatives on $\mathcal{U}$ for all values of $\alpha \in A_n$.*

Assumption 3.2.4. *A univariate kernel function $k(\cdot)$ and its first derivative $k^{(1)}(\cdot)$ are supported on the interval $(-1, 1)$ and $k(\cdot)$ is a symmetric probability density with $k^{(1)}(\cdot)$ being bounded.*

Assumption 3.2.5. *$E(e|x, \eta) = 0$ and $E(e^2|x, \eta) = \sigma^2(x, \eta)$, $E(u|x, \eta) = 0$ and $E(u^2|x, \eta) = \boldsymbol{u}^2(x, \eta)$ almost surely, and $\sup_i E|Y_i|^l < \infty$ and $\sup_i E||X_i||^l < \infty$ for some $l > 2$.* ■

Assumption 3.2.1 is imposed to permit estimation of the functions in the regions of $\mathcal{A}^\varepsilon$ and $\mathcal{A}_z^{\varepsilon_z}$ in order to avoid the random denominator problem. A similar set of conditions is imposed in Härdle et al. (1993) and Xia et al. (1999). Assumptions 3.2.2 and 3.2.3 are needed to ensure that the symmetric kernel function in Assumption 3.2.4 leads to a second-order bias in kernel smoothing. A higher-order bias can be achieved by imposing more restrictive conditions on the smoothness of functions. For instance, Robinson (1988) reduces the bias sufficiently by employing a higher-order kernel function with strong smoothness conditions on the functions. The condition of the first derivative of the kernel function in Assumption 3.2.4 is required because we employ the Taylor argument to address the generated regressors, $\hat{\eta}$. A similar condition on the rth derivative of the kernel function can be found in Hansen (2008). Assumption 3.2.5 is imposed so that the Chebyshev inequality can be applied as in Härdle et al. (1993) and Xia et al. (1999).

Let us define the following:

$$\begin{aligned} B_v(v_0,\eta) &= \frac{K_{v,2}}{f(v_0,\eta)}\left\{f_v^{(1)}(v_0,\eta)m_0^{(1)}(v_0)+f(v_0,\eta)m_0^{(2)}(v_0)\right\} \\ B_{\eta,j}(v_0,\eta) &= \frac{K_{\eta,2}}{f(v_0,\eta)}\left\{f_{\eta,j}^{(1)}(v_0,\eta)m_j^{(1)}(\eta)+f(v_0,\eta)m_j^{(2)}(\eta)\right\}, \end{aligned}$$

where $\mathcal{K}_{v,2} = \int v_0^2 k_{h_v}(v_0)dv_0$, $\mathcal{K}_{\eta,2} = \int \eta^2 K_{h_\eta}(\eta)d\eta$ with $K_{h_\eta} = k_{h_{\eta_1}}(\cdot)\times\cdots\times k_{h_{\eta_{q_2}}}(\cdot)$, $f_v^{(r)}$ and $f_\eta^{(r)}$ are rth derivatives of the joint density function of $f(v_0,\eta)$ with respect to v_0 and η, respectively, and $m_0^{(r)}(v_0)$ and $m_j^{(r)}(\eta)$ are the rth partial derivatives of the function $m(v_0,\eta)$ with respect to v_0 and η_j, respectively, where $j=1,\ldots,q_2$. Also, let $\mathcal{K}=\mathcal{K}_v\mathcal{K}_\eta^{q_2}$, where $\mathcal{K}_v=\int k_{h_v}(v_0)^2dv_0$ and $\mathcal{K}_\eta=\int k_{\eta,j}(\eta)^2d\eta$. In these notations, the "integrated mean squared error (IMSE)" is:

$$\begin{aligned} &IMSE(h_v,h_\eta) \\ &\asymp \int\left\{\left[B_v(v_0,\eta)h_v^2+\sum_{j=1}^{q_2}B_\eta(v_0,\eta)h_{\eta,j}^2\right]^2+\frac{\mathcal{K}}{nh_vh_{\eta,1}\ldots h_{\eta,q_2}}\frac{\sigma^2(v_0,\eta)}{f(v_0,\eta)}\right\}f(x,\eta)dxd\eta, \end{aligned}$$

where $\asymp$ means that the quotient of the two sides tends to 1 and $n\to\infty$. Now let us define the following:

$$\tilde{J}(\alpha)=\frac{1}{n}\sum_{i=1}^n\left\{W_i-U_i'\hat{\beta}\right\}^2 \quad\text{and}\quad T(h_v,h_\eta)=\frac{1}{n}\sum_{i=1}^n\left\{\hat{m}_i(V_{0i},\eta_i)-m(V_{0i},\eta_i)\right\}^2,$$

where $\hat{m}_i(\dot{})$ is the leave-one-out kernel estimator of $m(\cdot)$. Hence, we have the result shown in Theorem 3.2.1.

Theorem 3.2.1. *Under Assumptions 3.2.1 to 3.2.5, we can write:*

$$\hat{J}(\alpha,h_v,h_{\hat{\eta}})=\tilde{J}(\alpha)+T(h_v,h_\eta)+R_1(\alpha,h_v,h_\eta,h_z)+R_2(\alpha,h_v,h_\eta), \tag{3.2.20}$$

$$T(h_v,h_\eta)=IMSE(h_v,h_\eta)+R_3(h_v,h_\eta), \tag{3.2.21}$$

where $R_3(h_v, h_\eta)$ does not depend on α, and:

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

and:

$$\sup_{h_v, h_\eta \in \mathcal{H}_n} |R_3(h_v, h_\eta)| = o_p(1).$$

■

The above theorem is a direct extension of the work of Xia et al. (1999) to a more complicated model associated with endogeneity. Now, let us define the following:

$$\Phi_{U_0} = X_i - E(X_i|V_{0i}, \eta_i) \text{ and } m_0^{(1)} = \partial m(v_0, \eta)/\partial v_0.$$

As the results of Theorem 3.2.1 show, we have the asymptotic results for the estimators of α_0 and β_0 shown in Corollary 3.2.2 below.

Corollary 3.2.2. *Under the assumptions of Theorem 3.2.1, we obtain the following:*

$$\sqrt{n}(\hat{\beta} - \beta_0) \rightarrow_D (0, var_1),$$

where $var_1 = \sigma^2 \left[\Phi_{U_0}^{-} - \left(m_0^{(1)} \Phi_{U_0}\right)^{-} \Phi_{U_0} \left\{m_0^{(1)}\right\}^2 \left(m_0^{(1)} \Phi_{U_0}\right)^{-} \right]$ *and:*

$$\sqrt{n}(\hat{\alpha} - \alpha_0) \rightarrow_D (0, var_2),$$

where $var_2 = \sigma^2 \left[\left\{(m_0^{(1)})^2 \Phi_{U_0}\right\}^{-} - \left\{m_0^{(1)} \Phi_{U_0}\right\}^{-} \Phi_{U_0} \left\{m_0^{(1)} \Phi_{U_0}\right\}^{-} \right].$ ■

As for the estimator of the unknown structural function, i.e. $g(\cdot)$, we have the asymptotic properties shown in Theorem 3.2.2.

Theorem 3.2.2. *Under Assumptions 3.2.1 to 3.2.5, we show that:*

$$\sqrt{nh_v}\left(\hat{g}(\hat{v}) - g(v_0) - bias\right) \rightarrow_D N(0, var),$$

where $bias = h_v^2 \mathcal{B}_v(v_0, \eta) + \sum_{s=1}^{q_2} h_{\eta,s}^2 B_{\eta,s}(v_0, \eta)$ *and* $var = f(v_0) \mathcal{K}_v \int \frac{\sigma^2(v_0,\eta) f^2(\eta)}{f^2(v_0,\eta)} dQ(\eta).$

■

The proofs of Theorems 3.2.1 and 3.2.2 as well as Corollary 3.2.2 are given in the Appendix 3.5.

3.3 Simulation Studies

The purposes of the simulation exercises conducted in this section are threefold. Firstly, we aim to investigate whether experimental evidence can be found to support the various points made in the theoretical discussion presented in the previous sections. Secondly, we aim to provide finite sample evidence for the usefulness of the newly introduced method for addressing endogeneity in the estimation of semiparametric SI models. Finally, we aims to examine the empirical importance of the magnitude of endogeneity and nature of the instrumental variable on the use of our Procedure 3.2.2.

3.3.1 Initial Investigation

In this section, we consider two illustrative models, namely the GPLSI-type and the EGPLSI-type, as defined in Examples 3.3.1 and 3.3.2 below. In practice, endogeneity is introduced to the models first and then Procedure 3.2.2 is applied. The finite sample performance of Procedure 3.2.2 is subsequently compared to that of Procedure 3.2.1.

Example 3.3.1: *GPLSI-type* The baseline model without endogeneity is:

$$Y_i = 1.2X_{1i} + g(V_{0i}) + \epsilon_i \text{ with } V_{0i} = \alpha_0 X_{2i} = \frac{1}{\sqrt{2}}(X_{2i}), \tag{3.3.1}$$

such that:

$$g(V_{0i}) = \frac{1}{2}\left\{\frac{\frac{1}{\sqrt{2}}(X_{2i})}{1+\left[\frac{1}{\sqrt{2}}(X_{2i})\right]^2}\right\},$$

where X_1 and X_2 are independently and uniformly distributed on $[-1, 1]$ and $\epsilon_i \sim N(0, 1)$. Clearly, (3.3.1) is a GPLSI type of model such that the perpendicularity of the parameter vectors (see Xia et al. (1999), for instance) is not required. Furthermore, since X_2 can be written as $X_{21} + X_{22}$, where X_{21} and X_{22} are independently and uniformly distributed on $[-1, 1]$, we obtain $||\alpha_0|| = 1$. In this example, we introduce endogeneity into the nonparametric regressor by letting $X_{2i} = Z_i + \eta_i$, where Z and η are independently and uniformly distributed on $[-0.5, 0.5]$ and $[-1, 1]$, respectively, and $\epsilon_i = \eta_i + e_i$ and $e_i \sim N(0, 1)$. ■

Example 3.3.2: *EGPLSI-type* Due to the complexity of the model and conditions involved, first let us express our example model of the EGPLSI-type in general as:

$$Y_i = \beta_{01}X_{1i} + \beta_{02}X_{2i} + \beta_{03}X_{3i} + g(V_{0i}) + 0.1\epsilon_i, \tag{3.3.2}$$

where $V_{0i} = \alpha_{01}X_{1i}+\alpha_{02}X_{2i}+\alpha_{03}X_{3i}$, $g(V_{0i}) = \exp\left\{-2(\alpha_{01}X_{1i} + \alpha_{02}X_{2i} + \alpha_{03}X_{3i})^2\right\}$, X_j is independently and uniformly distributed on $[-1, 1]$ for $j = 1, 2, 3$ and $\epsilon_i \sim N(0, 1)$. It is required that (i) β_0 and α_0 are perpendicular to each other with (ii) $\| \alpha_0 \| = 1$. In order for these conditions to be satisfied, we define $\beta_{02} = 0.4$, $\beta_{03} = 0$, $\alpha_{01} = 0.7$, $\alpha_{02} = -0.6$, then write:

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

In this example, we introduce endogeneity into the nonparametric regressor by letting $X_{3i} = Z_i + \eta_i$, where Z and η are independently and uniformly distributed on $[-0.5, 0.5]$ and $[-1, 1]$, respectively, and $\epsilon_i = \eta_i + e_i$ and $e_i \sim N(0, 1)$. ■

Throughout this section, optimisation is implemented using a limited memory Broyden–Fletcher–Goldfarb–Shanno algorithm for the bound constrained optimisation of Byrd et al. (1995). All simulation exercises are conducted in R with the Gaussian kernel function and the number of replications $Q = 200$. To compare and evaluate the finite sample performances of the estimation procedures introduced above, we compute the mean and mean absolute errors of the estimates of both coefficients across Q replications as tabulated in Tables 3.1 to 3.4. We also compare the averaged absolute error (ae) of the estimates the unknown structural function which is computed for Procedure 3.2.1 and for Procedure 3.2.2 using the following:

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

where n is the number of samples.

Table 3.1: GPLSI-type model with nonparametric endogeneity: Procedure 3.2.1

n	$\hat{\beta}$	$\hat{\alpha}$	$\lvert\hat{\beta} - 1.2\rvert$	$\lvert\hat{\alpha} - 1/\sqrt{2}\rvert$	$ae_{\hat{g}}$
50	1.1997	0.8980	0.0060	0.1980	0.0438
150	1.1994	0.8592	0.0031	0.1592	0.0443
300	1.1999	0.7306	0.0024	0.0402	0.0443
500	1.2001	0.6523	0.0016	0.0708	0.0446

Table 3.2: GPLSI-type model with nonparametric endogeneity: Procedure 3.2.2

n	$\hat{\beta}$	$\hat{\alpha}$	$\lvert\hat{\beta} - 1.2\rvert$	$\lvert\hat{\alpha} - 1/\sqrt{2}\rvert$	$ae_{\hat{g}}$
50	1.2000	0.8272	0.0033	0.1436	0.0266
150	1.1999	0.7784	0.0015	0.0796	0.0176
300	1.2000	0.7527	0.0082	0.0578	0.0148
500	1.9999	0.7502	0.0006	0.0580	0.0118

n	$\lvert\hat{\beta}_1 - 0.3\rvert$	$\lvert\hat{\beta}_2 - 0.4\rvert$	$\lvert\hat{\alpha}_1 - 0.8\rvert$	$\lvert\hat{\alpha}_2 - (-0.6)\rvert$	$\lvert\hat{\alpha}_3 - 0.5\rvert$	$ae_{\hat{g}}$
50	0.0656	0.0714	0.1905	0.1772	0.1727	0.0905
150	0.0428	0.04572	0.1999	0.1671	0.1406	0.0891
300	0.0331	0.03377	0.1988	0.1674	0.1352	0.0895
500	0.0306	0.0319	0.1960	0.1653	0.1306	0.0906

Table 3.3: EGPLSI-type model with nonparametric endogeneity: Procedure 3.2.1.

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

Table 3.4: EGPLSI-type model with nonparametric endogeneity: Procedure 3.2.2.

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

n	$\lvert\hat{\beta}_1 - 0.3\rvert$	$\lvert\hat{\beta}_2 - 0.4\rvert$	$\lvert\hat{\alpha}_1 - 0.8\rvert$	$\lvert\hat{\alpha}_2 - (-0.6)\rvert$	$\lvert\hat{\alpha}_3 - 0.5\rvert$	$\mathrm{ae}_{\hat{g}}$
50	0.0816	0.0684	0.2640	0.2329	0.1782	0.0632
150	0.0307	0.0264	0.2143	0.1854	0.1267	0.0265
300	0.0213	0.0183	0.1203	0.1005	0.0702	0.0160
500	0.0189	0.0159	0.0377	0.0319	0.0251	0.0124

Let us now present some important findings based on the results in Tables 3.1 to 3.4. Since endogeneity is introduced to the nonparametric regressor only, we expect the LS estimators of the unknown parameters in the parametric component to be consistent in all cases. Strong experimental evidence of such consistency can be clearly seen in all of the tables; see the fourth column of Tables 3.1 and 3.2, and the eighth to tenth columns of Tables 3.3 and 3.4 in particular. The simulation results in Tables 3.1 and 3.3 show strong evidence against the use of Procedure 3.2.1 of Xia et al. (1999) to estimate the EGPLSI model when endogeneity is a possibility. Most importantly, such evidence is clearly seen when the averaged absolute errors in the sixth column of Table 3.1 and the thirteenth column of Table 3.3 are considered. Basically, the results suggest that the procedure is unable to provide a consistent estimate of the unknown structural function. On the other hands, the simulation results in Table 3.2 and Table 3.4 suggest that Procedure 3.2.2 is able to provide consistent estimates of the index parameters and is also able to identify the unknown structural function when endogeneity is present. In our view, such a conclusion should provide sufficient motivation for use of our newly established procedure in practice. However, in the next section, let us conduct a further investigation into the important of the magnitude of endogeneity and nature of the instrumental variable on the use of Procedure 3.2.2.

3.3.2 More Detailed Analysis

For the sake of clarity in illustrating the importance of some particular characteristics of endogeneity, the model used in the analysis that follows will be structurally similar to that of Example 3.3.1. However, some modifications will be made to ensure that the experimental design is suitable to the objectives of the exercise. In this section, we will conduct two types of analysis, which are referred to hereafter as *Type A* and *Type B*, respectively.

Type A: The objective of the experimental analysis that follows is to study the importance of the conditional expectation of ϵ given η, i.e. denoted previously as $\iota(\cdot)$, for the performance of Procedure 3.2.1, which was originally introduced in Xia et al. (1999), in the presence of endogeneity. In such an experiment, the magnitude of endogeneity is clearly an important parameter that must be carefully controlled. In this current analysis, in order to best illustrate the impact of endogeneity, let us consider an extreme case, i.e. by defining:

$$X_{2i} = \eta_i, \tag{3.3.3}$$

where η_i is independently and uniformly distributed on $[-1, 1]$. Defining X_{2i} as in (3.3.3) enables specification of three related types of models, namely “exogeneity”,

Table 3.5: Nonparametric exogeneity, i.e. $\iota 1$.

n	$\hat{\beta}$	Bias	Var	$\lvert\hat{\beta}-\beta\rvert$	$\hat{\alpha}$	Bias	Var	$\lvert\hat{\alpha}-\alpha\rvert$	$ae_{\hat{g}}$
100	1.1997	0.0002	0.0003	0.0143	0.9660	0.2660	0.0008	0.2660	0.0150
300	1.1996	0.0004	0.0001	0.0079	0.7989	0.0989	0.0012	0.0989	0.0108
500	1.2001	0.0001	0.0000	0.0055	0.7740	0.0740	0.0056	0.0740	0.0084
700	1.2005	0.0005	0.0000	0.0055	0.7330	0.0330	0.0045	0.0332	0.0073

"linear endogeneity" and "nonlinear endogeneity". In the current sections, these models can be respectively obtained by introducing the following:

$$\iota 1(\eta) = 0 \times \eta, \tag{3.3.4}$$

$$\iota 2(\eta) = 0.5 \times \eta, \tag{3.3.5}$$

$$\iota 3(\eta) = \frac{\eta}{\frac{1}{2}\left(4+\eta^2\right)}. \tag{3.3.6}$$

For example, (3.3.4) suggests that the conditional expectation of ϵ given η is zero and the model is exogenous. An example of $g(\cdot)$, $\iota 1(\cdot)$, $\iota 2(\cdot)$ and $\iota 3(\cdot)$ with $n = 500$ is presented in Figure 3.1. The simulation results in this section are presented in Tables 3.5 to 3.7.

Table 3.6: Linear endogeneity, i.e. $\iota 2$.

n	$Corr_L$	$\hat{\beta}$	Bias	Var	$\lvert\hat{\beta}-\beta\rvert$	$\hat{\alpha}$	Bias	Var	$\lvert\hat{\alpha}-\alpha\rvert$	$ae_{\hat{g}}$
100	0.9852	1.1998	0.0001	0.0003	0.0145	0.9910	0.2910	0.0001	0.2910	0.2474
300	0.9852	1.1994	0.0005	0.0001	0.0079	0.8039	0.1039	0.0049	0.1039	0.2492
500	0.9853	1.2000	0.0000	0.0000	0.0057	0.8092	0.1093	0.0128	0.1092	0.2496
700	0.9853	1.2001	0.0005	0.0000	0.0056	0.7721	0.0721	0.0124	0.0898	0.2491
900	0.9853	1.1997	0.0002	0.0000	0.0043	0.8072	0.1072	0.0199	0.1341	0.2492
1,100	0.9853	1.2003	0.0003	0.0000	0.0040	0.7595	0.0595	0.0115	0.0932	0.2494
1,300	0.9853	1.1995	0.0004	0.0000	0.0035	0.7591	0.0591	0.0133	0.0982	0.2495

Table 3.7: Nonlinear endogeneity, i.e. $\iota 3$.

n	$Corr_{NL}$	$\hat{\beta}$	Bias	Var	$\lvert\hat{\beta}-\beta\rvert$	$\hat{\alpha}$	Bias	Var	$\lvert\hat{\alpha}-\alpha\rvert$	$ae_{\hat{g}}$
100	0.9514	1.1998	0.0001	0.0003	0.0146	0.9852	0.2852	0.0001	0.2852	0.3748
300	0.9505	1.1995	0.0004	0.0001	0.0079	0.8573	0.1573	0.0079	0.1573	0.3771
500	0.9513	1.2000	0.0000	0.0000	0.0057	0.8882	0.1882	0.0099	0.1883	0.3777
700	0.9514	1.2005	0.0005	0.0000	0.0056	0.8592	0.1592	0.0099	0.1602	0.3771

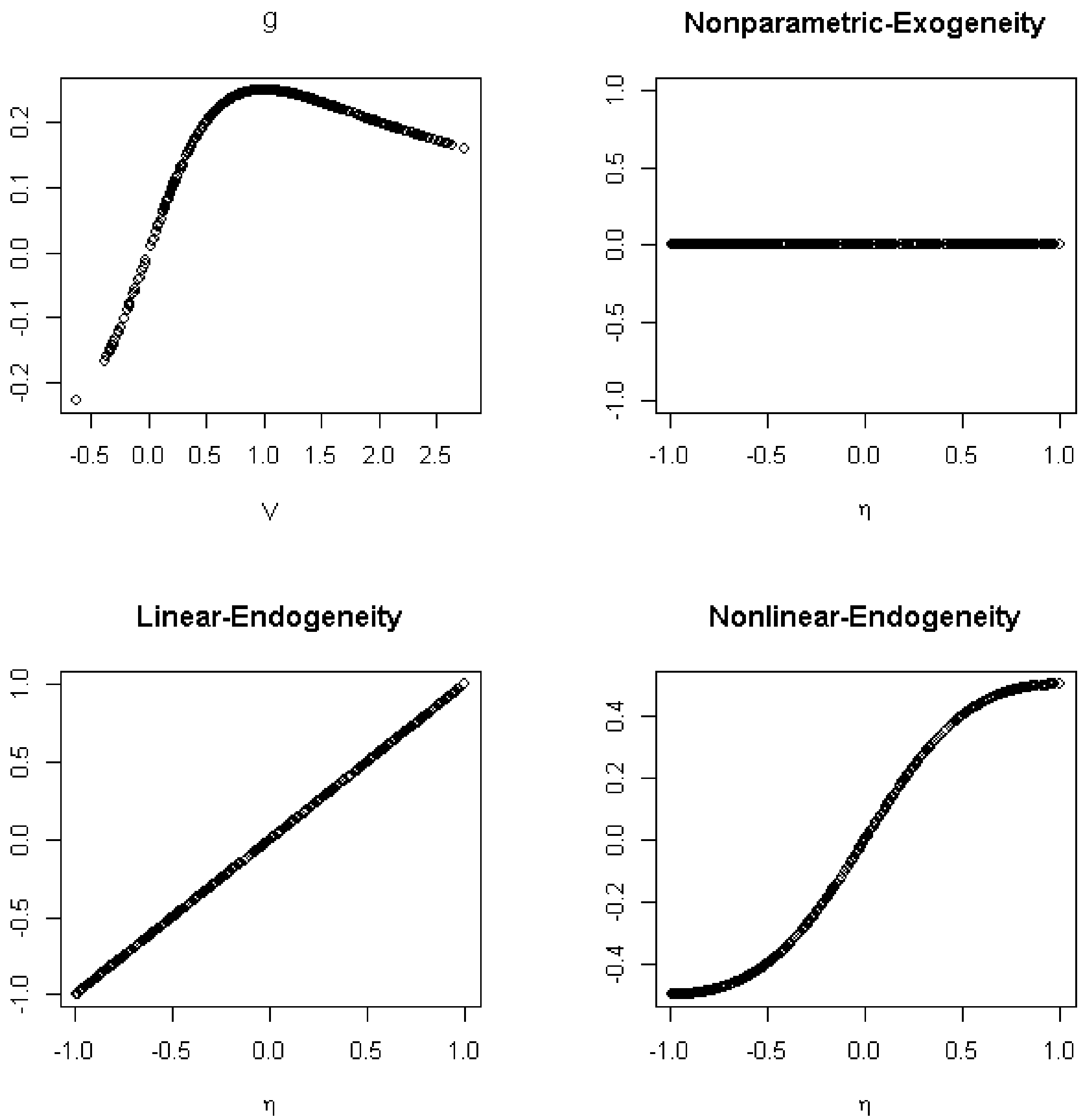
Table 3.8: $Corr_{X_{2i},Z_i}$

n	100	300	500	700
$Corr_{X_{2i},Z_i}$	0.8278	0.8302	0.8326	0.8330

Below, let us discuss some important findings. Note firstly that $E[\epsilon] = 0$, which implies that $E[\epsilon|\eta] = E[\epsilon] = 0$ when η and ϵ are independent. Therefore, in this case, we are able to measure the magnitude of endogeneity by simply considering the dependency between ϵ and η. The second columns of Tables 3.6 and 3.7, present averages over $Q = 200$ replications of the empirical correlation coefficients, which is a measure the linear dependence between ϵ and η. It is clear that even in such a controlled case, the functional forms of $\iota(\cdot)$ give rise to different magnitudes of endogeneity, which are measured by $Corr_L$ and $Corr_{NL}$. Since endogeneity is introduced to the nonparametric regressor only, the LS estimators of the unknown parameters in the parametric component seem to be consistent in all cases, as expected. Compared to the simulation results in Table 3.5, those in Tables 3.6 and 3.7 show clearly that Procedure 3.2.1 does not work well in the presence of endogeneity. Under linear endogeneity, the procedure seems to work quite well in estimating the index coefficient up to about 700 observations. By extending the number of observations to 900, 1,100 and 1,300, it becomes clear that $|\hat{\alpha} - \alpha|$ shows no sign of converging to zero. Furthermore, the evidence suggests that the procedure is incapable of identifying the unknown structural function when (either linear or nonlinear) endogeneity is present. Overall, nonlinear endogeneity seems to have somewhat more severe consequences when compared to its linear counterpart.

Type B: The objective of the analysis that follows is to investigate the finite-sample performance of our newly introduced Procedure 3.2.2 in the presence of endogeneity. In practice, whether Z_i is a weak or a strong instrument may significantly affect the estimation outcomes. In order to control for such an effect, let

Figure 3.1: $g(\cdot)$, $\iota 1(\cdot)$, $\iota 2(\cdot)$ and $\iota 3(\cdot)$.



us define the following:

$$X_{2i} = Z_i + \eta_i, \tag{3.3.7}$$

where Z and η are independently and uniformly distributed on $[0, 3]$ and $[-1, 1]$, respectively. Furthermore, we consider two cases of $\iota(\cdot)$, namely linear endogeneity and nonlinear endogeneity defined respectively as:

$$\iota_2(\eta) = 1 \times \eta \quad \text{and} \quad \iota_3(\eta) = \frac{\eta}{1+\eta^2}. \tag{3.3.8}$$

While Table 3.8 presents the averaged correlation coefficient of X_{2i} and Z_i at $Q = 200$ replications for $n = 100, 300, 500$ and 700, Tables 3.9 and 3.10 provide simulation results.

Below, let us discuss some important findings. Once again, the functional forms of $\iota(\cdot)$ seem to be important factors which determines the nature of endogeneity. With an instrument of a particular explanatory power in (3.3.7), linear-endogeneity tends to give a higher $Corr_L$ than $Corr_{NL}$ obtained from its nonlinear counterpart. An important observation which can be brought forward is that even for cases in which we are able to identify a strong instrument (with strong explanatory power), the impact of endogeneity is still determined by the relationship between ϵ and η, i.e. the conditional expectation of the former with respect to the latter. Furthermore, compared the results in Tables 3.9 and 3.10 to those presented in Tables 3.6 and 3.7, it is clear that our newly developed Procedure 3.2.2 performs much better than its Procedure 3.2.1 counterpart in the presence of endogeneity. Procedure 3.2.2 seems to be capable of obtaining consistent estimators of all the unknowns, including the parametric and index coefficients, and the unknown structural function.

Table 3.9: Linear endogeneity, i.e. $\iota 2$.

n	$Corr_L$	$\hat{\beta}$	Bias	Var	$\lvert\hat{\beta}-\beta\rvert$	$\hat{\alpha}$	Bias	Var	$\lvert\hat{\alpha}-\alpha\rvert$	$ae_{\hat{g}}$
100	0.9852	1.1972	0.0027	0.0008	0.0226	0.7803	0.0803	0.0076	0.0999	0.0827
300	0.9852	1.2009	0.0009	0.0001	0.0099	0.7372	0.0372	0.0009	0.0406	0.0511
500	0.9854	1.2003	0.0003	0.0001	0.0085	0.7137	0.0137	0.0005	0.0212	0.0439
700	0.9853	1.2008	0.0008	0.0000	0.0054	0.6948	0.0051	0.0002	0.0135	0.0385

Table 3.10: Nonlinear endogeneity, i.e. $\iota 3$.

n	$Corr_{NL}$	$\hat{\beta}$	Bias	Var	$\lvert\hat{\beta}-\beta\rvert$	$\hat{\alpha}$	Bias	Var	$\lvert\hat{\alpha}-\alpha\rvert$	$ae_{\hat{g}}$
100	0.6744 7	1.2004	0.0004	0.0003	0.0156	0.7863	0.0863	0.0021	0.0869	0.0326
300	0.6743	1.2001	0.0002	0.0001	0.0086	0.7248	0.0248	0.0005	0.0296	0.0230
500	0.6767	1.1998	0.0002	0.0000	0.0069	0.7082	0.0082	0.0001	0.0118	0.0196
700	0.6768	1.2008	0.0008	0.0000	0.0052	0.7016	0.0016	0.0000	0.0053	0.0171

3.4 Conclusions

Although the GPLSI model by Carroll et al. (1997), and Xia & Härdle (2006) has great flexibility and advantages from both a PL model and a SI model perspective, it is not appropriate for modelling the shape-invariant empirical Engel curve, since it does not allow the coefficient of the household equivalence scale to be included. Hence we consider the EGPLSI model of Xia et al. (1999) and Gao (2007) in order to take the shape-invariant specification into account. However, the estimation method and procedure of the existing EGPLSI model are not able to address endogeneity. Hence we establish the CF approach in the EGPLSI model to address endogeneity instead of the nonparametric IV estimation of Ai & Chen (2003); see Blundell et al. (2007) for its application to an semiparametric analysis of empirical Engel curves. The attractive feature of the proposed estimation procedure in the current chapter is that the practicality of Xia et al. (1999) approach is still applicable, despite the endogeneity control variable generated. The same bandwidth parameters are used for the estimation of α and $g(\cdot)$-function.

We also provide Monte Carlo simulation studies. The simulation studies illustrate the performance of CF approach which indicates that the proposed estimation procedure performs well and is able to address endogeneity in the estimation of the EGPLSI model.

3.5 Appendix

In this section, we provide the necessary mathematical proofs of the main theoretical results of the current paper. In Section 3.5.1, we show the proofs of Theorem 3.2.1 and Corollary 3.2.2 in two main steps. In Section 3.5.2, we present the proofs of Theorem 3.2.2.

3.5.1 Proofs of Theorem 3.2.1 and Corollary 3.2.2

Step 1. Proofs of Theorem 3.2.1: Given α, $\hat\eta$ and $\hat\beta$, the feasible objective function (3.2.17) is expanded as follows:

$$\begin{aligned}\hat J(\alpha,h_v,h_{\hat\eta}) &= \frac{1}{n}\sum_{i=1}^{n}\left[Y_i-\hat Y_i+\hat Y_i-\hat Y_{2i}-\left\{X_i-\hat X_i+\hat X_i-\hat X_{2i}\right\}'\hat\beta\right]^2\\ &\equiv \frac{1}{n}\sum_{i=1}^{n}\left[Y_i-\hat Y_i-\hat\delta_{Y,i}-\left\{X_i-\hat X_i-\hat\delta_{X,i}\right\}'\hat\beta\right]^2\\ &= \hat J^*(\alpha,h_v,h_\eta)+R_1(\alpha,h_v,h_\eta,h_z), \qquad (5.A.1)\end{aligned}$$

where $\hat Y_{2i}=\hat m_y(V_i,\hat\eta_i)+\hat W_{2i}$ and $\hat X_{2i}=\hat m_x(V_i,\hat\eta_i)+\hat U_{2i}$ with $\hat W_{2i}=\frac{\sum_{j\neq i}W_jL_{2,ij}}{\sum_{j\neq i}W_jL_{2,ij}}$, $\hat U_{2i}=\frac{\sum_{j\neq i}U_jL_{2,ij}}{\sum_{j\neq i}W_jL_{2,ij}}$ and $L_{2,ij}=L_{h_v,h_\eta}(V_i-V_j,\hat\eta_i-\hat\eta_j)$, and $\hat\delta_{Y,i}=\hat Y_{2i}-\hat Y_i$, $\hat\delta_{X,i}=\hat X_{2i}-\hat X_i$ with $\hat Y_i=\hat m_y(V_i,\eta_i)+\hat W_i$, $\hat X_i=\hat m_x(V_i,\eta_i)+\hat U_i$ with $\hat W_i=\frac{\sum_{j\neq i}W_jL_{ij}}{\sum_{j\neq i}L_{il}}$, $\hat U_i=\frac{\sum_{j\neq i}^{n}U_jL_{ij}}{\sum_{j\neq i}^{n}L_{il}}$ and $L_{ij}=L_{h_v,h_\eta}(V_i-V_j,\eta_i-\eta_j)$. Let us note that $m=m(v_0,\eta)$ and $\tilde m=E(m|\alpha)$. Note that the term in the last equation of (5.A.1), $\hat J^*(\alpha,h_v,h_\eta)$, is further expanded, as shown below:

$$\hat J^*(\alpha,h_v,h_\eta)=\frac{1}{n}\sum_{i=1}^{n}\left[Y_i-\hat Y_i-\left\{X_i-\hat X_i\right\}'\hat\beta\right]^2=\tilde J(\alpha)+T(h_v,h_\eta)+R_2(\alpha,h_v,h_\eta),$$

where:

$$\begin{aligned}&R_2(\alpha,h_v,h_\eta)=(\hat\beta-\beta_0)'S_{\tilde m_x-\hat m_x}(\beta_0-\hat\beta)+(\hat\beta-\beta_0)'S_{\hat U}(\beta_0-\hat\beta)\\ &-2(\hat\beta-\beta_0)'S_{m_x-\tilde m_x,\tilde m_x-\hat m_x}(\beta_0-\hat\beta)-\ 2(\hat\beta-\beta_0)'S_{m_x-\tilde m_x,\hat U}(\beta_0-\hat\beta)\\ &+2(\hat\beta-\beta_0)'S_{\tilde m_x-\hat m_x,U}(\beta_0-\hat\beta)-2(\hat\beta-\beta_0)'S_{\tilde m_x-\hat m_x,\hat U}(\beta_0-\hat\beta)\\ &-\ 2(\hat\beta-\beta_0)'S_{U\hat U}(\beta_0-\hat\beta)+S_{\tilde m-\hat m}+2S_{m-\tilde m,\tilde m-\hat m}-2S_{e\hat e}+S_{\hat e}-2(\hat\beta-\beta_0)'S_{\tilde m_x-\hat m_x,m-\tilde m}\\ &+\ 2(\hat\beta-\beta_0)'S_{\hat U,m-\tilde m}-2(\hat\beta-\beta_0)'S_{m_x-\tilde m_x,\tilde m-\hat m}-2(\hat\beta-\beta_0)S_{\tilde m_x-\hat m_x,\tilde m-\hat m}-2(\hat\beta-\beta_0)'S_{U,\tilde m-\hat m}\\ &+\ 2(\hat\beta-\beta_0)'S_{\hat U,\tilde m-\hat m}-2(\hat\beta-\beta_0)'S_{\tilde m_x-\hat m_x,e}+2(\hat\beta-\beta_0)'S_{\hat Ue}+2(\hat\beta-\beta_0)'S_{U\hat e}\\ &+2(\hat\beta-\beta_0)'S_{m_x-\tilde m_x,\hat e}+\ 2(\hat\beta-\beta_0)'S_{\tilde m_x-\hat m_x,\hat e}+2(\hat\beta-\beta_0)'S_{U\hat e}-2(\hat\beta-\beta_0)S_{\hat U\hat e}+2S_{\tilde m-\hat m,e}\\ &-2S_{m-\tilde m,\hat e}-2S_{\tilde m-\hat m,\hat e}-\ S_{m-\hat m_0}\end{aligned}$$

with $\tilde Y_i=\tilde m_{y,i}$ and $\tilde X_i=\tilde m_{x,i}$ since $E(w|x,\eta)=0$ and $E(u|x,\eta)=0$, and $\hat m_0=\frac{\sum_{j\neq i}m_jL_{0,ij}}{\sum_{j\neq i}L_{0,ij}}$, with $L_{0,ij}=L_{h_v,h_\eta}(V_{0i}-V_{0j},\eta_i-\eta_j)$. The results of $\sup_{\alpha\in\mathcal A_n,h_v,h_\eta\in\mathcal H_n}|R_2(\alpha,h_v,h_\eta)|=$

$o_p(n^{-1/2})$ are easily shown using the fact that $\beta_0 - \hat{\beta} = o_p(n^{-1/2})$ shown below, and Propositions 5.A.7, 5.A.8, 5.A.9, 5.A.12, 5.A.13, 5.A.15, 5.A.18, 5.A.19, and 5.A.20. The last term in R_2 is $S_{m-\hat{m}_0} = O_p(n^{-1}h_v^{-1}h_\eta^{-q_2}) + O_p((h_v^2 + h_\eta^2)^2)$ by a simple nonparametric analysis. This is a simple extension of the results in Xia et al. (1999). Hence the objective function (5.A.1) is rewritten as:

$$\hat{J}(\alpha, h_v, h_{\hat{\eta}}) = \tilde{J}(\alpha) + T(h_v, h_\eta) + R_1(\alpha, h_v, h_\eta, h_z) + o_p(n^{-1/2}),$$

where:

$$\begin{aligned}
&R_1(\alpha, h_z, h_v, h_\eta) = (\hat{\beta} - \beta_0)' S_{\hat{\delta}_X}(\hat{\beta} - \beta_0) + S_{\hat{\delta}_m} + S_{\hat{\delta}_e} + 2(\hat{\beta} - \beta_0)' S_{\hat{\delta}_X \hat{\delta}_m} \\
&+2(\hat{\beta} - \beta_0)' S_{\hat{\delta}_X \hat{\delta}_e} - 2S_{\hat{\delta}_m \hat{\delta}_e} - 2(\hat{\beta} - \beta_0)' S_{\hat{\delta}_X, m_x - \tilde{m}_x}(\hat{\beta} - \hat{\beta}_0) \\
&-2(\hat{\beta} - \beta_0)' S_{\hat{\delta}_X, \tilde{m}_x - \hat{m}_x}(\hat{\beta} - \beta_0) - 2(\hat{\beta} - \beta_0)' S_{\hat{\delta}_X U}(\hat{\beta} - \beta_0) \\
&+ 2(\hat{\beta} - \beta_0)' S_{\hat{\delta}_X \hat{U}}(\hat{\beta} - \beta) + 2(\hat{\beta} - \beta_0)' S_{\hat{\delta}_X, m - \tilde{m}} + 2(\hat{\beta} - \beta_0)' S_{\hat{\delta}_X, \tilde{m} - \hat{m}} \\
&+2(\hat{\beta} - \beta_0)' S_{\hat{\delta}_X e} - 2(\hat{\beta} - \beta_0)' S_{\hat{\delta}_X \hat{e}} + 2(\hat{\beta} - \beta_0)' S_{\hat{\delta}_m, m_x - \tilde{m}_x} \\
&+2(\hat{\beta} - \beta_0) S_{\hat{\delta}_m, \tilde{m}_x - \hat{m}_x} + 2(\hat{\beta} - \beta_0)' S_{\hat{\delta}_m U} - 2(\hat{\beta} - \beta_0)' S_{\hat{\delta}_m \hat{U}} - 2S_{\hat{\delta}_m, m - \tilde{m}} \\
&-2S_{\hat{\delta}_m, \tilde{m} - \hat{m}} - 2S_{\hat{\delta}_m e} + 2S_{\hat{\delta}_m \hat{e}} + 2(\hat{\beta} - \beta_0) S_{\hat{\delta}_e, m_x - \tilde{m}_x} + 2(\hat{\beta} - \beta_0)' S_{\hat{\delta}_e, \tilde{m}_x - \hat{m}_x} \\
&+2(\hat{\beta} - \beta_0) S_{\hat{\delta}_e U} - 2(\hat{\beta} - \beta_0)' S_{\hat{\delta}_e \hat{U}} - 2S_{\hat{\delta}_e, m - \tilde{m}} - 2S_{\hat{\delta}_e, \tilde{m} - \hat{m}} - 2S_{\hat{\delta}_e e} + 2S_{\hat{\delta}_e \hat{e}}.
\end{aligned}$$

In particular, we show that $\sup_{\alpha \in A_n, h_v, h_\eta, h_z \in \mathcal{H}_n} |R_1(\alpha, h_z, h_v, h_\eta)| = o_p(n^{-1/2})$ by using the fact that $\hat{\beta} = \beta_0 + o_p(n^{-1/2})$ and Propositions 5.A.10, 5.A.11, 5.A.14, 5.A.16, 5.A.17 and 5.A.21 below. Hence we have:

$$\hat{J}(\alpha, h_v, h_{\hat{\eta}}) = \tilde{J}(\alpha) + T(h_v, h_\eta) + o_p(n^{-1/2}).$$

■

Step 2. Proofs of Corollary 3.2.2: We can now present the proofs of asymptotic properties of $\hat{\alpha}$ and $\hat{\beta}$. In view of the representation of $||\alpha - \alpha_0|| \leq Cn^{-1/2}$, we may write, for bounded values of x:

$$m(v_0, \eta) = m(v, \eta) - x'(\alpha - \alpha_0) m_0^{(1)} + O(n^{-1}), \quad (5.A.2)$$

$$m(v_0, \eta | v, \eta) = m(v, \eta) - m_x(x | v, \eta)'(\alpha - \alpha_0) m_0^{(1)} + O(n^{-1}), \quad (5.A.3)$$

where $m_x(x|v,\eta) = E(X_{\mathcal{A}_x}|v,\eta)$. Firstly, let us consider the asymptotic properties of $\hat{\alpha}$. Using (5.A.2) and (5.A.3), we have the expansion of $\tilde{J}(\alpha)$ below:

$$\begin{aligned}
\tilde{J}(\alpha) &= \frac{1}{n}\sum_{i=1}^{n}\Big[m_i - \tilde{m}_i + U_i'(\beta_0 - \hat{\beta}) + e_i\Big]^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\{m_i - \tilde{m}_i\}^2 + \frac{2}{n}\sum_{i=1}^{n}\{m_i - \tilde{m}_i\}e_i + \frac{2}{n}\sum_{i=1}^{n}\{m_i - \tilde{m}_i\}U_i'(\beta_0 - \hat{\beta}) \\
&+ \text{ terms independent of } \alpha + o_p(n^{-1/2}) \\
&= (\alpha_0 - \alpha)'\left[\frac{1}{n}\sum_{i=1}^{n}\left\{m_0^{(1)}\right\}^2 U_iU_i'\right](\alpha_0 - \alpha) + 2\frac{1}{n}\sum_{i=1}^{n} e_i m_0^{(1)}U_i'(\alpha_0 - \alpha) \\
&+ 2(\beta_0 - \hat{\beta})'\left[\frac{1}{n}\sum_{i=1}^{n} m_0^{(1)}U_iU_i'\right](\alpha_0 - \alpha) + o_p(n^{-1/2}) \\
&= (\alpha_0 - \alpha)'\left[\frac{1}{n}\sum_{i=1}^{n}\left\{m_0^{(1)}\right\}^2 U_{0i}U_{0i}'\right](\alpha_0 - \alpha) + 2\frac{1}{n}\sum_{i=1}^{n} e_i m_0^{(1)}U_{0i}'(\alpha_0 - \alpha) \\
&+ 2(\beta_0 - \hat{\beta})'\left[\frac{1}{n}\sum_{i=1}^{n} m_0^{(1)}U_{0i}U_{0i}'\right](\alpha_0 - \alpha) + o_p(1) + O_p(n^{-1/2}), \qquad (5.A.4)
\end{aligned}$$

where $U_{0i} = \{X_i - E(X_i|V_{0i},\eta_i)\}$.

Given α_0, $(\beta_0 - \hat{\beta}) = -\left(\frac{1}{n}\sum_{i=1}^{n} U_{0i}U_{0i}'\right)^{-}\frac{1}{n}\sum_{i=1}^{n} U_{0i}e_i$ (see the last equation of (5.A.8) below). Hence we have:

$$\begin{aligned}
\tilde{J}(\alpha) &= (\alpha_0 - \alpha)'\left[\frac{1}{n}\sum_{i=1}\left\{m_0^{(1)}\right\}^2 U_{0i}U_{0i}'\right](\alpha_0 - \alpha) + 2\frac{1}{n}\sum_{i=1}^{n} e_i m_0^{(1)}U_{0i}'(\alpha_0 - \alpha) \\
&- 2\left[\left(\frac{1}{n}\sum_{i=1}^{n} U_{0i}U_{0i}'\right)^{-}\frac{1}{n}\sum_{i=1}^{n} e_iU_{0i}'\right]\left[\frac{1}{n}\sum_{i=1}^{n} m_0^{(1)}U_{0i}U_{0i}'\right](\alpha_0 - \alpha) + o_p(1) + O_p(n^{-1/2}) \\
&= (\alpha_0 - \alpha)'\left\{m_0^{(1)}\right\}^2 S_{U_0}(\alpha_0 - \alpha) \\
&+ 2m_0^{(1)}S_{eU_0}(\alpha_0 - \alpha) - 2\left\{(S_{U_0})^{-} S_{eU_0}\right\}\left\{m_0^{(1)}S_{U_0}(\alpha_0 - \alpha)\right\} + o_p(1).
\end{aligned}$$

Given $\hat{\eta}$ and α, we write the linear reduced form from Robinson (1988) as follows:

$$Y_i - \hat{Y}_{3i} = (X_i - \hat{X}_{3i})'\beta_0 + (m_i - \hat{m}_{3i}) + (e_i - \hat{e}_{3i}), \qquad (5.A.5)$$

where $\hat{Y}_{3i} = \hat{m}_y(\hat{V}_i,\hat{\eta}_i) + \hat{W}_{3i}$, $\hat{X}_{3i} = \hat{m}_x(\hat{V}_i,\hat{\eta}_i) + \hat{U}_{3i}$, $\hat{m}_{3i} = \frac{\sum_{j=1}^{n}\tilde{m}_jL_{3,ij}}{\sum_{l=1}^{n}L_{3,il}}$, $\hat{e}_{3i} = \frac{\sum_{j=1}^{n}e_jL_{3,ij}}{\sum_{l=1}^{n}L_{3,il}}$ with $\hat{W}_{3i} = \frac{\sum_{j=1}^{n}W_jL_{3,ij}}{\sum_{l=1}^{n}L_{3,il}}$ and $\hat{U}_{3i} = \frac{\sum_{j=1}^{n}U_jL_{3,ij}}{\sum_{l=1}^{n}L_{3,il}}$ with $L_{3,ij} = L_{h_v,h_\eta}(\hat{V}_i - \hat{V}_j, \hat{\eta}_i - \hat{\eta}_j)$.

Hence from (5.A.5), we obtain:

$$\hat{\beta} - \beta_0 = S^{-1}_{X-\hat{X}_3}\left(S_{X-\hat{X}_3,m-\hat{m}_3} + S_{X-\hat{X}_3,e-\hat{e}_3}\right). \qquad (5.A.6)$$

We further decompose (5.A.5), as shown below:

$$\begin{aligned}
Y_i - \hat{Y}_{1i} + \hat{Y}_{1i} - \hat{Y}_{3i} &= \left\{X_i - \hat{X}_{1i} + \hat{X}_{1i} - \hat{X}_{3i}\right\}'\beta_0 + m_i - \hat{m}_{1i} + \hat{m}_{1i} - \hat{m}_{3i} \\
&+ e_i - \hat{e}_{1i} + \hat{e}_{1i} - \hat{e}_{3i} \\
Y_i - \hat{Y}_{1i} - \breve{\delta}_{y,i} &\equiv (X_i - \hat{X}_{1i} - \breve{\delta}_{x,i})'\beta_0 + (m_i - \hat{m}_{1i} - \breve{\delta}_{m,i}) + (e_i - \hat{e}_{1i} - \breve{\delta}_{e,i}) \\
Y_i - \tilde{Y}_i + \tilde{Y}_i - \hat{Y}_{1i} - \breve{\delta}_{y,i} &\equiv (X_i - \tilde{X}_i + \tilde{X}_i - \hat{X}_{1i} - \breve{\delta}_{x,i})'\beta_0 \\
&+ (m_i - \tilde{m}_i + \tilde{m}_i - \hat{m}_{1i} - \breve{\delta}_{mi}) + (e_i - \hat{e}_{1i} - \breve{\delta}_{e,i}). \qquad (5.A.7)
\end{aligned}$$

The last term of the right-hand side in (5.A.7) is from $E(e|x,\eta) = 0$, where $\breve{\delta}_{y,i} = \hat{Y}_{3i} - \hat{Y}_{1i}$, $\breve{\delta}_{x,i} = \hat{X}_{3i} - \hat{X}_{1i}$, $\breve{\delta}_{m,i} = \hat{m}_{3i} - \hat{m}_{1i}$, $\breve{\delta}_{e,i} = \hat{e}_{3i} - \hat{e}_{1i}$, $\hat{Y}_{1i} = \hat{m}_y(\hat{V}_i, \eta_i) + \hat{W}_{1i}$, $\hat{X}_{1i} = \hat{m}_x(\hat{V}_i, \eta_i) + \hat{U}_{1i}$, $\hat{m}_{1i} = \frac{\sum_{j=1}^n \tilde{m}_j L_{1,ij}}{\sum_{l=1}^n L_{1,il}}$, $\hat{e}_{1i} = \frac{\sum_{j=1}^n e_j L_{1,ij}}{\sum_{l=1}^n L_{il}}$, $\hat{W}_{1i} = \frac{\sum_{j=1}^n W_j L_{1,ij}}{\sum_{l=1}^n L_{1,il}}$ and $\hat{U}_{1i} = \frac{\sum_{j=1}^n U_j L_{1,ij}}{\sum_{l=1}^n L_{1,il}}$ with $L_{1,ij} = L_{h_v,h_\eta}(\hat{V}_i - \hat{V}_j, \eta_i - \eta_j)$. By utilising the decomposition in (5.A.7), we have:

$$\begin{aligned}
&S_{X-\hat{X}_3} = S_{X-\tilde{X}} + S_{\tilde{X}-\hat{X}_1} + S_{\breve{\delta}_X} + 2S_{X-\tilde{X},\tilde{X}-\hat{X}_1} - 2S_{X-\tilde{X},\breve{\delta}_X} - 2S_{\tilde{X}-\hat{X}_1,\breve{\delta}_X} \\
&= S_{m_x-\tilde{m}_x} + S_{\tilde{m}_x-\hat{m}_{x_1}} + S_U + S_{\hat{U}_1} + S_{\breve{\delta}_X} + 2S_{m_x-\tilde{m}_x,\tilde{m}_x-\hat{m}_{x_1}} + 2S_{m_x-\tilde{m}_x,U} \\
&-2S_{m_x-\tilde{m}_x,\hat{U}_1} - S_{m_x-\tilde{m}_x,\breve{\delta}_X} - 2S_{\tilde{m}_x-\hat{m}_{x_1},U} - 2S_{\tilde{m}_x-\hat{m}_{x_1},\hat{U}_1} - 2S_{m_x-\tilde{m}_x,\breve{\delta}_X} \\
&-2S_{U\hat{U}_1} - 2S_{U\breve{\delta}_X} + 2S_{\hat{U}_1\breve{\delta}_X}; \\
&S_{m-\hat{m}_3} = S_{m-\tilde{m}} + S_{\tilde{m}-\hat{m}_1} + S_{\breve{\delta}_m} + 2S_{m-\tilde{m},\tilde{m}-\hat{m}_1} - 2S_{m-\tilde{m},\breve{\delta}_m} - 2S_{\tilde{m}-\hat{m}_1,\breve{\delta}_m}; \\
&S_{e-\hat{e}_3} = S_e + S_{\hat{e}_1} + S_{\breve{\delta}_e} - 2S_{e\hat{e}_1} - 2S_{e\breve{\delta}_e} + 2S_{\hat{e}_1\breve{\delta}e}; \\
&S_{X-\hat{X}_3,m-\hat{m}_3} = S_{m_x-\tilde{m}_x,m-\tilde{m}} + S_{m_x-\tilde{m}_x,\tilde{m}-\hat{m}_1} - S_{m_x-\tilde{m}_x,\breve{\delta}_m} + S_{\tilde{m}_x-\hat{m}_{x_1},m-\tilde{m}} \\
&+S_{\tilde{m}_x-\hat{m}_{x_1},\tilde{m}-\hat{m}_1} - S_{\tilde{m}_x-\hat{m}_{x_1},\breve{\delta}_m} + S_{m-\tilde{m},U} + S_{\tilde{m}-\hat{m}_1,U} - S_{U\breve{\delta}_m} - S_{m-\tilde{m},\hat{U}_1} \\
&-S_{\tilde{m}-\hat{m}_1,\hat{U}_1} + S_{\hat{U}_1\breve{\delta}_m} - S_{m-\tilde{m},\breve{\delta}_X} - S_{\tilde{m}-\hat{m}_1,\breve{\delta}_X} + S_{\breve{\delta}_X\breve{\delta}_m}; \\
&S_{X-\hat{X}_3,e-\hat{e}_3} = S_{m_x-\tilde{m}_x,e} - S_{m_x-\tilde{m}_x,\hat{e}_1} - S_{m_x-\tilde{m}_x,\breve{\delta}e} + S_{\tilde{m}_x-\hat{m}_{x_1},e} - S_{\tilde{m}_x-\hat{m}_{x_1},\hat{e}_1} \\
&-S_{\tilde{m}_x-\hat{m}_{x_1},\breve{\delta}_e} + S_{Ue} - S_{U\hat{e}_1} - S_{U\breve{\delta}_e} - S_{\hat{U}_1e} + S_{\hat{U}_1\hat{e}_1} + S_{\hat{U}_1\breve{\delta}_e} - S_{\breve{\delta}_Xe} + S_{\breve{\delta}_X\hat{e}_1} + S_{\breve{\delta}X\breve{\delta}e}; \\
&S_{m-\hat{m}_3,e-\hat{e}_3} = S_{m-\tilde{m},e} - S_{m-\tilde{m},\hat{e}_1} - S_{m-\tilde{m},\breve{\delta}e} + S_{\tilde{m}-\hat{m}_1,e} - S_{\tilde{m}-\hat{m}_1,\hat{e}_1} - S_{\tilde{m}-\hat{m}_1,\breve{\delta}_e} \\
&-S_{\breve{\delta}_me} + S_{\breve{\delta}_m\hat{e}_1} + S_{\breve{\delta}_e\breve{\delta}_m}.
\end{aligned}$$

Note that we approximate two kernel functions to be $L_{3,ij} = L_{2,ij}+O_p(n^{-1/2}h_v^{-1})$ and $L_{1,ij} = L_{ij} + O_p(n^{-1/2}h_v^{-1})$ uniformly in i. Hence, we employ $L_{2,ij}$ and L_{ij} instead of $L_{3,ij}$ and $L_{1,ij}$, respectively, for the case of $\hat{\beta}$ in Propositions 5.A.7 to 5.A.21.

By Propositions 5.A.7 to 5.A.21, and (5.A.2) and (5.A.3), we obtain that (5.A.6) is:

$$\begin{aligned}
(\hat{\beta}-\beta_0) &= \left(\frac{1}{n}\sum_{i=1}^{n} U_iU_i'\right)^{-}\left\{\frac{1}{n}\sum_{i=1}^{n} U_ie_i - \frac{1}{n}\sum_{i=1}^{n} U_i(m_i-\tilde{m}_i)\right\} + o_p(n^{-1/2})\\
&= \left(\frac{1}{n}\sum_{i=1}^{n} U_iU_i'\right)^{-}\left\{\frac{1}{n}\sum_{i=1}^{n} U_ie_i - \frac{1}{n}\sum_{i=1}^{n} m_0^{(1)}U_iU_i'(\alpha_0-\alpha)\right\} + o_p(n^{-1/2})\\
&= \left(\frac{1}{n}\sum_{i=1}^{n} U_{0i}U_{0i}'\right)^{-}\left\{\frac{1}{n}\sum_{i=1}^{n} U_{0i}e_i - \frac{1}{n}\sum_{i=1}^{n} m_0^{(1)}U_{0i}U_{0i}'(\alpha_0-\alpha)\right\}\\
&\quad +o_p(1)+O_p(n^{-1/2}). \qquad (5.A.8)
\end{aligned}$$

Given β_0, $\alpha_0-\alpha = \left(\frac{1}{n}\sum_{i=1}^{n}\left\{m_0^{(1)}\right\}^2 U_{0i}U_{0i}'\right)^{-}\frac{1}{n}\sum_{i=1}^{n} m_0^{(1)}e_iU_{0i}'$ (see the last equation in (5.A.4)). Hence we have:

$$\begin{aligned}
&(\hat{\beta}-\beta_0)\\
&= \left(\frac{1}{n}\sum_{i=1}^{n} U_{0i}U_{0i}'\right)^{-}\left\{\frac{1}{n}\sum_{i=1}^{n} U_{0i}e_i - \frac{1}{n}\sum_{i=1}^{n} m_0^{(1)}U_{0i}U_{0i}'\left(\frac{1}{n}\sum_{i=1}^{n}\left\{m_0^{(1)}\right\}^2 U_{0i}U_{0i}'\right)^{-}\frac{1}{n}\sum_{i=1}^{n} m_0^{(1)}e_iU_{0i}'\right\}\\
&+o_p(1)\\
&= (S_{U_0})^{-}\left\{S_{U_0e} - m_0^{(1)}S_{U_0}\left(\left\{m_0^{(1)}\right\}^2 S_{U_0}\right)^{-} m_0^{(1)}S_{eU_0}\right\} + o_p(1).
\end{aligned}$$

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

$$\begin{aligned}
\hat{\sigma}^2 &= S_{e-\hat{e}_3} + S_{m-\hat{m}_3} + (\hat{\beta}-\beta_0)'S_{(X-\hat{X}_3)}(\hat{\beta}-\beta_0) - 2(\hat{\beta}-\beta_0)'S_{(X-\hat{X}_3)',e-\hat{e}_3}\\
&- 2(\hat{\beta}-\beta_0)'S_{(X-\hat{X}_3),m-\hat{m}_3} + 2S_{m-\hat{m}_3,e-\hat{e}_3} \qquad (5.A.9)\\
&= S_e + o_p(1) \xrightarrow{p} \sigma^2,
\end{aligned}$$

by Propositions 5.A.7 to 5.A.21 below, the law of large numbers, and the *i.i.d.* assumption of e_i . The other nine terms are $(\hat{\beta}-\beta_0)'S_{m_x-\tilde{m}_x}(\hat{\beta}-\beta_0)$;

$(\hat{\beta}-\beta_0)'S_{m_x-\tilde{m}_x,U}(\hat{\beta}-\beta_0)$; $(\hat{\beta}-\beta_0)'S_U(\hat{\beta}-\beta_0)$; $S_{m-\tilde{m}}$; $S_{m_x-\tilde{m}_x,m-\tilde{m}}$; $S_{m-\tilde{m},U}$; $S_{m_x-\tilde{m}_x,e}$; S_{Ue} and $S_{m-\tilde{m},e}$ equal to $o_p(n^{-1/2})$.

By the central limit theorem and the law of large numbers, the asymptotic normalities of $\hat{\alpha}$ and $\hat{\beta}$ are:

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

■

Note that the stated orders of the remainder term $R_1(\alpha,h_v,h_\eta,h_z)$ are available uniformly in $\alpha\in A_n$ and $h_v,h_\eta,h_z\in\mathcal{H}_n$, using the uniform bounds in Härdle et al. (1993). Let $\varphi_n(\alpha,h_v,h_\eta,h_z)$ be a possible quantity for which we show that:

$$\sup_{\alpha\in A_n,h_v,h_\eta,h_z\in\mathcal{H}_n}|\varphi_n(\alpha,h_v,h_\eta,h_z)|=o_p(n^a), \tag{5.A.10}$$

since we have:

$$\sup_{\alpha\in A_n,h_v,h_\eta,h_z\in\mathcal{H}_n}E\left(\varphi_n(\alpha,h_v,h_\eta,h_z)/n^b\right)^{2l}=O(1), \tag{5.A.11}$$

for all integers $l\geq 1$ and where $b<a$. For details of the equations (5.A.10) and (5.A.11), see step (ii) of the proof section 4 in Härdle et al. (1993). For proofs of Propositions 5.A.7 to 5.A.21, we assume that $h_{\eta,1}=\cdots=h_{\eta,q_2}=h_\eta$ and $h_{z,1}=\cdots=h_{z,q_2}=h_z$ for expositional simplicity.

Proposition 5.A.7.

(i) $\sqrt{n}S_{\tilde{m}_x-\hat{m}_x}=O_p(n^{-1/2}h_v^{-1}h_\eta^{-q_2})+O_p(n^{1/2}(h_v^2+h_\eta^2)^2)$;

(ii) $\sqrt{n}S_{\tilde{m}-\hat{m}}=O_p(n^{-1/2}h_v^{-1}h_\eta^{-q_2})+O_p(n^{1/2}(h_v^2+h_\eta^2)^2)$.

Proof: *Let* $\varphi(\cdot)$ *denote* $m(\cdot)$ *and* $m_x(\cdot)$, *and* $\tilde{\varphi}(\cdot)$ *denote* $\tilde{m}(\cdot)$ *and* $\tilde{m}_x(\cdot)$. *We deduce from (5.A.2) and (5.A.3) that, uniformly in* i, *we have:*

$$\begin{aligned}
\tilde{\varphi}_i - \hat{\varphi}_i &= \frac{\sum_{j\neq i}\left\{\varphi(X_i'\alpha_0, \eta_i|v,\eta) - \varphi(X_j'\alpha_0,\eta_j)\right\} L_{ij}}{\sum_{j\neq i} L_{ij}} \\
&= \frac{\sum_{j\neq i}\left\{\tilde{\varphi}_i - \tilde{\varphi}_j + U_j'(\alpha-\alpha_0)\varphi_0^{(1)}\right\} L_{ij}}{\sum_{j\neq i} L_{ij}} + O(n^{-1}) \\
&= \frac{(nh_v h_\eta^{q_2})^{-1}\sum_{j\neq i}\left\{\tilde{\varphi}_i - \tilde{\varphi}_j + U_j'(\alpha-\alpha_0)\varphi_0^{(1)}\right\} L_{ij}}{f(v,\eta)}\left[1 - \frac{\hat{f}(v,\eta) - f(v,\eta)}{\hat{f}(v,\eta)}\right] + o(1),
\end{aligned}$$

where $\varphi_0^{(1)} = \partial\varphi(v_0,\eta)/\partial v_0$. *Note that since* $\left(\hat{f}(v,\eta) - f(v,\eta)\right)$ *is* $O_p(nh_v h_\eta^{q_2})^{-1/2} + O_p(h_v^2 + h_\eta^2)$ *so* $\left[1 - \frac{\hat{f}(v,\eta)-f(v,\eta)}{\hat{f}(v,\eta)}\right]$ *can be dropped, hence we consider only the numerator terms in the rest of the section.*

By identical distribution, $E(S_{\tilde{\varphi}-\hat{\varphi}}) = E\{(\tilde{\varphi}_i - \hat{\varphi}_i)^2\}$. *We can easily obtain those* $E(\tilde{\varphi}_i - \hat{\varphi}_i) = O(h_v^2 + h_\eta^2)$ *and* $Var(\tilde{\varphi}_i - \hat{\varphi}_i) = O(nh_v h_\eta^{q_2})^{-1}$, *where:*

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

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

$$Var\left(\frac{1}{nh_v h_\eta^{q_2}}\sum_{j\neq i}(\tilde{\varphi}_i - \tilde{\varphi}_j)L_{ij}\right) = O(nh_v h_\eta^{q_2})^{-1},$$

$$Var\left(\frac{1}{nh_v h_\eta^{q_2}}\sum_{j\neq i}U_j'(\alpha-\alpha_0)\varphi_0^{(1)}\right) = O(n^2 h_v h_\eta^{q_2})^{-1},$$

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

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

Proposition 5.A.8.

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

Proof: *Proposition 5.A.7 (i) and (ii), and the Cauchy inequality provide the proof.* ■

Proposition 5.A.9.

(i) $\sqrt{n}S_{\hat{U}} = O_p(n^{-1/2}h_v^{-1}h_\eta^{-q_2})$;

(ii) $\sqrt{n}S_{\hat{e}} = O_p(n^{-1/2}h_v^{-1}h_\eta^{-q_2})$.

Proof: *Let* ϱ_i *denote* U_i *and* e_i, *and* $E(\varrho_i|\mathcal{L}) = 0$ *almost surely, where* $\mathcal{L} = (X, \eta)$, *hence* $E(S_{\hat{\varrho}}) = E(\hat{\varrho}_i^2)$. *Then we have:*

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

■

Proposition 5.A.10.

(i) $\sqrt{n}S_{\hat{\delta}_X} = O_p(n^{-1}h_z^{-q_z/2}h_v^{-1/2}h_\eta^{-(q_2+2)/2}) + O_p(n^{1/2}h_z^2(h_v^2 + h_\eta^2))$;

(ii) $\sqrt{n}S_{\hat{\delta}_m} = O_p(n^{-1}h_z^{-q_z/2}h_v^{-1/2}h_\eta^{-(q_2+2)/2}) + O_p(n^{1/2}h_z^2(h_v^2 + h_\eta^2))$;

(iii) $\sqrt{n}S_{\hat{\delta}_e} = O_p(n^{-1}h_z^{-q_z/2}h_v^{-1/2}h_\eta^{-(q_2+2)/2}) + O_p(n^{1/2}h_z^2(h_v^2 + h_\eta^2))$.

Proof: *Let* δ *denote* δ_X, δ_m *and* δ_e.*Then we have:*

$$\hat{\delta}_i = \delta_{2,i} - \delta_{1,i} = \frac{\sum_{j\neq i}\delta_jL_{2,ij}}{\sum_{j\neq i}L_{2,ij}} - \frac{\sum_{j\neq i}\delta_jL_{ij}}{\sum_{j\neq i}L_{ij}}.$$

The Taylor expansion of the kernel function, $L_{2,ij}$, *is:*

$$L_{2,ij} = L_{ij} + L_{ij}^{(1)}\left(\frac{\triangle_{ij}}{h_\eta}\right) + L_{ij}^{(2)}(\tau)\left(\frac{\triangle_{ij}}{h_\eta}\right)^2,$$

where $L_{ij}^{(r)}$ *is the* r*th derivative of* L_{ij} *with respect to* η *with* $r = 1$ *or* 2, $\triangle_{ij} = \{\hat{m}_x(Z_j) - m_x(Z_j)\} - \{\hat{m}_x(Z_i) - m_x(Z_i)\}$ *and* τ *is between the segment line of* $\eta_j - \eta_i$ *and* $\hat{\eta}_j - \hat{\eta}_i$. *Hence, the denominator of* $\delta_{2,i}$ *is:*

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

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

$$\begin{aligned}
&E\left(\frac{1}{nh_v h_\eta^{q_2+1}}\sum_{j\neq i} L_{ij}^{(1)}\triangle_{ij}\right)^2\\
&=\frac{1}{n^2h_v^2h_\eta^{2(q_2+1)}}E\left(\sum_{j\neq i}\left(L_{ij}^{(1)}\right)^2\triangle_{ij}^2\right)+\frac{2}{n^2h_v^2h_\eta^{2(q_2+1)}}E\left(\sum_{j\neq i}\sum_{k\neq i,j}L_{ij}^{(1)}L_{ik}^{(1)}\triangle_{ij}\triangle_{ik}\right)\\
&=\frac{1}{n^4h_z^{2q_z}h_v^2h_\eta^{2(q_2+1)}}E\left(\sum_{j\neq i}\left(L_{ij}^{(1)}\right)^2\left\{\sum_{l\neq j}C_{(l,j;K)}-\sum_{l\neq i}C_{(l,i;K)}\right\}^2\right)\\
&+\frac{2}{n^4h_z^{2q_z}h_v^2h_\eta^{2(q_2+1)}}E\left(\sum_{j\neq i}\left(L_{ij}^{(1)}\right)^2\left\{\sum_{l\neq j}C_{(l,j;K)}-\sum_{l\neq i}C_{(l,i;K)}\right\}\left\{\sum_{m\neq j,l}C_{(m,j;K)}-\sum_{m\neq i,l}C_{(m,i;K)}\right\}\right)\\
&+\frac{2}{n^4h_z^{2q_z}h_v^2h_\eta^{2(q_2+1)}}E\left(\sum_{j\neq i}\sum_{k\neq i,j}L_{ij}^{(1)}L_{ik}^{(1)}\left\{\sum_{l\neq j}C_{(l,j;K)}-\sum_{l\neq i}C_{(l,i;K)}\right\}\right.\\
&\left.\times\left\{\sum_{m\neq k,l}C_{(m,k;K)}-\sum_{m\neq i,l}C_{(m,i;K)}\right\}\right)\\
&=O\left(n^{-2}h_z^{-q_z}h_v^{-1}h_\eta^{-(q_2+2)}\right)+O\left(n^{-1}h_z^4h_v^{-1}h_\eta^{-(q_2+2)}\right)+O\left(h_z^4(h_v^2+h_\eta^2)^2\right),
\end{aligned}\tag{5.A.12}$$

where $C_{(l,j;K)}=\{m_x(Z_l)-m_x(Z_j)\}K_{jl}$. Hence $\hat{\delta}_i=\frac{(nh_vh_\eta^{q_2+1})^{-1}\sum_{j\neq i}\delta_j L_{ij}^{(1)}\triangle_{ij}}{(nh_vh_\eta)^{-1}\sum_{j\neq i}L_{ij}+o_p(n^{-1/2})}$.

Now consider $E(\sqrt{n}S_{\hat{\delta}})$, we have:

$$E(\sqrt{n}S_{\hat{\delta}})=\frac{1}{n}\sum_{i=1}^{n}E(\hat{\delta}_i^2)+\frac{2}{n}\sum_{i=1}^{n}\sum_{j=1,\neq i}^{n}E(\hat{\delta}_i\hat{\delta}_j).\tag{5.A.13}$$

Using a similar argument to the above, the two terms in the right-hand side of (5.A.13) are:

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

and

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

■

Proposition 5.A.11.

(i) $\sqrt{n} S_{\hat{\delta}_X \hat{\delta}_m} = O_p(n^{-1} h_z^{-q_z/2} h_v^{-1/2} h_\eta^{-(q_2+2)/2}) + O_p(n^{1/2} h_z^2 (h_v^2 + h_\eta^2))$;

(ii) $\sqrt{n} S_{\hat{\delta}_X \hat{\delta}_e} = O_p(n^{-1} h_z^{-q_z/2} h_v^{-1/2} h_\eta^{-(q_2+2)/2}) + O_p(n^{1/2} h_z^2 (h_v^2 + h_\eta^2))$;

(iii) $\sqrt{n} S_{\hat{\delta}_m \hat{\delta}_e} = O_p(n^{-1} h_z^{-q_z/2} h_v^{-1/2} h_\eta^{-(q_2+2)/2}) + O_p(n^{1/2} h_z^2 (h_v^2 + h_\eta^2))$.

Proof: *Proposition 5.A.10 (i), (ii) and (iii), and the Cauchy inequality provide the proof.* ■

Proposition 5.A.12.

(i) $\sqrt{n} S_{U\hat{U}} = O_p(n^{-1/2} h_v^{-1/2} h_\eta^{-q_2/2})$;

(ii) $\sqrt{n}S_{\hat{U}e} = O_p(n^{-1/2}h_v^{-1/2}h_\eta^{-q_2/2})$;

(iii) $\sqrt{n}S_{e\hat{e}} = O_p(n^{-1/2}h_v^{-1/2}h_\eta^{-q_2/2})$;

(iv) $\sqrt{n}S_{U\hat{e}} = O_p(n^{-1/2}h_v^{-1/2}h_\eta^{-q_2/2})$.

Proof: *Since* $E(\varrho_i|\mathcal{L}) = 0$, *we have:*

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

where:

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

■

Proposition 5.A.13.

(i) $\sqrt{n}S_{\tilde{m}_x-\hat{m}_x,U} = O_p(n^{-1/2}h_v^{-1/2}h_\eta^{-q_2/2})$;

(ii) $\sqrt{n}S_{\tilde{m}-\hat{m},U} = O_p(n^{-1/2}h_v^{-1/2}h_\eta^{-q_2/2})$;

(iii) $\sqrt{n}S_{\tilde{m}_x-\hat{m}_x,e} = O_p(n^{-1/2}h_v^{-1/2}h_\eta^{-q_2/2})$;

(iv) $\sqrt{n}S_{\tilde{m}-\hat{m},e} = O_p(n^{-1/2}h_v^{-1/2}h_\eta^{-q_2/2})$.

Proof: *Since* $E(\varrho_i|\mathcal{L}) = 0$, *we have:*

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

where:

$$\begin{aligned} E\left\{\varrho_i^2(\tilde{\varphi}_i-\hat{\varphi}_i)^2\right\} &= \frac{1}{n^2h_v^2h_\eta^{2q_2}}E\left(\varrho_i^2\sum_{j\neq i}\left(C^*_{(i,j;L)}\right)^2\right) + \frac{2}{n^2h_v^2h_\eta^{2q_2}}E\left(\varrho_i^2\sum_{j\neq i}\sum_{l\neq i,j}C^*_{(i,j;L)}C^*_{(i,l;L)}\right) \\ &= O(n^{-1}h_v^{-1}h_\eta^{-q_2}) + O((h_v^2+q_2h_\eta^2)^2) \end{aligned}$$

with $C^*_{(i,j;L)} = \left\{\tilde{\varphi}_i - \tilde{\varphi}_j + U_j'(\alpha-\alpha_0)\varphi_0^{(1)}\right\}L_{ij}$. ■

Proposition 5.A.14.

(i) $\sqrt{n}S_{U\hat{\delta}_X} = O_p(n^{-1}h_z^{-q_z/2}h_v^{-1/2}h_\eta^{-(q_2+2)/2}) + O_p(n^{-1/2}h_z^2h_v^{-1/2}h_\eta^{-(q_2+2)/2})$;

(ii) $\sqrt{n}S_{e\hat{\delta}_e} = O_p(n^{-1}h_z^{-q_z/2}h_v^{-1/2}h_\eta^{-(q_2+2)/2}) + O_p(n^{-1/2}h_z^2h_v^{-1/2}h_\eta^{-(q_2+2)/2})$;

(iii) $\sqrt{n}S_{e\hat{\delta}_m} = O_p(n^{-1}h_z^{-q_z/2}h_v^{-1/2}h_\eta^{-(q_2+2)/2}) + O_p(n^{-1/2}h_z^2h_v^{-1/2}h_\eta^{-(q_2+2)/2})$;

(iv) $\sqrt{n}S_{e\hat{\delta}_X} = O_p(n^{-1}h_z^{-q_z/2}h_v^{-1/2}h_\eta^{-(q_2+2)/2}) + O_p(n^{-1/2}h_z^2h_v^{-1/2}h_\eta^{-(q_2+2)/2})$;

(v) $\sqrt{n}S_{U\hat{\delta}_m} = O_p(n^{-1}h_z^{-q_z/2}h_v^{-1/2}h_\eta^{-(q_2+2)/2}) + O_p(n^{-1/2}h_z^2h_v^{-1/2}h_\eta^{-(q_2+2)/2})$;

(vi) $\sqrt{n}S_{U\hat{\delta}_e} = O_p(n^{-1}h_z^{-q_z/2}h_v^{-1/2}h_\eta^{-(q_2+2)/2}) + O_p(n^{-1/2}h_z^2h_v^{-1/2}h_\eta^{-(q_2+2)/2})$.

Proof: *Since $E(\varrho_i|\mathcal{L}) = 0$, we have:*

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

where:

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

using similar arguments to those in Proposition 5.A.10. ■

Proposition 5.A.15.

(i) $\sqrt{n}S_{\tilde{m}_x-\hat{m}_x,\hat{U}} = O_p(nh_vh_\eta^{q_2})^{-1} + O_p(n^{1/2}(h_v^2+h_\eta^2)^2)$;

(ii) $\sqrt{n}S_{\tilde{m}-\hat{m},\hat{U}} = O_p(nh_vh_\eta^{q_2})^{-1} + +O_p(n^{1/2}(h_v^2+h_\eta^2)^2)$;

(iii) $\sqrt{n}S_{\tilde{m}_x-\hat{m}_x,\hat{e}} = O_p(nh_vh_\eta^{q_2})^{-1} + O_p(n^{1/2}(h_v^2+h_\eta^2)^2)$;

(iv) $\sqrt{n}S_{\tilde{m}-\hat{m},\hat{e}} = O_p(nh_v h_\eta^{q_2})^{-1} + O_p(n^{1/2}(h_v^2+h_\eta^2)^2)$.

Proof:

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

where:

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

■

Proposition 5.A.16.

(i) $\sqrt{n}S_{\hat{U}\hat{\delta}_X} = O_p\left(n^{-1}h_z^{-q_z/2}h_v^{-1}h_\eta^{-(q_2+1)}\right) + O_p\left(n^{1/2}h_z^2(h_v^2+h_\eta^2)\right)$;

(ii) $\sqrt{n}S_{\hat{e}\hat{\delta}_e} = O_p\left(n^{-1}h_z^{-q_z/2}h_v^{-1}h_\eta^{-(q_2+1)}\right) + O_p\left(n^{1/2}h_z^2(h_v^2+h_\eta^2)\right)$;

(iii) $\sqrt{n}S_{\hat{U}\hat{\delta}_m} = O_p\left(n^{-1}h_z^{-q_z/2}h_v^{-1}h_\eta^{-(q_2+1)}\right) + O_p\left(n^{1/2}h_z^2(h_v^2+h_\eta^2)\right)$;

(iv) $\sqrt{n}S_{\hat{e}\hat{\delta}_X} = O_p\left(n^{-1}h_z^{-q_z/2}h_v^{-1}h_\eta^{-(q_2+1)}\right) + O_p\left(n^{1/2}h_z^2(h_v^2+h_\eta^2)\right)$;

(v) $\sqrt{n}S_{\hat{e}\hat{\delta}_m} = O_p\left(n^{-1}h_z^{-q_z/2}h_v^{-1}h_\eta^{-(q_2+1)}\right) + O_p\left(n^{1/2}h_z^2(h_v^2+h_\eta^2)\right)$;

(vi) $\sqrt{n}S_{\hat{U}\hat{\delta}_e} = O_p\left(n^{-1}h_z^{-q_z/2}h_v^{-1}h_\eta^{-(q_2+1)}\right) + O_p\left(n^{1/2}h_z^2(h_v^2+h_\eta^2)\right)$.

Proof:

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

where:

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

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

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

$$\left.\times\left\{\sum_{m\neq l,k} C_{(m,l;K)} - \sum_{m\neq i,k} C_{(m,i;K)}\right\}\right)$$

$$= O\left(n^{-2}h_z^{-q_z}h_v^{-2}h_\eta^{-(2q_2+2)}\right) + O\left(n^{-2}h_z^4 h_v^{-2}h_\eta^{(-2q_2+2)}\right),$$

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

$$E\left(\sum_{j\neq i}\sum_{s\neq j}\sum_{l\neq i}\sum_{t\neq j}\varrho_j\varrho_s L_{ij}L_{js}\delta_l\delta_t L_{il}^{(1)}L_{jt}^{(1)}\left\{\sum_{k\neq l} C_{(k,l;K)} - \sum_{k\neq i} C_{(k,i;K)}\right\}\right.$$

$$\left.\times\left\{\sum_{m\neq t} C_{(m,t;K)} - \sum_{m\neq j} C_{(m,j;K)}\right\}\right).$$

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

■

Proposition 5.A.17.

(i) $\sqrt{n}S_{\tilde{m}_x-\hat{m}_x,\hat{\delta}_X} = O_p\left(n^{-1}h_z^{-q_z/2}h_v^{-1}h_\eta^{-(q_2+1)}\right) + O_p\left(n^{1/2}h_z^2(h_v^2+h_\eta^2)^2\right)$;

(ii) $\sqrt{n}S_{\tilde{m}_x-\hat{m}_x,\hat{\delta}_m} = O_p\left(n^{-1}h_z^{-q_z/2}h_v^{-1}h_\eta^{-(q_2+1)}\right) + O_p\left(n^{1/2}h_z^2(h_v^2+h_\eta^2)^2\right)$;

(iii) $\sqrt{n}S_{\tilde{m}_x-\hat{m}_x,\hat{\delta}_e} = O_p\left(n^{-1}h_z^{-q_z/2}h_v^{-1}h_\eta^{-(q_2+1)}\right) + O_p\left(n^{1/2}h_z^2(h_v^2+h_\eta^2)^2\right)$;

(iv) $\sqrt{n}S_{\tilde{m}-\hat{m},\hat{\delta}_m} = O_p\left(n^{-1}h_z^{-q_z/2}h_v^{-1}h_\eta^{-(q_2+1)}\right) + O_p\left(n^{1/2}h_z^2(h_v^2+h_\eta^2)^2\right)$;

(v) $\sqrt{n}S_{\tilde{m}-\hat{m},\hat{\delta}_X} = O_p\left(n^{-1}h_z^{-q_z/2}h_v^{-1}h_\eta^{-(q_2+1)}\right) + O_p\left(n^{1/2}h_z^2(h_v^2+h_\eta^2)^2\right)$;

(vi) $\sqrt{n}S_{\tilde{m}-\hat{m},\hat{\delta}_e} = O_p\left(n^{-1}h_z^{-q_z/2}h_v^{-1}h_\eta^{-(q_2+1)}\right) + O_p\left(n^{1/2}h_z^2(h_v^2+h_\eta^2)^2\right)$.

Proof:

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

where:

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

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

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

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

Proposition 5.A.18.

$\sqrt{n}S_{\hat{U}\hat{e}} = O_p(nh_v h_\eta^{q_2})^{-1} + O_p(n^{1/2}(h_v^2+h_\eta^2)^2)$.

Proof:

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

where:

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

and:

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

■

Proposition 5.A.19.

(i) $\sqrt{n}S_{m_x-\tilde{m}_x,\tilde{m}_x-\hat{m}_x} = O_p\left(n^{-1}h_v^{-1/2}h_\eta^{-q_2/2}\right)$;

(ii) $\sqrt{n}S_{m-\tilde{m},\tilde{m}-\hat{m}} = O_p\left(n^{-1}h_v^{-1/2}h_\eta^{-q_2/2}\right)$;

(iii) $\sqrt{n}S_{m_x-\tilde{m}_x,\tilde{m}-\hat{m}} = O_p\left(n^{-1}h_v^{-1/2}h_\eta^{-q_2/2}\right)$;

(iv) $\sqrt{n}S_{m-\tilde{m},\tilde{m}_x-\hat{m}_x} = O_p\left(n^{-1}h_v^{-1/2}h_\eta^{-q_2/2}\right)$.

Proof: *By (5.A.2) and (5.A.3) we deduce that, uniformly in i, we have:*

$$\varphi_i - \tilde{\varphi}_i = U_i'(\alpha_0-\alpha)\varphi_0^{(1)}(X_i'\alpha_0,\eta_i) + O(n^{-1}). \tag{5.A.14}$$

Hence we have:

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

where $t_i = U_i'(\alpha_0-\alpha)\varphi_0^{(1)}$.

For the rest of proofs, we use similar arguments to those in Proposition 5.A.13 because $E(U_i|\mathcal{L})=0$. Hence we have:

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

where:

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

■

Proposition 5.A.20.

(i) $\sqrt{n}S_{m_x-\tilde{m}_x,\hat{U}} = O_p\left(n^{-1}h_v^{-1/2}h_\eta^{-q_2/2}\right)$;

(ii) $\sqrt{n}S_{m_x-\tilde{m}_x,\hat{e}} = O_p\left(n^{-1}h_v^{-1/2}h_\eta^{-q_2/2}\right)$;

(iii) $\sqrt{n}S_{m-\tilde{m},\hat{U}} = O_p\left(n^{-1}h_v^{-1/2}h_\eta^{-q_2/2}\right)$;

(iv) $\sqrt{n}S_{m-\tilde{m},\hat{e}} = O_p\left(n^{-1}h_v^{-1/2}h_\eta^{-q_2/2}\right)$.

Proof: *By (5.A.14) and* $E(\varrho_i|\mathcal{L}) = 0$*, we use similar arguments to those in Proposition 5.A.12 for the rest of the proofs.*

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

where:

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

■

Proposition 5.A.21.

(i) $\sqrt{n}S_{m_x-\tilde{m}_x,\hat{\delta}_X} = O_p\left(n^{-2}h_z^{-q_z/2}h_v^{-1/2}h_\eta^{-(q_2+2)/2}\right)$;

(ii) $\sqrt{n}S_{m-\tilde{m},\hat{\delta}_m} = O_p\left(n^{-2}h_z^{-q_z/2}h_v^{-1/2}h_\eta^{-(q_2+2)/2}\right)$;

(iii) $\sqrt{n}S_{m-\tilde{m},\hat{\delta}_X} = O_p\left(n^{-2}h_z^{-q_z/2}h_v^{-1/2}h_\eta^{-(q_2+2)/2}\right)$;

(iv) $\sqrt{n}S_{m_x-\tilde{m}_x,\hat{\delta}_m} = O_p\left(n^{-2}h_z^{-q_z/2}h_v^{-1/2}h_\eta^{-(q_2+2)/2}\right)$;

(v) $\sqrt{n}S_{m_x-\tilde{m}_x,\hat{\delta}_e} = O_p\left(n^{-2}h_z^{-q_z/2}h_v^{-1/2}h_\eta^{-(q_2+2)/2}\right)$;

(vi) $\sqrt{n}S_{m-\tilde{m},\hat{\delta}_e} = O_p\left(n^{-2}h_z^{-q_z/2}h_v^{-1/2}h_\eta^{-(q_2+2)/2}\right)$.

Proof: *By (5.A.14) and* $E(U_i|\mathcal{L}) = 0$*, the rest of proofs is similar to that of Proposition 5.A.14.*

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

where:

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

■

3.5.2 Proof of Theorem 3.2.2

Given $\hat{\beta}$ and $\hat{\alpha}$, we have:

$$\hat{m}(\hat{v},\hat{\eta}_i) - m(v_0,\eta_i) = \left\{\hat{m}_{y**}(\hat{v},\eta_i) - \widetilde{m}_{y**}(\hat{v},\eta_i) + \widetilde{m}_{y**}(\hat{v},\eta_i) - m_{y**}(v_0,\eta_i) + \breve{\delta}_{y**,i}\right\}$$
$$- \left\{\hat{m}_x(\hat{v},\eta_i) - \widetilde{m}_x(\hat{v},\eta_i) + \widetilde{m}_x(\hat{v},\eta_i) - m_x(v_0,\eta_i) + \breve{\delta}_{x,i}\right\}' (\hat{\beta} - \beta_0), \tag{5.B.1}$$

where $Y_i^{**} = Y_i - X_i'\beta_0$, $\breve{\delta}_{y^{**}} = \breve{\delta}_y - \breve{\delta}_x'\beta_0$ and $\widetilde{m}(\hat{v},\eta) = E(m|\hat{\alpha})$. As the results of Section 3.5.1, the second term in the right-hand side of (5.B.1) is $o_p(n^{-1/2})$, uniformly in i, by applying (5.A.10) and (5.A.11) as $\sup_{X_i,\eta_i\in\mathcal{A},Z_i\in\mathcal{A}_z} |\varphi_i| = o_p(n^a)$ since $\sup_{X_i,\eta_i\in\mathcal{A},Z_i\in\mathcal{A}_z} E|\varphi_i/n^b|^{2l} = O(1)$. Hence (5.B.1) is:

$$\hat{m}(\hat{v},\hat{\eta}_i) - m(v_0,\eta_i) = \left\{\hat{m}_{y**}(\hat{v},\eta_i) - \widetilde{m}_{y**}(\hat{v},\eta_i) + \widetilde{m}_{y**}(\hat{v},\eta_i) - m_{y**}(v_0,\eta_i) + \breve{\delta}_{y**,i}\right\} + o_p(1), \tag{5.B.2}$$

where $\breve{\delta}_{y^{**}} = o_p(n^{-1/2})$ by similar arguments to those in Proposition 5.A.10 and $\widetilde{m}_{y^{**}}(\hat{v},\eta_i) - m_{y^{**}}(v_0,\eta_i) = O_p(n^{-1/2})$ by (5.A.2) and (5.A.3), uniformly in i. Hence (5.B.2) is:

$$\begin{aligned}\hat{m}(\hat{v},\hat{\eta}_i) - m(v_0,\eta_i) &= \hat{m}_{y^{**}}(\hat{v},\eta_i) - \widetilde{m}_{y^{**}}(\hat{v},\eta_i) + o_p(1) \\ &\equiv \hat{m}(\hat{v},\eta_i) - \widetilde{m}(\hat{v},\eta_i) + o_p(1),\end{aligned} \tag{5.B.3}$$

where:

$$\begin{aligned}
\hat{m}(\hat{v},\eta_i)-\widetilde{m}(\hat{v},\eta_i) &= \frac{\sum_{j\neq i}\{m(v_0,\eta_j)-\widetilde{m}(\hat{v},\eta_i)\}L_{1,ij}}{\sum_{j\neq i}L_{1,ij}}\\
&= \frac{\sum_{j\neq i}\{m(v_0,\eta_j)-m(v_0,\eta_i)\}L_{1,ij}}{\sum_{j\neq i}L_{1,ij}}+U_i'(\hat{\alpha}-\alpha_0)m_0^{(1)}+O(n^{-1})\\
&= \frac{\sum_{j\neq i}\{m(v_0,\eta_j)-m(v_0,\eta_i)\}\left\{L_{0,ij}+O(n^{-1/2}h_v^{-1})\right\}}{\sum_{j\neq i}L_{0,ij}+o(1)}+O_p(n^{-1/2}).
\end{aligned}$$

Hence (5.B.3) is:

$$\hat{m}(\hat{v},\hat{\eta}_i)-m(v_0,\eta_i)=\hat{m}(v_0,\eta_i)-m(v_0,\eta_i)+o_p(1). \quad (5.B.4)$$

Let us define $\check{m}(v_0,\eta_i)=\hat{m}(v_0,\eta_i)\hat{f}(v_0,\eta_i)$. Then we can rewrite the term in the right-hand side of (5.B.4) as follows:

$$\begin{aligned}
\hat{m}(v_0,\eta_i)-m(v_0,\eta_i) &= \frac{\check{m}(v_0,\eta_i)-m(v_0,\eta_i)\hat{f}(v_0,\eta_i)}{\hat{f}(v_0,\eta_i)}\\
&= \frac{\check{m}(v_0,\eta_i)-m(v_0,\eta_i)\hat{f}(v_0,\eta_i)}{f(v_0,\eta_i)}\left[1-\frac{\hat{f}(v_0,\eta_i)-f(v_0,\eta_i)}{f(v_0,\eta_i)}\right]. \quad (5.B.5)
\end{aligned}$$

First, we consider the bias term
$E(\hat{m}(v_0,\eta_i)-m(v_0,\eta_i))=f^{-1}(v_0,\eta_i)\left(E\check{m}(v_0,\eta_i)-m(v_0,\eta_i)E(\hat{f}(v_0,\eta_i))\right)$, where:

$$\begin{aligned}
E\check{m}(v_0,\eta_i) &= E\left[\frac{1}{nh_vh_\eta^{q_2}}\sum_{j=1}^{n}K_v\left(\frac{V_{0,j}-v_0}{h_v}\right)K_\eta\left(\frac{\eta_j-\eta_i}{h_\eta}\right)Y_j^{**}\right]\\
&= E\left[E_{v_0,\eta_i}\left\{\frac{1}{nh_vh_\eta^{q_2}}\sum_{j=1}^{n}K_v\left(\frac{V_{0,j}-v_0}{h_v}\right)K_\eta\left(\frac{\eta_j-\eta_i}{h_\eta}\right)Y_j^{**}\right\}\right]\\
&= E\left[\frac{1}{nh_vh_\eta^{q_2}}\sum_{j=1}^{n}K_v\left(\frac{V_{0,j}-v_0}{h_v}\right)K_\eta\left(\frac{\eta_j-\eta_i}{h_\eta}\right)m(V_{0,j},\eta_j)\right]\\
&= f(v_0,\eta_i)m(v_0,\eta_i)+\mathcal{K}_{v,2}h_v^2\left\{f_v^{(1)}(v_0,\eta_i)m_0^{(1)}(v_0)+f(v_0,\eta_i)m_0^{(2)}(v_0)\right\}\\
&+ \mathcal{K}_{\eta,2}\sum_{s=1}^{q_2}h_{\eta,s}^2\left\{f_{\eta,s}^{(1)}(v_0,\eta_i)m^{(1)}(\eta_{s,i})+f(v_0,\eta_i)m_{\eta,s}^{(2)}(\eta_i)\right\}+O(h_v^3)+O\left(\sum_s h_{\eta,s}^3\right).
\end{aligned}$$

In the expression above, E_{v_0,η_i} is the expectation conditional on v_0 and η_i. Hence we obtain:

$$E(\hat{m}(v_0,\eta_i)-m(v_0,\eta_i))=\left\{h_v^2B_v(v_0,\eta_i)+\sum_{s=1}^{q_2}h_{\eta,s}^2B_{\eta,s}(v_0,\eta_i)\right\}+o(1). \quad (5.B.6)$$

The single sum of (5.B.6) converges to its population mean by the Chebyshev's law of large numbers (see Linton & Härdle (1996)).

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

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

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

$$\begin{aligned}
& V\left(\frac{1}{n}\sum_{i=1}^{n}\check{m}(v_0,\eta_i)\right) = E\left(V_{v_0,\eta_i}\left\{\frac{1}{n}\sum_{i=1}^{n}\check{m}(v_0,\eta_i)\right\}\right) + V\left(E_{v_0,\eta_i}\left\{\frac{1}{n}\sum_{i=1}^{n}\check{m}(v_0,\eta_i)\right\}\right) \\
&= \sigma^2 f(\eta)^2 E\left[\frac{1}{nh_v^q}\sum_{j=1}^{n}K_v\left(\frac{V_{j,0}-v_0}{h_v}\right)\right]^2 + f(\eta)^2 V\left[\frac{1}{nh_v^q}\sum_{j=1}^{n}K_v\left(\frac{V_{0,j}-v_0}{h_v}\right)m(V_{0,j},\eta_j)\right] \\
&= \frac{\sigma^2 f(\eta)^2}{nh_v^q}\mathcal{K}_v + \frac{m(v_0,\eta)^2 f(\eta)^2 f(v_0)}{nh_v^q}\mathcal{K}_v + O(n^{-1}), \\
& V\left(\frac{1}{n}\sum_{i=1}^{n}\hat{f}(v_0,\eta_i)\right) = \frac{f(\eta)^2 f(v)\mathcal{K}_v}{nh_v^q} + O(n^{-1}) \\
& \text{Cov}\left(\frac{1}{n}\sum_{i=1}^{n}\check{m}(v_0,\eta_i), \frac{1}{n}\sum_{i=1}^{n}\hat{f}(v_0,\eta_i)\right) = E\left\{\frac{1}{n}\sum_{i=1}^{n}\check{m}(v_0,\eta_i)\frac{1}{n}\sum_{i=1}^{n}\hat{f}(v_0,\eta_i)\right\} \\
&- E\left\{\frac{1}{n}\sum_{i=1}^{n}\check{m}(v_0,\eta_i)\right\}E\left\{\frac{1}{n}\sum_{i=1}^{n}\hat{f}(v_0,\eta_i)\right\} \\
&= \frac{m(v_0,\eta)f(\eta)^2 f(v)\mathcal{K}_v}{nh_v^q} + O(n^{-1}),
\end{aligned}$$

where V_{v_0,η_i} denotes the variance conditional on v_0 and η_i. Hence we have:

$$\sqrt{nh_v^q}(\hat{m}(\hat{v}) - m(v_0) - bias) \to_D N(0, var).$$

The consistency of $\hat{g}(\hat{v})$ and its asymptotic normality are argued in the same way as above, since $m(v_0) = g(v_0) + c_1$. ■

Chapter 4

Semiparametric Analysis of Empirical Engel Curves in Australia

Consumer demand presents an important area for the application of semiparametric methods. In analysis of the cross-section behavior of consumers, nonparametric analysis of the Engel curve relationship is now common place.

Richard Blundell, Alan Duncan and Krishna Pendakur (1998)

4.1 Introduction

In this chapter, we intend to provide a further contribution to the economic literature, particularly on the cross sectional relationships between expenditure on specific goods and the level of total expenditure. To achieve this objective, we employ our newly established methods to conduct a semiparametric analysis of shape-invariant Engel curves in Australia. It should be noted that within the context of the empirical demand study, Blundell et al. (2007) address the endogeneity of the total expenditure by using the nonparametric IV method through which some regularity conditions are imposed on the inversion matrix and a constraint is placed on the space of the reduced relation to make it compact. Blundell et al.

(2007) show the $\sqrt{n}$-consistency of the estimators of both the scale and the shift coefficients. On the other hand, Blundell et al. (1998) address endogeneity by using the CF approach by a parametrically generated endogeneity control variable. We will clearly explain the difference between our method and that of Blundell et al. (1998) below. Furthermore, because of the importance of this topic, even though an effective tool is lacking for testing endogeneity in semiparametrics, an additional advantage of our method is that it enables a rather simple procedure to be established for the purpose. This is brought about mainly by its ability to identify and disentangle the effect of endogeneity in the model. This simple tool relies on the variability bands being constructed over the estimates of the endogeneity measures (to be defined below) as the means of testing their statistical significance.

In this section, we will study the relationships between expenditure on specific goods and the level of total expenditure by using our newly established method in Chapter 3 to conduct a semiparametric analysis of shape-invariant Engel curves in the Australian context. The data used is based on the Household, Income and Labor Dynamics in Australia (HILDA) Survey, which is Australia's household-based panel study that began in 2001. The goal of such a survey is to collect information about economic and subjective well-being, the labour market and family dynamics. The survey consists of more than 7,500 households with just below 20,000 individuals. The current release, i.e. Release 8, covers the first eight waves (out of 11) of data, which has recently become publicly available. The current section consists of four subsections. In Section 4.2, we explain the empirical model which our analysis will be based on. Section 4.2.1 discusses the details of the relevance of endogeneity in the study at hand. In Section 4.2.2, we then discuss the empirical estimation of the shape-invariant Engel curves and we present a number of important findings in Section 4.2.3.

4.2 The Empirical Model

Hereafter, let $\{Y_{1il}, X_{1i}, X_{2i}\}_{i=1}^{n}$ represent an i.i.d. sequence of n household observations on the budget share Y_{1il} of good $l = 1, \ldots, L \geq 1$ for each household i facing the same relative prices, the log of total expenditure X_{1i}, and a vector of household composition variables X_{2i}. For each commodity l, budget shares and total outlay are related by the general stochastic Engel curve $Y_{1il} = G_l(X_{1i}) + \epsilon_{il}$, where G_l is an unknown function that can be estimated using a standard nonparametric regression method under the exogeneity assumption of X_1, i.e. $E(\epsilon_l|x_1) = 0$. Furthermore, a number of previous studies have reported that household expenditures typically display a large variation with demographic composition. When X_2 is discrete, a simple approach for model estimation is to stratify by each distinct discrete outcome of X_2 and then estimate using nonparametric regression within each cell. At some point, however, it may be useful to pool the Engel curves across household demographic types and to allow X_1 to enter each Engel curve semiparametrically. This idea leads to the following specification:

$$Y_{1il} = g_l(X_{1i} - \phi(X_{2i}'\gamma_0)) + X_{2i}'\beta_{0l} + \epsilon_{il}, \tag{4.2.1}$$

where $g_l(\cdot)$ is an unknown function and $\phi(X_{2i}'\gamma_0)$ is a known function up to a finite set of unknown parameters γ_0 that can be interpreted as the log of general equivalence scales for household i.

The functional form specification in (4.2.1) deserves a few remarks. To this end, Blundell et al. (2003) show that such the functional form specification is consistent with consumer optimisation theory; see also the discussion of Lemma 3.2 of Blundell et al. (1998). Furthermore, in the current chapter, we choose $\phi(X_{2i}'\gamma_0) = X_{2i}'\gamma_0$, where X_{2i} is a vector of the demographic variables that represent different household types and γ_0 is the vector of the corresponding equivalence scales. Hence we have the following EGPLSI specification:

$$Y_{1il} = g_l(X_{1i} - X_{2i}'\gamma_0) + X_{2i}'\beta_{0l} + \epsilon_{il}. \tag{4.2.2}$$

In our application, we consider six broad categories of goods, namely food, clothing, alcohol, electricity and gas, transportation and other goods. In order to

Table 4.1: Descriptive statistics.

	Couples 1 child		**Couples 2 children**	
	Mean	**Std. Dev**	**Mean**	**Std. Dev**
Budget shares:				
Alcohol	0.03373	0.03608	0.02918	0.03409
Clothing	0.03060	0.02343	0.03212	0.02788
Electricity and gas	0.04077	0.16236	0.03850	0.14124
Food	0.31515	0.02600	0.31303	0.02872
Transportation	0.04076	0.00153	0.04385	0.00124
Other	0.56870	0.03060	0.57263	0.02308
Expenditure and income:				
log (total expenditure)	4.53302	0.20566	4.58983	0.17854
log (income)	4.92124	0.23414	4.96652	0.23769
Sample size	286		531	

preserve a degree of demographic homogeneity, we select a subset of married (or cohabiting) couples with one or two dependent children aged less than 16 years, in five Australian territory capital cities, namely Adelaide, Brisbane, Melbourne, Perth and Sydney. Therefore, our demographic variable, X_2, is simply a binary dummy variable that reflects whether the couple has one child ($X_2 = 0$) or two children ($X_2 = 1$). This leaves us with 817 observations, including 286 couples with one child.

The budget shares of these goods are presented in Table 4.1. The log of total expenditure on the these goods is our measure of the continuous endogenous explanatory variable X_1. Furthermore, Table 4.1 also presents descriptive statistics for the main variables used in this study. The table shows larger expenditure shares for alcohol, electricity and gas, and food for the couples with one child, but larger expenditure shares for clothing, transportation and other goods for the couples with two children. This indicates the differences in the consumption patterns between the two demographic groups, and we expect the estimators of the scale and shift coefficients to reflect these patterns.

4.2.1 A Simple Test of Endogeneity

Regarding the empirical study in the current section, in order to see the reason why the log of total expenditure X_1 is likely to be endogenous, i.e. $E(\epsilon_l|x_1) \neq 0$, let us note firstly that the system of budget shares can be thought of as the second stage in a two-stage budgeting model (see Gorman (1959) for details), in which total expenditure and savings are first determined conditional on total expenditure, and individual commodity shares are chosen at the second stage; see Blundell (1988) for example. Hence X_1 is a variable which reflects savings and other consumption decisions made at the same time as the budget shares Y_1 are chosen. In our analysis that follows, we consider an earning variable, which is the amount that a household earned before tax in the chosen year, as an instrument.

Figures 4.1 and 4.2 present a plot of the kernel estimates for the joint density of log(total expenditure) and log(earning) and a plot for $E(\log(\text{expenditure})|\log(\text{earning}))$, respectively. The two variables show strong positive correlation such that for the sample with one child, the correlation is 0.4882 and is 0.4056 for those with two children. As seen in the figure, the joint density is also smooth and, together with the conditional mean, confirms our belief that the gross earnings variable should be a good choice for our instrumental variable. Since the kernel estimate of the density of log earnings is close to normal, we have taken the instrumental variable $Z = \Phi(\log \text{earnings})$ in the empirical applications and write:

$$\eta_i = X_{1i} - m_{X1}(Z_i). \tag{4.2.3}$$

Our model, which consists of the index model in (4.2.2) and the specification of the endogeneity control regressor in (4.2.3), is appropriate for the application since it is coherent with the economic theory and it allows for the endogeneity of total expenditure as discussed earlier.

Figure 4.1: Kernel joint density estimates with a full bandwidth matrix.

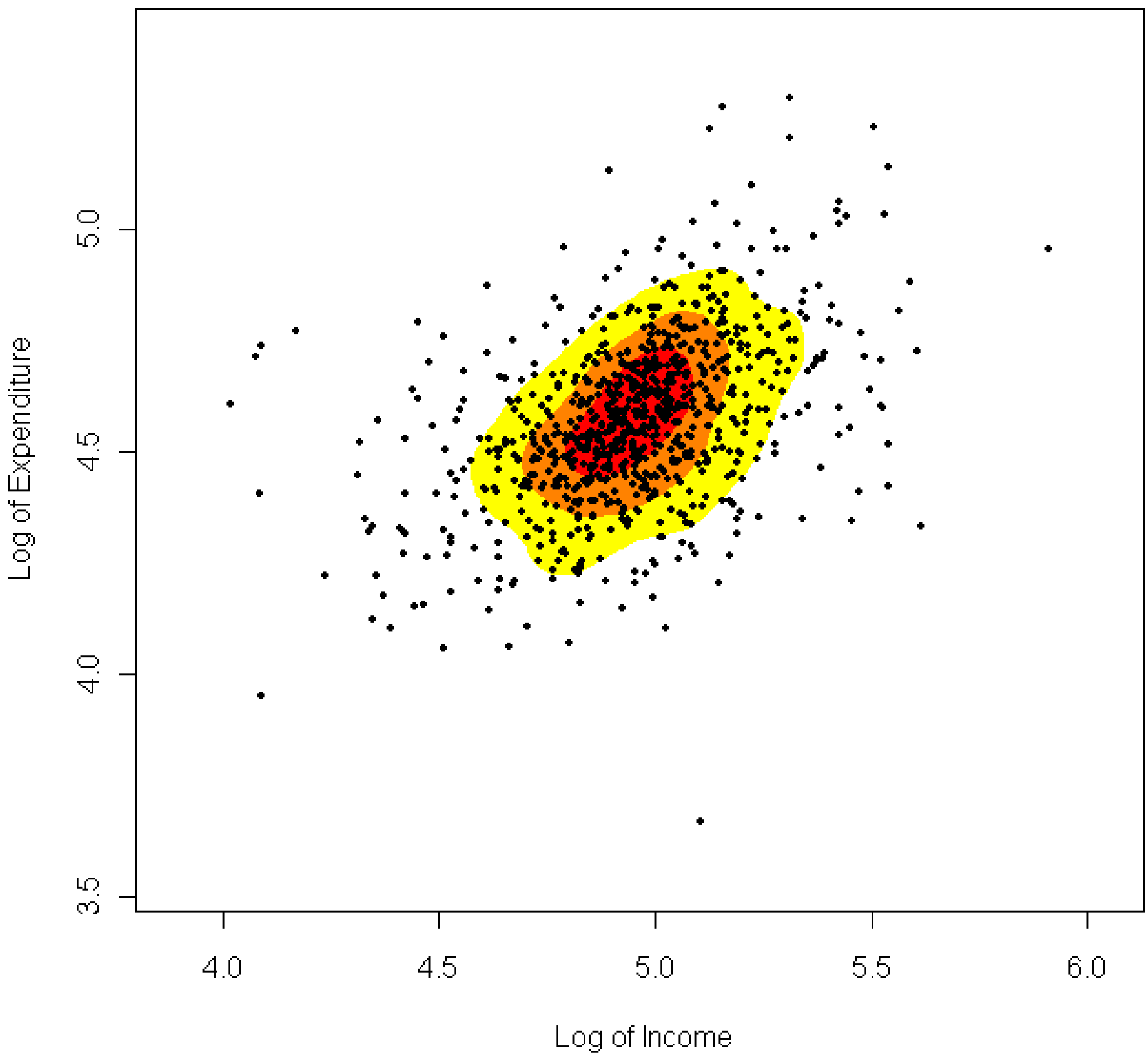


Figure 4.2: Kernel estimates of conditional expectation of log(expenditure) with respect to log(income).

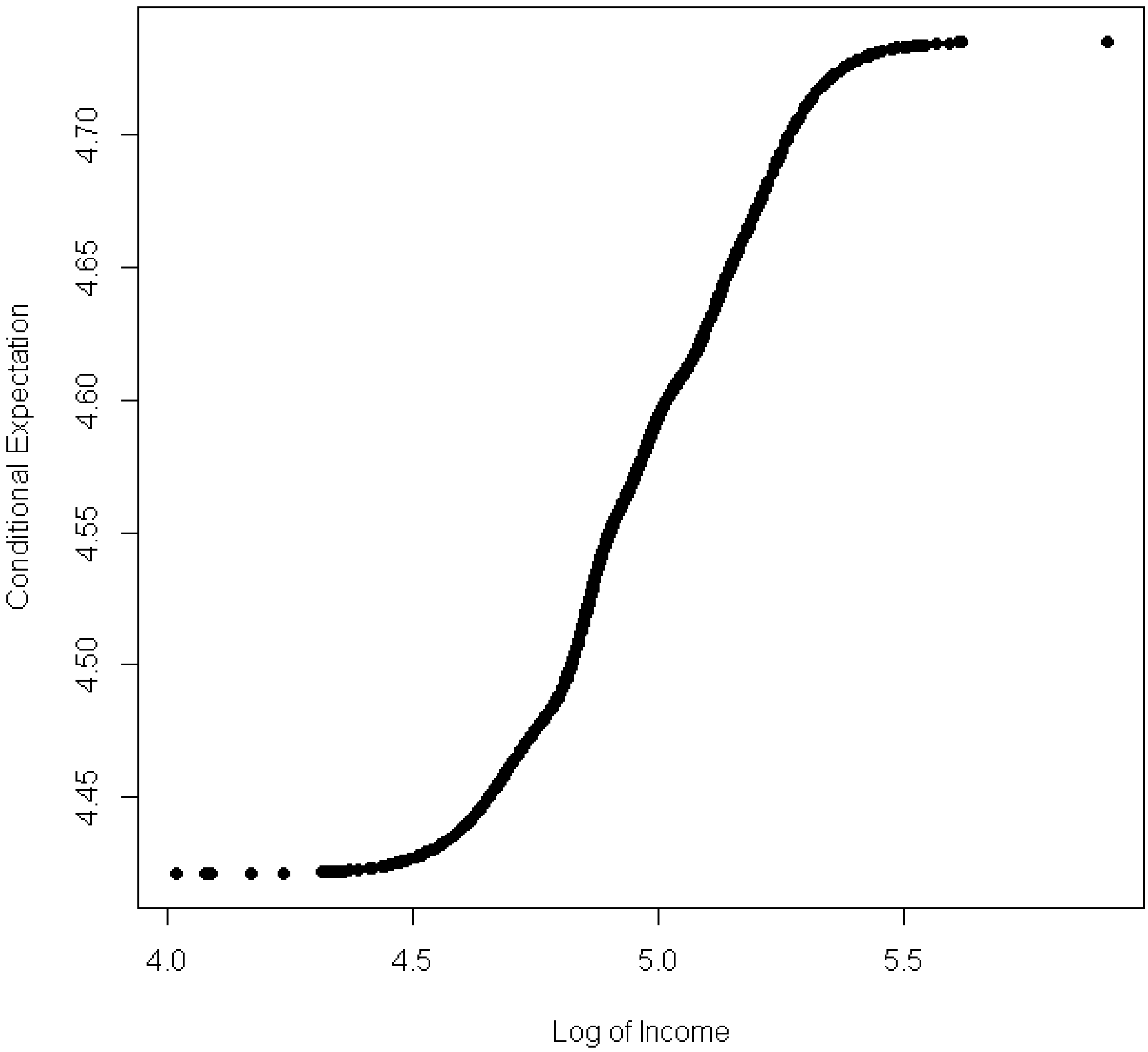


Figure 4.3 shows log(expenditure) (black line), m_{X1} (red line) and η (blue line). In the view of this triangular structure, the figure stresses that the endogenous variable, X_1, may be decomposed into the exogenous (i.e. Z) and the endogenous (i.e. η) components. An important observation to be noted is that even for cases in which we are able to identify a strong instrument (with strong explanatory power), the impact of endogeneity is still determined by the relationship between ϵ_l and η, i.e. the conditional expectation of the former with respect to the latter. We will explore this point further below.

In the following, we discuss the construction of variability bands in our analysis and how they can be used as a preliminary test of exogeneity. For convenience, let us first restate the triangular structure as:

$$\begin{aligned} Y_{1il} &= g_l(X_{1i} - X_{2i}'\gamma_0) + X_{2i}'\beta_{0l} + \epsilon_{il}, && (4.2.4) \\ X_{1i} &= m_{X1}(Z_i) + \eta_i, && (4.2.5) \end{aligned}$$

where $m_{X1}(z) = E(x_1|z)$, under the assumptions of the following:

$$E(\eta|z) = 0 \ \text{ and } E(\epsilon_l|z,\eta) = E(\epsilon_l|\eta) \neq 0. \quad (4.2.6)$$

The structure described in (4.2.4) to (4.2.6) suggests that we have:

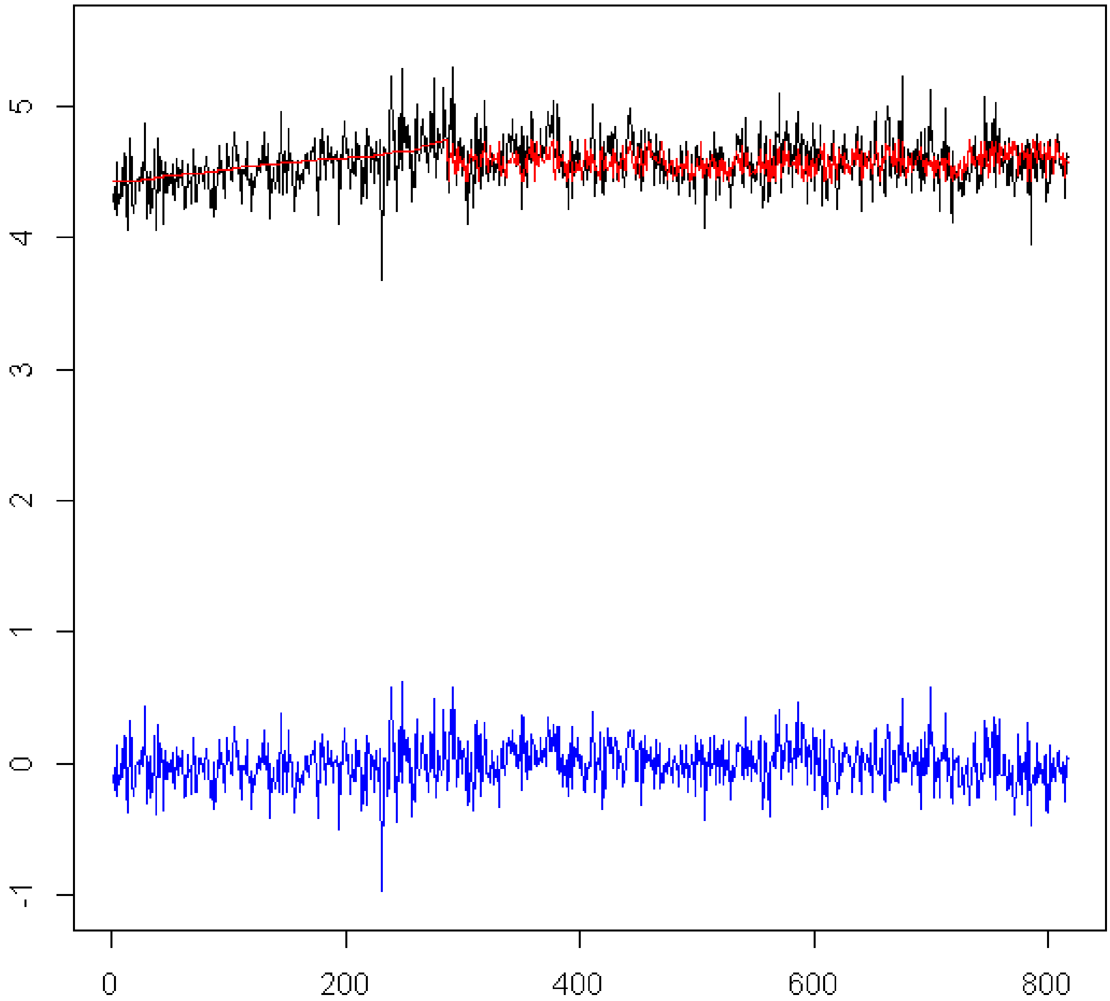
$$E[y_{1l}|(x_1 - x_2'\gamma_0),\eta] - E[x_2|(x_1 - x_2'\gamma_0),\eta]'\beta_{0l} = g_l(x_1 - x_2'\gamma_0) + \iota_l(\eta), \quad (4.2.7)$$

where $E[\epsilon_l|(x_1 - x_2'\gamma),\eta] = E[\epsilon_l|x_2,\eta] = E[\epsilon_l|\eta] \equiv \iota_l(\eta) \neq 0$. Expression (4.2.7) implies, however, that we have:

$$\begin{aligned} Y_{1il} &= X_{2i}'\beta_{0l} + g_l(X_{1i} - X_{2i}'\gamma_0) + \iota_l(\eta_i) + e_{il}, && (4.2.8) \\ X_{1i} &= m_{X1}(Z_i) + \eta_i, && (4.2.9) \end{aligned}$$

where $E(e_l|\eta) = 0$. Let $M_l[(X_{1i} - X_{2i}'\gamma_0),\eta_i] = g_l(X_{1i} - X_{2i}'\gamma_0) + \iota_l(\eta_i)$. In order to use (4.2.8), it is important to note that:

Figure 4.3: log(expenditure), m_{X1} and η.



$$\begin{aligned} m_l(x_1 - x_2'\gamma_0) &= \int M_l(\{x_1 - x_2'\gamma_0\}, \eta)\, d\eta = g_l(x_1 - x_2'\gamma_0) + c_1 \\ g_l(x_1 - x_2'\gamma_0) &= m_l(x_1 - x_2'\gamma_0) - c_1, \end{aligned} \tag{4.2.10}$$

where $c_1 = \int \iota(\eta) dQ(\eta)$ and $E(g_l(\cdot)) = 0$; the estimation of which can be done based on the marginal integration technique in Step 3.2.2.5 of Procedure 3.2.2.

Now, observe that if we were to impose a linear specification on $\iota_l(\cdot)$, (4.2.8) would be closely similar to the extended partially linear (EPL) model discussed in Blundell et al. (1998). In this case, Blundell et al. (1998) showed that a test of the exogeneity null can be constructed by testing $H_0 : \iota_l = 0$, where ι_l is an unknown parameter. To allow for more flexibility on the functional form between the total expenditure and its instrument, as an alternative, one may apply an existing test of a parametric mean regression model against a nonparametric alternative; see Horowitz & Spokoiny (2001), for example. However, in the current chapter, we suggest that it is more convenient to simply construct the variability bands for $\iota_l(\cdot)$ since its estimate is readily available. To do so, we use the following procedure.

Procedure 4.2.1

Step 4.2.1.1: Obtain an empirical estimate of $g_l(x_1 - x_2'\gamma_0)$ in (4.2.10); see also Remark 4.1.

Step 4.2.1.2: Regress (4.2.9) using the estimates in Step 4.2.1.1 to obtain the nonparametric estimates of $\iota_l(\cdot)$.

Step 4.2.1.3: Compute the bias-corrected confidence bands for the nonparametric regression using the procedure introduced in Xia (1998). Finally, the above mentioned (Bonferroni-type) variability bands are obtained using a similar procedure discussed in Eubank & Speckman (1993).

Remark 4.1. *To complete Step 4.2.1.1, Procedure 3.2.2 in Section 3.2.2 can be useful. However, some modifications are required to take the index coefficient* γ_0 *into account, which can be interpreted as a general equivalence scale for household i. Steps 3.2.2.1 and 3.2.2.2 are directly applicable since they are implemented using a given* γ *across* $l = 1, 2, \ldots, 6$ *commodities. In this case, the objective* *function (3.2.17) in Step 3.2.2.3 is only used for the particular l commodity. A new objective function is the summation of these individual functions, i.e.* $\min_{\gamma \in A_n, h_{v,l}, h_{\hat{\eta},l} \in \mathcal{H}_n} \hat{J}(\gamma, h_{v,l}, h_{\hat{\eta},l})$*, which is minimised with respect to γ and 12 bandwidth parameters, i.e. two for each commodity. Finally, Steps 3.2.2.4 and 3.2.2.5 are directly applicable using $\hat{\gamma}$ as well as $\hat{h}_{v,l}$ and $\hat{h}_{\hat{\eta},l}$.* ■

4.2.2 Shape-Invariant Engel Curves

First, observe that (4.2.8) can also be re-stated as:

$$\tilde{Y}_{1il} = g_l(X_{1i} - X_{2i}'\gamma_0) + e_{il}, \tag{4.2.11}$$

where $\tilde{Y}_{1il} = Y_{1il} - X_{2i}'\beta_{0l} - \iota_l(\eta_i)$. The use of (4.2.11) relies on the following corresponding expression of (4.2.10):

$$\begin{aligned} m_l(\eta) &= \int M_l(v, \eta)\, dv = \iota_l(\eta) + c_2 \\ \iota_l(\eta) &= m_l(\eta) - c_2, \end{aligned} \tag{4.2.12}$$

where $v = x_1 - x_2'\gamma$, $c_2 = \int g(v)dQ(v)$ and $E(\iota_l(\cdot)) = 0$. Hence (4.2.11) suggests that we are able to employ *Procedure 4.2.2* below in order to obtain the estimates of the shape-invariant Engel curves and the related confidence bands.

Procedure 4.2.2

Step 4.2.2.1: Obtain empirical estimates of $\iota_l(\eta)$ in (4.2.12).

Step 4.2.2.2: Regress (4.2.11) using the estimates in Step 4.2.2.1 to obtain the nonparametric estimates of $g_l(\cdot)$.

Step 4.2.2.3: Compute the bias-corrected confidence bands about the nonparametric estimator in Step 4.2.2.2 using the procedure introduced in Xia (1998).

4.2.3 Empirical Findings

Prior to presenting our empirical findings, let us recapitulate our empirical model of shape-invariant Engel curves and made a final remark on the estimation of the model through the EGPLSI structure. Recall from the above discussion that the empirical model we are attempting to estimate is of the following:

$$\begin{aligned} Y_{1il} &= g_l(X_{1i} - \gamma_0 X_{2i}) + \beta_{0l} X_{2i} + \epsilon_{il}, \\ X_{1i} &= m_{X1}(Z_i) + \eta_i. \end{aligned} \tag{4.2.13}$$

The EGPLSI structure which provides a direct lead to (4.2.13) is:

$$Y_{1il} = g_l(\alpha_{01} X_{1i} + \alpha_{02} X_{2i}) + \beta_{01,l} X_{1i} + \beta_{0l} X_{2i} + \epsilon_{il}, \tag{4.2.14}$$

which satisfies all the conditions given beneath (3.2.1), such that $V_{0i} = \alpha_{01} X_{1i} + \alpha_{02} X_{2i}$ and $\overline{V}_{0i} = X_{1i} + (\alpha_{02}/\alpha_{01}) X_{2i}$. Hence, the estimation of the model can be based either on the standardised (4.2.13) or the EGPLS version in (4.2.14) by which the estimates of α_{01} and α_{02} can also be obtianed. The estimatability of the model can be further assured by imposing an assumption which is based closely on Assumption of Ai & Chen (2003) as follows.

Assumption 3.2.6. *Assume that* $E\left[(y_{1l} - M_l(\alpha_{01}x_1 + \alpha_{02}x_2) - \beta_{01,l}x_1 - \beta_{0l}x_2)\,|x_2, z\right]$ $= E\left[(y_{1l} - M_l(\alpha_{01}x_1 + \alpha_{02}x_2) - \beta_{0l}x_2)\,|x_2, z\right] = 0$ *so that*

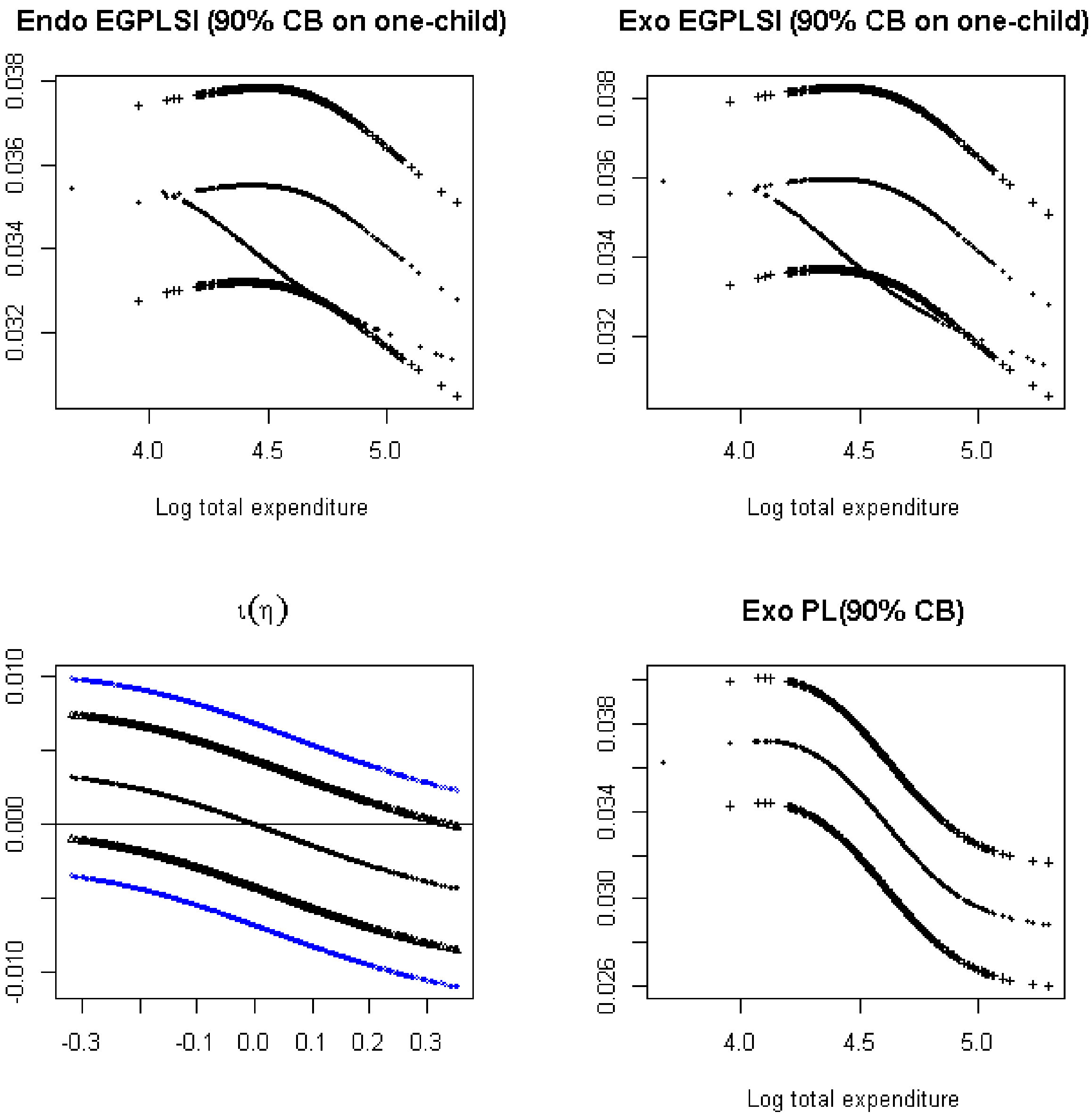
$$E[y_{1l}|x_2, z] = E\left[(M_l(\alpha_{01}x_1 + \alpha_{02}x_2) + \beta_{0l}x_2)\,|x_2, z\right],$$

and, therefore, that $E[\beta_{01,l}x_1|x_2, z] = 0$. ■

Hereafter, let us use $\hat{g}_{1,l}(\cdot)$ and $\hat{\iota}_{1,l}(\cdot)$ to denote the empirical estimates of $g_l(\cdot)$ and $\iota_l(\cdot)$ based on the marginal integration techniques, i.e. those obtained from Steps 4.2.1.1 and 4.2.2.1, respectively. Furthermore, let us use $\hat{g}_{2,l}(\cdot)$ and $\hat{\iota}_{2,l}(\cdot)$ to denote the empirical estimates of $g_l(\cdot)$ and $\iota_l(\cdot)$ which are obtained from Steps 4.2.1.2 and 4.2.2.2, respectively. Table 4.2 below presents the empirical estimates of the unknown parameters γ_0 and β_{0l} (4.2.4). In addition, to demonstrate the validity of our Procedures 4.2.1 and 4.2.2 above, in the table we also present in the following average squared difference:

$$d_{gl} = \frac{1}{n}\sum_{i=1}^{n}\left\{\hat{g}_{1,l}(\hat{v}) - \hat{g}_{2,l}(\hat{v})\right\}^2 \quad \text{and} \quad d_{\iota l} = \frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\iota}_{1,l}(\hat{\eta}) - \hat{\iota}_{2,l}(\hat{\eta})\right\}^2,$$

Figure 4.4: Engel curves for alcohol



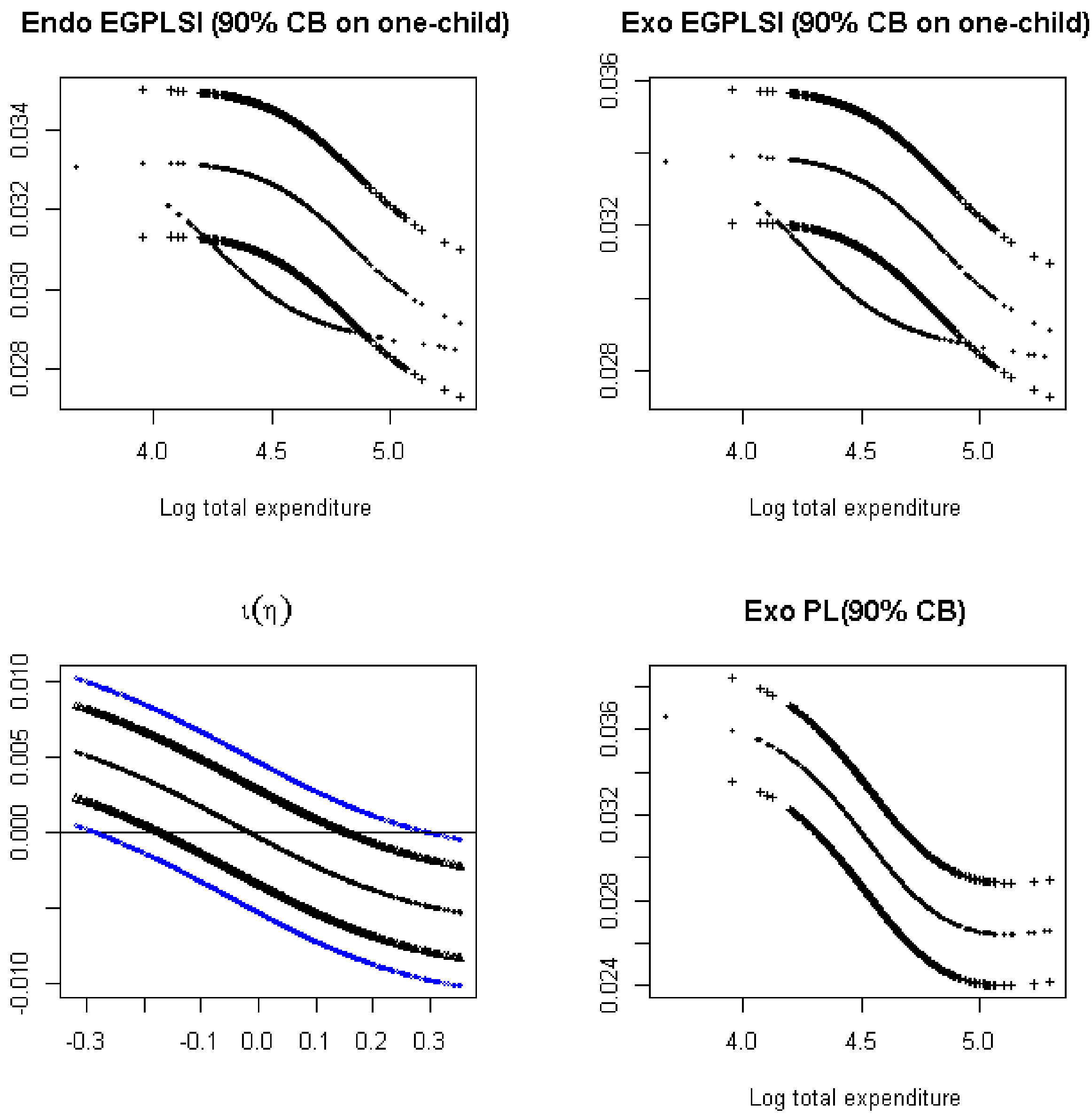
where $\hat{v} = x_1 - \hat{\gamma} x_2$.

We will now summarise a number of important findings based on the empirical results in Table 4.2 and Figures 4.4 to 4.9.

Firstly, the average squared errors reported in the fourth and the fifth columns of Table 4.2 are virtually zero, which provides strong evidence in support of the procedures discussed in Sections 4.2.1 and 4.2.2. Secondly, the signs and magnitudes of the estimates of the parameters reported in the first and the third columns are consistent with what is reported in the existing literature; see Blundell et al. (1998) for example. Furthermore, Figures 4.4 to 4.9, present the Engel curves for the six budget shares in our HILDA sample, each of which consists of four panels. The first and second panels present estimates of the Engel curves (for couples with one child and couples with two children) based on the EGPLSI model with the endogeneity being controlled using Procedure 3.2.2 and the endogeneity not being controlled by Procedure 3.2.1 in Chapter 3, respectively. Xia's (1998) confidence bands are constructed for the Engel curves of couples with one child. Furthermore, the fourth panels present estimates of the Engel curves computed using the partially linear model of Robinson (1988) in Chapter 2 for the sake of comparison with the EGPLSI model. They show clear evidence that the partially linear model restricts the empirical Engel curves to be within the same specification; see Blundell et al. (1998) and Blundell et al. (2003) for example, where all empirical Engel curves are similar to the quadratic functional form.

Finally, the third panel of each graph presents the nonparametric estimates of $\iota_l(\cdot)$ with two sets of bands, namely the bias-corrected confidence bands for the nonparametric regression of Xia (1998) (black) and the Bonferroni-type variability bands discussed in Eubank & Speckman (1993) (blue). Regarding alcohol, clothing and transportation, $\iota_l(\cdot)$ for these cases do not seem to be statistically significant. These findings can be linked to the fact that the shapes of the Engel curves presented in the top two panels are similar. In other words, we show that the seriousness of the effect of the endogeneity problem, given an instrument, depends very much on the relationship between the disturbances in the structural

Figure 4.5: Engel curves for clothing



and the reduced relations, i.e. the relationship between ϵ and η, which, in this case, is summarised by $\iota_l(\eta)$. For a given instrument and therefore the corresponding η, $\iota_l(\cdot)$ can be a function such that the impact of endogeneity is minimal, e.g. in the case of alcohol, clothing and transportation. Otherwise, they may be functions which make the effect of the endogeneity severe, such as the case of electricity and gas.

Some of these Engel curves, e.g. those of alcohol, clothing and transportation, appear to demonstrate that the Working-Leser linear logarithmic (Piglog) formulation is a reasonable approximation. Nonetheless, for other shares, particularly electricity and gas, and food and other goods, a more nonlinear relationship between the shares and the log expenditure is evident. Regarding alcohol, clothing and transportation, although the Engel curves for our two demographic groups both slope downward a broadly parallel shift in the Engel curves does not seem to appear. In fact, the Engel curves of families with two children tend to decline at a much faster rate as the log total expenditure increases.

On the contrary, it is interesting to note how similar the shapes of the Engel curves are for our two demographic groups for food and other goods. In these cases, there appears to be a parallel shift in the Engel curves. A couple with one child spends around 15% more of their budget on food than a couple with two children. However, couples with two children end up spending 4% more of their budget on other goods than couples with one child at the same level of expenditure. Such outcomes seem consistent with our intuitive belief about consumption behaviour in practice, i.e. a couple with two children incurs additional costs for having an extra child which are hidden within the other goods category.

Figure 4.6: Engel curves for electricity and gas

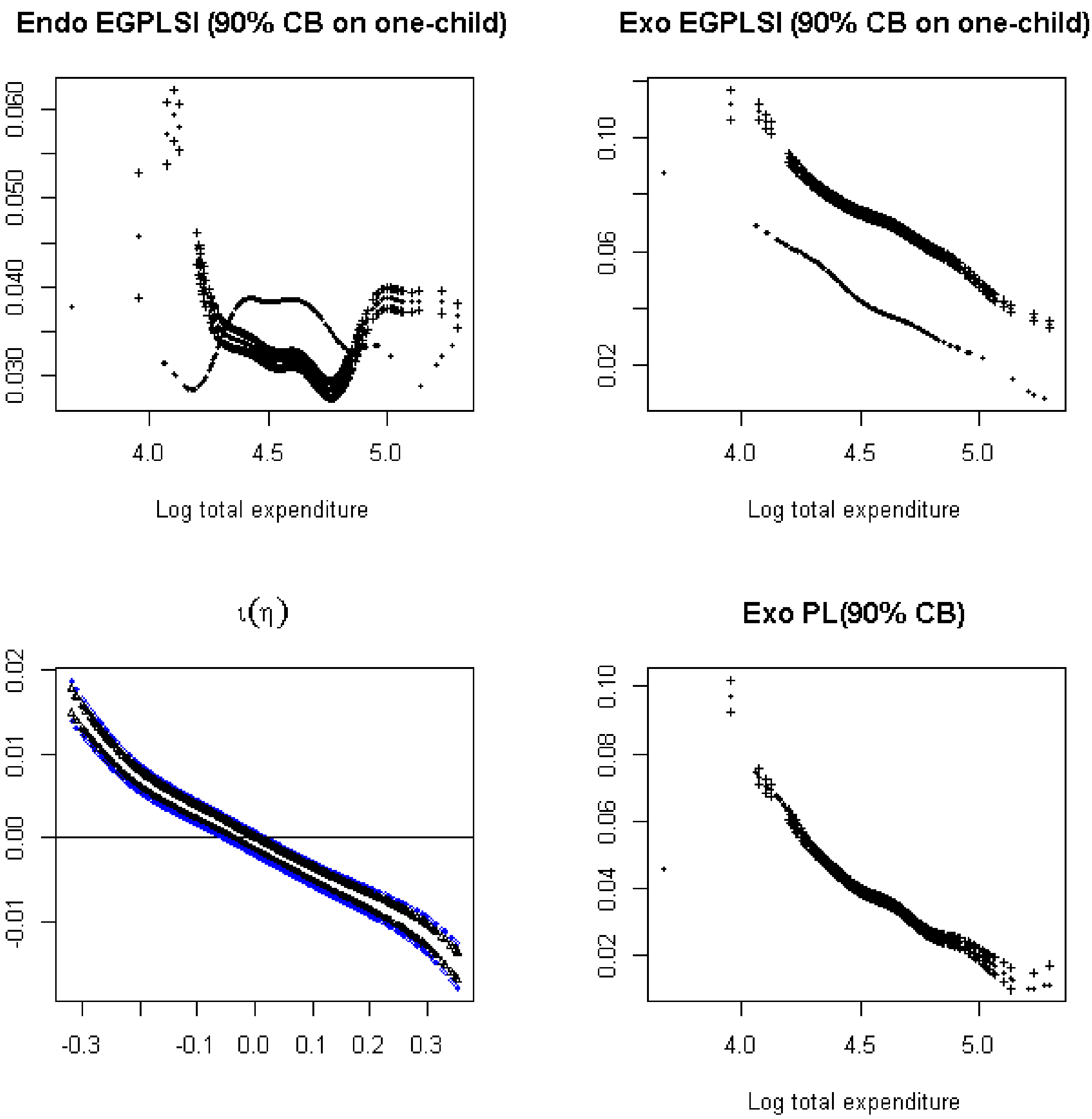


Figure 4.7: Engel curves for transportation

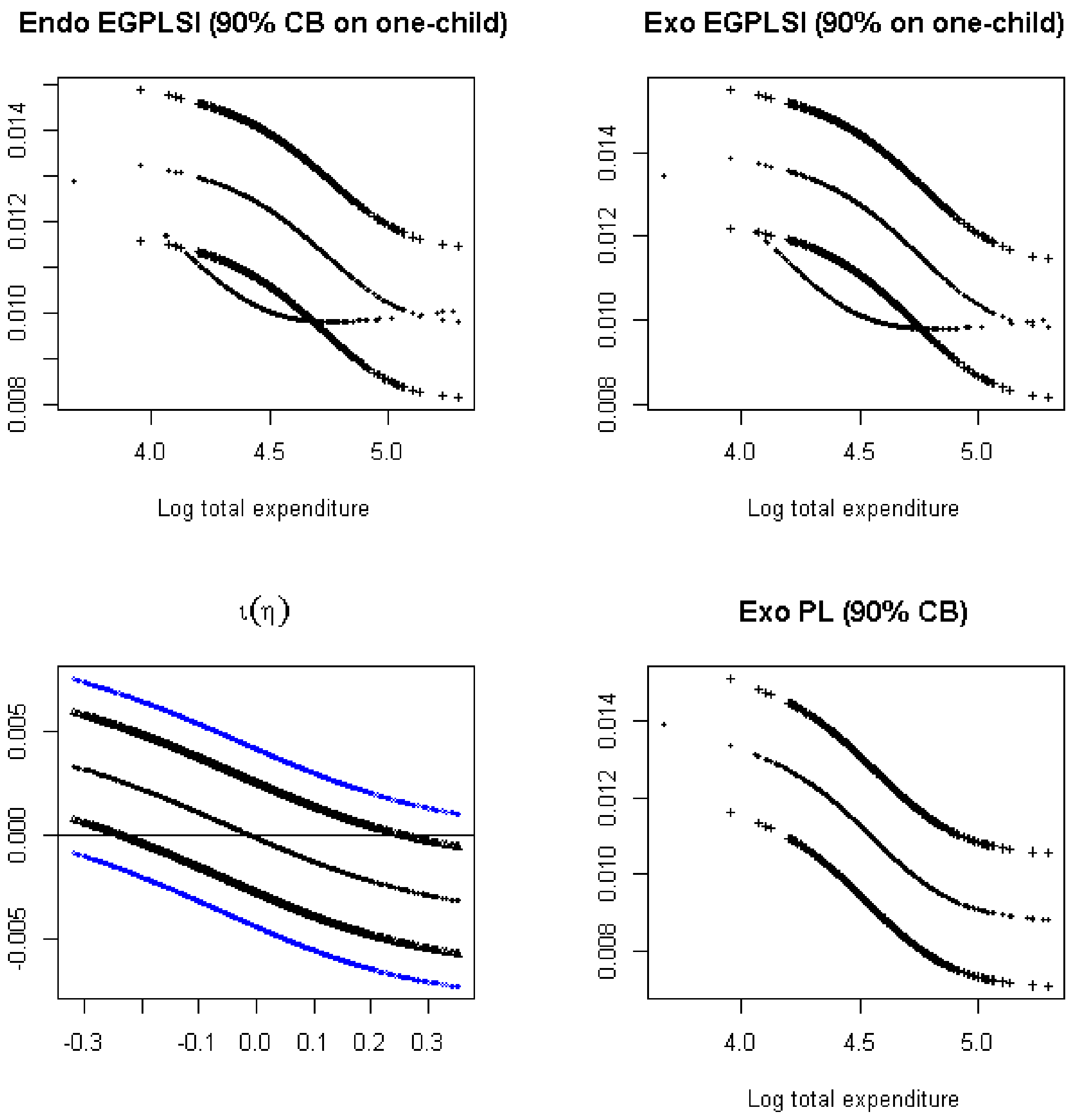


Figure 4.8: Engel curves for food

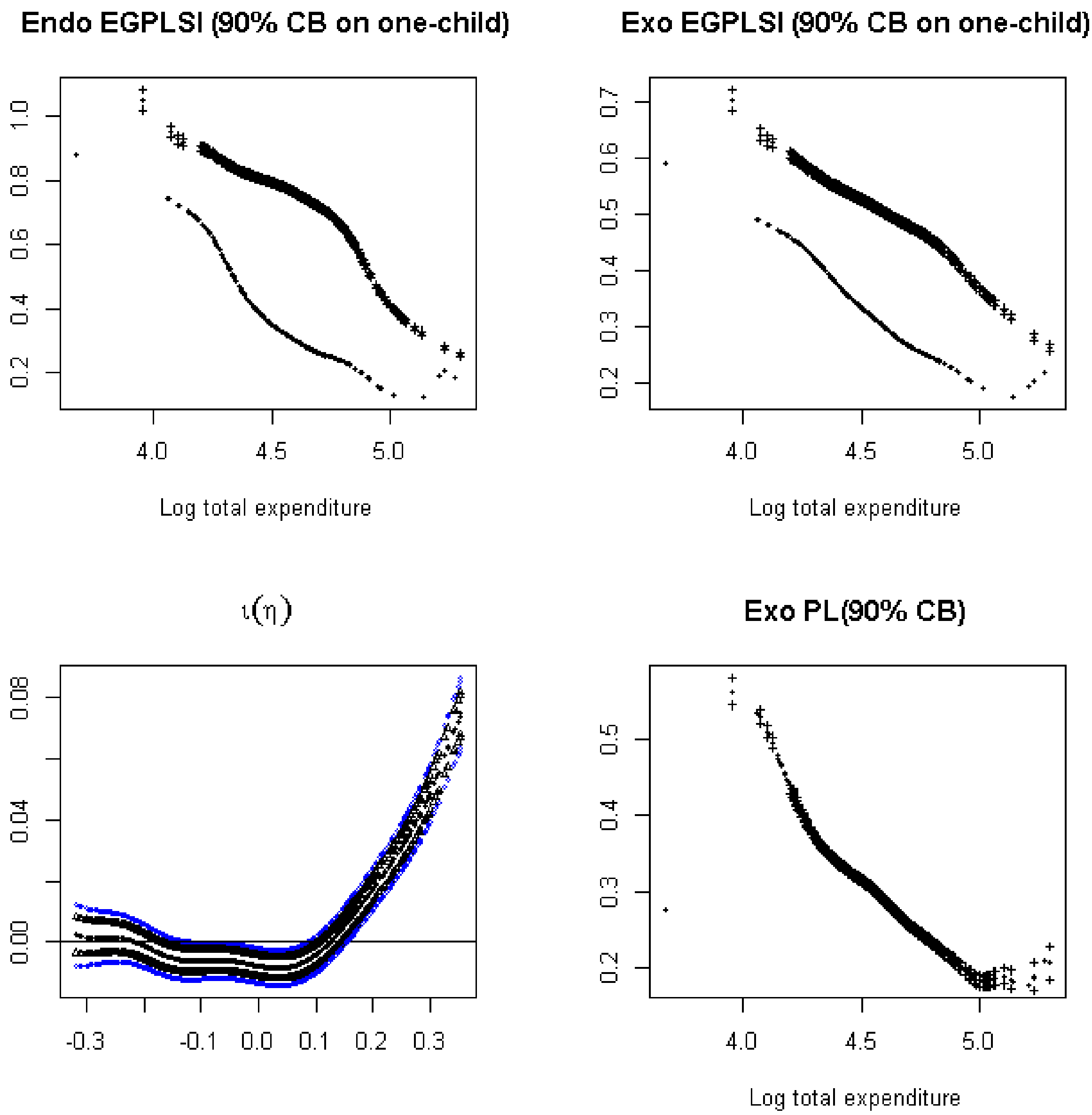


Figure 4.9: Engel curves for other goods

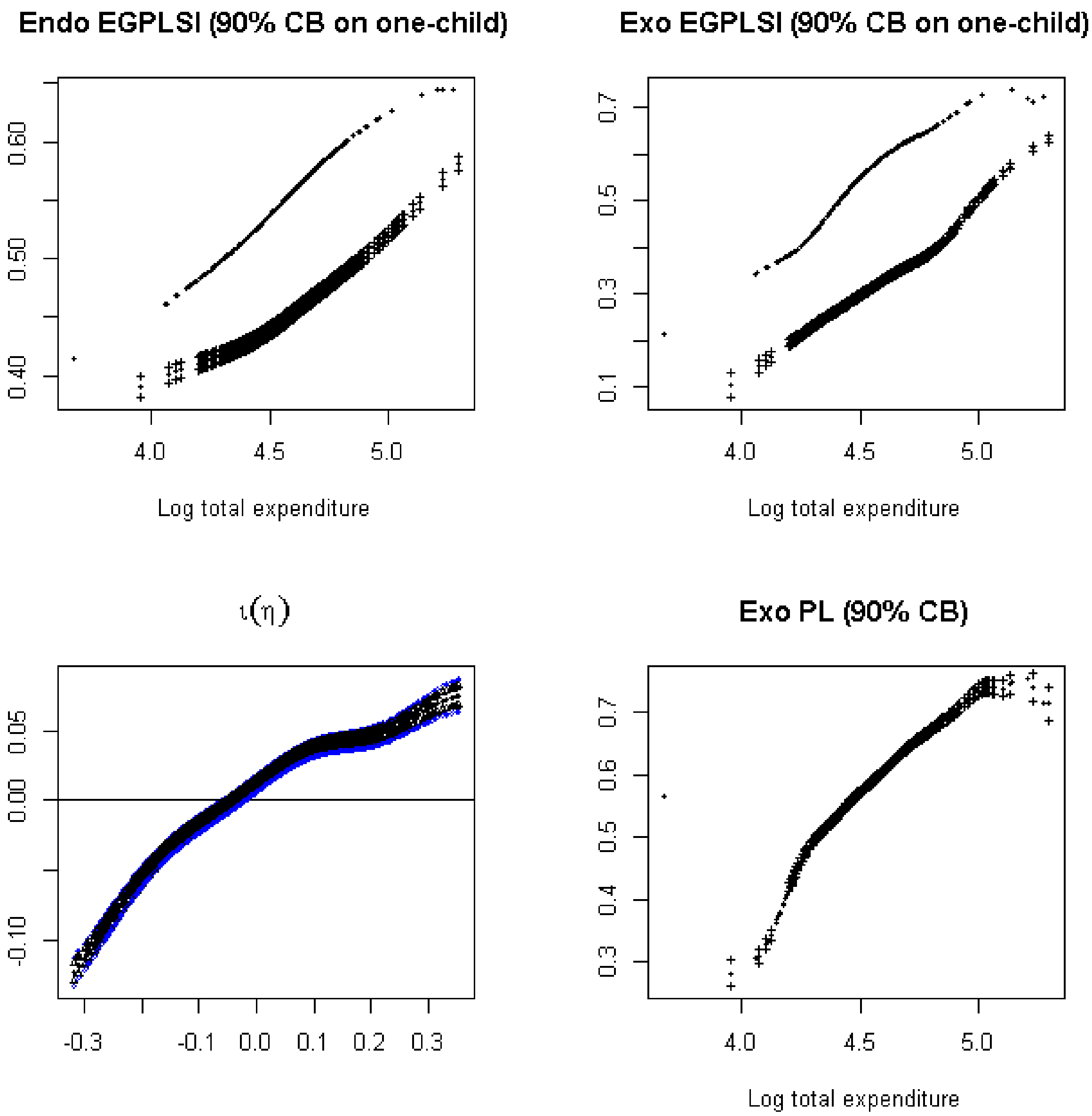


Table 4.2: Empirical results

$\hat{\gamma}$	Categories of goods	$\hat{\beta}_l$	d_{gl}	$d_{\iota l}$	$\hat{h}_{v,l}$	$\hat{h}_{\hat{\eta},l}$
0.5813	Alcohol	-0.0053	3.9781e-07	3.2355e-06	0.581334	0.581333
	Clothing	0.0005	7.8607e-07	6.4676e-06	0.581332	0.581330
	Food	-0.4541	3.4367e-04	1.7932e-04	0.065466	0.065465
	Electricity and Gas	0.0133	6.9226e-06	2.8772e-06	0.065465	0.065466
	Transportation	-0.0024	5.3794e-07	2.3716e-06	0.581335	0.581333
	Other	0.1245	1.6083e-04	2.8754e-04	0.065466	0.065465

4.3 Conclusions

In this chapter, we employ the estimation procedures and methods in the previous chapters to address the endogeneity of the total expenditure for a semiparametric analysis of empirical Engel curves. We particularly consider the "biased-adjusted" confidence band for the nonparametric structural function since the index coefficient is estimated and the endogeneity control regressor is generated when the EGPLSI model is considered. This corrected confidence band gives us useful information such as whether the effect of endogeneity is significant by analysing whether the band is significantly different from zero.

The application illustrates that the partially linear model restricts empirical Engel curves to be within the same specification, where all empirical Engel curves are similar to the quadratic functional form. However, the EGPLSI model, which is coherent with consumption theory, shows different functional forms for different commodities. Also, the EGPLSI model shows that the effect of endogeneity on total expenditure is nontrivial, the magnitude of the effects can be measured by the endogeneity control functions and they are statistically significantly different from zero.

Chapter 5

Conclusion

It is good to have an end to journey toward; but it is the journey that matters, in the end.

Ernest Hemingway

5.1 Summary

A PL semiparametric model allows us to conduct an empirical study with the benefits of both parametric and nonparametric modelling. However, careful treatment is required when addressing endogeneity in the model due to its complex feature. For instance, identification of the source(s) of endogeneity and an appropriate estimation procedure and methods accordingly are nontrivial issues as we discussed in Chapter 2. It is essential to construct the consistent estimators of the parametric coefficients given the dominance of the parametric part in the model. The endogeneity in parametric regressors can conveniently addressed without much difficulties using the conventional parametric alternatives such as PIV and P2SLS estimations. Thanks to the two-stage estimation procedure of Robinson (1988) and Speckman (1988) which partials out the nonparametric component from the structural one to obtain the linear reduced form then to estimate parametric coefficients, the estimators of the parametric coefficients is still consistent with the presence of only nonparametric endogeneity. However, it is a challenging task to address endogeneity particularly in the SI type of semiparametric models since the index coefficients also should be considered. The SI models does not identify the unknown structural function and also the estimators of the index coefficients are not consistent with the presence of endogeneity. Two most popular alternatives to address endogeneity are the NpIV estimation and the CF approach based on the nonparametric triangular structure. In this thesis, we employ the CF approach to address endogeneity in the PL and the EGPLSI models.

The imposition of the well-known nonparametric triangular structure of Newey et al. (1999) allows for imposing the exclusion restriction which leads to control endogeneity by introducing the endogeneity control function. As the result, we use the marginal integration technique to recover the unknown structural function since we have a simple additive nonparametric model. More importantly, this allows us to identify and disentangle the effects of endogeneity in the model. However, the generated regressor issue should be addressed since the endogeneity control variable is not observable in practice but instead estimated from the reduced form. Given this generated regressor issue, we show the asymptotic properties of the estimator of $g(\cdot)$ in both the PL and the EGPLSI models. We also show that the estimators of the parametric coefficients are still $\sqrt{n}$-consistent and asymptotically normal in both the PL and the EGPLSI models. Furthermore, we also show that estimators of index coefficients are $\sqrt{n}$-consistent and asymptotically normal, and the attractive feature of Xia et al. (1999) (the same bandwidth(s) are used for estimating the index coefficients and the unknown structural function) is still applicable in the EGPLSI model with the presence of endogeneity.

In recent years, the semiparametric technique becomes the important tool to analyse empirical Engel curves since it provides the flexibility of depicting any type of nonlinear relationships between budget shares and total expenditure and allows inclusion of the effects of demographic variables on demand. More importantly, the EGPLSI is able to provide an accurate demand analysis which is coherent with the consumer optimisation theory. Hence we conduct the analysis of empirical Engel curves with the methodologies proposed in Chapters 2 and 3 in order to take the endogeneity of total expenditure into account. In particular, the EGPLSI specification does not restrict overall shapes of Engel curves unlike the PL counterpart. More importantly, we show that the effects of endogeneity is not trivial. The magnitude of the effects of endogeneity is measured by an endogeneity control function by providing biased adjusted confidence band. The biased adjusted confidence band for the nonparametric function is needed since the scale (index) coefficients are estimated and the endogeneity control variable is generated. As the result of the correct confidence band, we are able to obtain the useful information such as whether the effect of endogeneity is significant by analysing that the band is statistically significantly different from zero.

5.2 Future Research

The EGPLSI model of (3.2.1) in Chapter 3 encompasses most of the popular econometric models as special cases such as parametric, nonparametric and semi-parametric models. The EGPLSI model with the presence of endogeneity in the index component is:

$$Y_i = X_i'\beta_0 + g(X_i'\alpha_0) + \epsilon_i, \tag{5.1}$$

where $\beta_0 \perp \alpha_0$ with $||\alpha_0|| = 1$, $E(\epsilon|x_1) = 0$ with $X_1 \subseteq \mathbb{R}^{q_1}$ and $q_1 < q$ is the parametric regressors belong to a subset of X and $E(\epsilon|x) \neq 0$. We can conveniently address endogeneity with the the nonlinear two-stage least squares estimation (NL2SLS) method of Amemiya (1974) in the case where $g(\cdot)$ is a known link function (a nonlinear model). The estimation method possesses relatively simple estimation procedures and an easy implementation, in practice. Hence we intend to generalise the NL2SLS estimation method of Amemiya (1974) to the case where $g(\cdot)$ is a unknown link function in order to address endogeneity in (5.1), in the future study. Note that the parametric endogeneity, i.e. $E(\epsilon|x_1) \neq 0$, is conveniently addressed using the conventional parametric treatments; see Chapter 2 for details. In this section, we firstly give a brief review on the NL2SLS estimation method of Amemiya (1974) then generalise the method to the case where $g(\cdot)$ is unknown. We also outline the main objectives and issues of the study.

To simplify the argument, we consider the simple nonlinear model without he parametric component in (5.1) as shown below:

$$Y_i = g(X_i'\alpha_0) + \epsilon_i, \tag{5.2}$$

where $g(\cdot)$ is a known link function and $E(\epsilon|x) \neq 0$. We assume that Z_i is a $\mathbb{R}^{q_2}$ IV vector for X_i, where $q_2 > q - q_1$ such that $E(\epsilon|z) = 0$ and $E(xz) \neq 0$. Then the NL2SLS estimator is the value of α that minimises the below objective function:

$$\begin{aligned}
J(\alpha) &= E\left[\{g(V_{0,i}) - g(V_i) + \epsilon_i\} P_Z \{g(V_{0,i}) - g(V_i) + \epsilon_i\}\right] \\
&= E\left[\{g(V_{0,i}) - g(V_i)\} P_Z \{g(V_{0,i}) - g(V_i)\}\right] + E\left(\epsilon_i P_Z \epsilon_i\right) \\
&+ 2E\left[\{g(V_{0,i}) - g(V_i)\} P_Z \epsilon_i\right] \\
&= A_{2,1} + A_{2,2} + A_{2,3},
\end{aligned} \tag{5.3}$$

where $V_{0,i} = X_i'\alpha_0$, $V_i = X_i'\alpha_0$ and $P_Z = Z(Z'Z)^{-1}Z'$. The first term in the right-hand side of (5.3), i.e. $A_{2,1}$, provides the value of α that minimises $J(\alpha)$ since $A_{2,3}$ converges to 0 in probability and $A_{2,2}$ is not relevant to α; see discussions in Chapter 3 and chapter 8 of Amemiya (1985) for details.

In the future study, we intend to generalise the above NL2SLS estimation to the EGPLSI case. We firstly present the generalised NL2SLS estimation procedures then outline the relevant issues to establish the methodology. We suppose that:

$$X_i = Z_i'\gamma + \eta_i, \tag{5.4}$$

where $\eta \perp z$, $E(\epsilon|z) = 0$ and $\hat{\gamma} = \left(\sum_i Z_i Z_i'\right)^{-1} \sum_i Z_i X_i'$. Then the generalised NL2SLS estimation procedure is summarised as below:

Generalised NL2SLS estimation Procedure

Step 0: Transform (5.1) using the IV projection matrix:

$$\begin{aligned}
P_Z Y_i &= P_Z X_i'\beta_0 + P_Z g(V_{0,i}) + P_Z \epsilon_i \\
\tilde{Y}_i &= \tilde{X}_i'\beta_0 + \tilde{g}(V_{0,i}) + e_i,
\end{aligned} \tag{5.5}$$

where $\tilde{Y}_i = Z_i'\left(\sum_i Z_i Z_i'\right)^{-1} \sum_i Z_i Y_i$, $\tilde{X}_i = Z_i'\left(\sum_i Z_i Z_i'\right)^{-1} \sum_i Z_i X_i'$, $\tilde{g}(V_{0,i}) = Z_i'\left(\sum_i Z_i Z_i'\right)^{-1} \sum_i Z_i g(V_{0,i})$ and $e_i = Z_i'\left(\sum_i Z_i Z_i'\right)^{-1} \sum_i Z_i \epsilon_i$.

Step 1: Perform Steps 3.2.1.1 to 3.2.1.4 of Procedure 3.2.1 in Chapter 3 on (5.5) to obtain the estimators of β_0 and α_0.
Step 2: Given $\hat{\beta}$ and $\hat{\alpha}$, obtain the estimate of the reduced relation, i.e. $\tilde{g}(\cdot)$. We have:

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

Step 3: Recover the structural relation, i.e. $g(\cdot)$, from the previous step using $P_Z^{-1} = \{Z(Z'Z)^{-1}Z'\}^{-1}$. We have:

$$\hat{g}(\hat{v}) = P_Z^{-1}\hat{\tilde{g}}(\hat{v}) = P_Z^{-1}\hat{E}(\tilde{y}|\hat{v}) - P_Z^{-1}\hat{E}(\tilde{x}|\hat{v})'\hat{\beta}.$$

In the following paragraphs, we discuss Step 3 of the above estimation procedure in greater detail. We consider the similar argument as the exclusion restriction ($\eta \perp z$) in the CF approach literature in order to recover the structural relation from the reduced one. Given β_0 and α_0, we rewrite (5.5) as follows:

$$\tilde{Y}_i^* = \tilde{g}(V_{0,i}) + e_i,$$

where $\tilde{Y}_i^* = \tilde{Y}_i - \tilde{X}_i'\beta_0$, then we have:

$$E(\tilde{y}^*|x) = E(P_Z y^*|z,\eta) = P_Z E(y^*|z,\eta) \equiv P_Z E(y^*|x).$$

Hence we obtain the unknown structural function as shown below:

$$\begin{aligned} g(v_0) &= P_Z^{-1}\tilde{g}(v_0) \\ &= P_Z^{-1}E(y|v_0) - P_Z^{-1}E(x|v_0). \end{aligned}$$

Although the proposed models, i.e. (5.1) and (5.4), are similar to that of the triangular structure discussed in Chapter 3, the proposed methodology is the generalisation of the NL2SLS estimation method to the semiparametric case which is distinctive to that of the CF approach.

We aim to investigate the asymptotic properties of the all unknown estimators of the generalised NL2SLS estimation methods. Furthermore, we also intend to investigate whether (5.4) can be extended to the nonparametric case to provide more flexibility which does not require a tight functional form relation between the endogenous regressors and their IV. More importantly, we investigate the practicality of Xia et al. (1999); which allows the same smoothing parameter in the estimation of the index coefficients and the structural unknown function, is still applicable in the proposed estimation procedure.

Bibliography

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

Amemiya, T. (1974), 'The non-linear two-stage least-squares estimator', *Journal of Econometrics* **2**, 105–110.

Amemiya, T. (1985), *Advanced Econometrics*, Harvard University Press, Cambridge, MA.

Banks, J., Blundell, R. & Lewbel, A. (1997), 'Quadratic Engel curves and consumer demand', *Review of Economics and Statistics* **79**(4), 527–539.

Blundell, R. (1988), 'Consumer behaviour: theory and empirical evidence', *Economic Journal* **98**(389), 16–65.

Blundell, R., Browning, M. & Crawford, I. (2003), 'Nonparametric Engel curves and revealed preference', *Econometrica* **71**(1), 205–240.

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

Blundell, R., Duncan, A. & Pendakur, K. (1998), 'Semiparametric estimation and consumer demand', *Journal of Applied Econometrics* **13**(5), 435–461.

Blundell, R. & Powell, J. (2003), *Endogeneity in Nonparametric and Semiparametric Regression Models: Advances in Economics and Econometrics, Eighth World Congress*, Vol. 2, Cambridge University Press, Cambridge, UK.

Blundell, R. & Powell, J. (2004), 'Endogeneity in semiparametric binary response models', *Review of Economics Studies* **71**(3), 655–679.

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

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

Deaton, A. & Muellbauer, J. (1980*a*), 'An almost ideal demand system', *The American Economic Review* **70**(3), 312–326.

Deaton, A. & Muellbauer, J. (1980*b*), *Economics and Consumer Behavior*, Cambridge University Press, Cambridge, UK.

Engle, R. F., Granger, C., Rice, J. & Weiss, A. (1986), 'Semiparametric estimates of the relationship between weather and electricity sales', *Journal of the American Statistical Association* **81**(394), 310–320.

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

Fan, J. & Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, Vol. 66, Chapman & Hall/CRC.

Fan, J. & Li, Q. (1999), 'Root-n-consistent estimation of partially linear times series models', *Journal of Nonparametric Statistics* **11**(1-3), 251–269.

Fan, J., Li, Q. & Stengos, T. (1995), 'Root-n-consistent semiparametric regression with conditional heteroscedastic disturbances', *Journal of Quantitative Economics* **11**, 140–229.

Gao, J. (2007), *Nonlinear Time Series*, Vol. 108, Chapman & Hall/CRC.

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

Gong, X., van Soest, A. & Zhang, P. (2005), 'The effects of the gender of children on expenditure patterns in rural China: a semiparametric analysis', *Journal of Applied Econometrics* **20**(4), 509–527.

Gorman, W. M. (1959), 'Separable utility and aggregation', *Econometrica: Journal of the Econometric Society* pp. 469–481.

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

Härdle, P. & Horowitz, J. (2005), 'Non-parametric method for inference in the presence of Instrumental Variables', *The Annals of Statistics* **33**(6), 2904–2929.

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

Härdle, W., Liang, H. & Gao, J. (2000), *Partially Linear Models. Springer Series in Economics and Statistics*, Springer/Physica-Verlag, New York.

Härdle, W. & Marron, J. (1990), 'Semiparametric comparison of regression curves', *The Annals of Statistics* pp. 63–89.

Hastie, T. & Tibishirani, R. (1991), *Generalized Additive Models*, Chapman and Hall, Boca raton, Florida.

Hausman, J. (1978), 'Specification tests in econometrics', *Econometrica* **46**(6), 1251–1271.

Horowitz, J. (2009), *Semiparametric and Nonparametric Methods in Econometrics*, Springer Science and Business Media, New York, USA.

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

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

Jorgenson, D. W. & Slesnick, D. T. (1987), 'Aggregate consumer behavior and household equivalence scales', *Journal of Business & Economic Statistics* **5**(2), 219–232.

Li, Q. (1999), 'Consistent model specification tests for time series econometric models', *Journal of Econometrics* **92**(1), 101–147.

Li, Q. & Racine, J. S. (2011), *Nonparametric Econometrics: Theory and Practice*, Princeton University Press, Princeton.

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

Linton, A. & Nielsen, J. (1995), 'A kernel method of estimating structural nonparametric regression based on marginal integration', *Biometrika* **82**(1), 93–100.

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

Nakamura, A. & Nakamura, M. (1998), 'Model specification and endogeneity', *Journal of Econometrics* **83**(1-2), 213–237.

Newey, W. & Powell, J. (2003), 'Instrumental variable estimation of nonparametric models', *Econometrica* **71**(5), 1565–1578.

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

O'Sullivan, F. (1986), 'A statistical perspective on ill-posed inverse problems', *Statistical Science* **1**(4), 502–518.

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

Pinkse, C. & Robinson, P. (1995), 'Pooling nonparametric estimates of regression functions with a similar shape. in: Advances in econometrics and quantitative economics', *In: Advances in Econometrics and Quantitative Economics. Wiley-Blackwell* pp. 172–195.

Pinkse, J. (2000), 'Nonparametric two-step regression estimation when regressors and error are dependent', *The Canadian Journal of Statistics* **28**(2), 289–300.

Pollack, R. & Wales, T. (1995), *Demand System Specification and Estimation*, Oxford University Press, New York, USA.

Robinson, P. (1988), 'Root-N-consistent semiparametric regression', *Econometrica* **56**(4), 931–954.

Saart, P. W. & Gao, J. (2013), 'Semiparametric methods in non-linear time series analysis: a selective review', *available at:* `https://sites.google.com/site/patrickwsaart/home` .

Sargan, J. D. & Desai, M. (1988), *Lectures on Advanced Econometric Theory*, B. Blackwell, Oxford.

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

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

Tjøstheim, D. & Austad, R. (1996), 'Nonparametric identification of nonlinear time series: projection', *Journal of American Statistical Associtation* **89**(428), 1398–1409.

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

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

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

Yatchew, A. (2003), *Semiparametric Regression for the Applied Econometrician*, Cambridge University Press, Cambridge, UK.