EXACT ASYMPTOTIC EXPANSIONS OF THE ESTIMATES OF MULTIVARIATE REGRESSIONS WITH OMITTED VARIABLES: A UNIFIED FRAMEWORK APPROACH

SHIMING HAO

Hunan University

In this paper, we investigate several parametric, nonparametric and semi-parametric estimators' explicit asymptotic biases attributed to the endogeneity caused by omitting variables in multivariate regressions. A unified framework to derive and compare these asymptotic biases are proposed, explicit estimations of these asymptotic biases can be then identified. The conditions are given out where the asymptotic bias of the Nadaraya-Watson's estimation of the nonparametric model is smaller than that of the least-squared estimation of the parametric linear model, and where the asymptotic bias of the Robinson's estimation of the partial linear model is smaller than that of the linear model. It is shown that the condition for the Robinson's estimator to dominate the OLS is weaker than the condition for the Nadaraya-Watson's estimator to dominate the OLS, and the partial linear model's IV estimation has a faster convergent speed to the true parameter value than the linear model as the partial linear model's IV estimation owns a smaller finite sample bias when dealing with omitted variables. Asymptotic biases of some other nonparametric estimators are also derived and compared under this framework where some explanatory variables are omitted. We show that the asymptotic bias of the Gasser-Müller's estimator is smaller than that of the Nadaraya-Watson's estimator, and the asymptotic bias of the local linear estimator is the same as that of the Gasser-Müller's estimator. In addition, regressions in which there are some explanatory variables omitted and in which there are location shifts in the endogenous regressors (Phillips and Liangjun Su, 2011) are also studied under this framework. We find that all the estimators are consistent without instrument variables and the nonparametric estimators with location shifts will always dominate the ones without location shifts under a weaker assumption. Monte Carlo simulations demonstrate the finite sample performances of these proposed properties.

1. INTRODUCTION

This study proposes a unified framework to derive, identify and compare the asymptotic performances of the parametric and nonparametric estimations when there are several explanatory variables omitted from regression models. As far as we know, the proposed approach has not been considered in the literature yet. This study is motivated by the fact that, under some certain conditions, the nonparametric methods have some properties dominating the parametric methods when dealing with endogeneity problems, see e.g., Wang and Phillips (2009a, 2009b, 2011), Phillips and Liangjun Su (2009, 2011), etc. It is well known that omitted variables, measurement errors, simultaneous causality, use of lagged values of the dependent variables as explanators in the presence of series correlation and selection bias are the main reasons for endogeneity. Among them, omitted variables will cause biased inconsistent

The author is deeply grateful to Xingyue Hao, Zhiping Ying for their guidance and help. The author is also grateful to an earlier review that has led to a substantial improvement of this paper. The author thanks Hans Mueller and Paul Schrimpf for helpful discussions. This research is supported by the Fundamental Research Funds for the Hunan University, and the Research Funds of Hunan University of China, #18XNF014. Address correspondence to Shiming Hao, Provincial Key Lab of Jurimetrics, Department of Mathematics and Econometrics, Hunan University, Changsha 410082, China; e-mail: haoshiming@hnu.edu.cn.

estimation and misleading inferences, which is a problem that nearly every empirical study has to face. Although the endogeneity problem has got much attention in modern econometric literatures, there is still a lack of a unified review and summary on the studies of omitted variables, and a unified comparison framework for the performances of different estimation methods under omitted variables when most studies mainly concern the roles of omitted variables in different model settings and specifications.

Classical textbooks assess the sensitivity of certain estimator $\hat{\beta}$ to the exclusion or inclusion of certain specific factors by tracing the effects of these omitted variables to its bias and inconsistency, and finding that the bias and inconsistency can be very huge even though the sample size is large enough, e.g., Green (1997), Wooldridge (2003) and Pearl (2000) etc. The degree and direction of bias or inconsistency can be evaluated by different kinds of ways, such as geometric approach (Stafford and Andrews, 1993), genetic algorithms approach (Sessions and Stevans, 2006) and exact sampling distribution approaches (Kinal and Lahiri, 1981). An asymptotic bias formula for the use of a generated vector with a dependent error structure compared to that of the estimating vector derived from a genetic algorithm under omitted variables was given by Sessions and Stevans (2006), who discovered that the estimation bias can be reduced significantly using the genetic algorithm proxy method. Toro-vizcarrondo and Wallace (1969, 2014) and Feldstein (1973) found that the MSE of the estimate of the parameter β can be reduced by omitting a nuisance variable if and only if the ratio of the nuisance parameter to the standard error of its estimate is less than one. Kinal and Lahiri (1981) derived the exact sampling distribution of the omitted variables estimator and found out the conditions under which the MSE and efficiency of the stochastic regressor dominate the non-stochastic one when some of the regressors are omitted. Apart from these, the finite and infinite sample behaviors of the omitted variables estimator and proxy estimator in measurement error model settings were compared by McCallum (1972), who pointed out that if the measurement errors occasioned by the use of the proxy variables are random and independent of the true regressor, then the asymptotic bias will be smaller if the proxy is used and the missing variable is simply omitted.

Beside parametric model specifications and estimations, some special properties of nonparametric models under endogeneity have also been explored by econometricians. Recent works by Wang and Phillips (2009a, 2009b, 2011) have shown that ill-posed inverse problem does not arise in either zero energy function of nonstationary time series or structural cointegrating non-parametric regressions and there is no need for non-parametric instrumental estimation. Simple Nadaraya-Watson nonparametric estimation is consistent irrespective of the endogeneity in regressors. Then a similar work by Phillips and Liangjun Su (2009, 2011) found that a nonparametric structural estimation will be $\sqrt{nh/L_n}-$consistent without using instrument variables if there are continuous location shifts in the endogenous regressors. These location shifts serve the role of a form of cross-section non-stationarity and in this sense the resulting consistency of nonparametric regression is analogous to that achieved in the case of unit root regressors. The results are also suitable for panel data models. All of these studies seem to prove that, under certain conditions, nonparametric methods are superior to parametric methods in dealing with endogeneity problems. As pointed out by Phillips and Liangjun Su (2011), nonparametric regression may display robustness to endogeneity in a regression by concentrating attention on local information and attennating tail information that may be more heavily subjected to endogeneity effects, the first-order bias under endogeneity can be eliminated in local regressions, while the parametric model uses full and global information. Hereat, this paper devotes to proposing a unified framework to derive and compare the accurate omitted variables regressions' asymptotic biases of the nonparametric and parametric methods, where different nonparametric methods and semiparametric methods are also investigated as well. According to the existing literature, it is difficult to derive the explicit asymptotic bias due to omitted variables for the parametric and nonparametric models in multivariate settings (Wooldridge, 2010). If there is a unified framework for us to compare their asymptotic biases with omitted variables, under what conditions the nonparametric and semiparametric approach dominate the parametric ones in dealing with omitted variables and what we can learn from these are still open questions in front of us.

In this paper, in order to solve the questions proposed above, we set up a unified framework under which the asymptotic biases of the parametric and nonparametric multivariate regressions with omitted variables can be derived and compared by introducing a kind of separable and additive regression error term with nonparametric structures. In this framework, the omitted variables enter the regression error term to describe the real situation, and then we give out the conditions under which the nonparametric methods dominate the parametric ones. These conditions can be easily satisfied by most social and economic empirical studies. Apart from this, the asymptotic biases of the local linear kernel estimation (Fan and Gijbels, 1992), Nadaraya-Watson's estimation (Nadaraya, 1989) as well as the Gasser- Müller's estimation (Gasser and Müller, 1979) are derived and compared respectively. The asymptotic bias of the Robinson's estimation (Robinson, 1988) of a partial linear model where there are several explanatory variables omitted is also derived and compared to a linear model using this framework. We then extend Phillips and Liangjun Su (2009, 2011)'s model to multivariate settings with location shifts in the endogenous regressors and where several variables are omitted. Our paper demonstrates some of the extraordinary and interesting properties of the nonparametric and semiparametric methods when dealing with omitted variables problems, and suggests that nonparametric or semiparametric methods should be well considered in empirical studies if the numbers of the included explanatory variables are finite and the sample size is large enough.

The rest of the paper is organized as follows. In section 2, we propose a kind of separable and additive nonparametric regression error structure where omitted variables enter, and then set up a unified framework under which the asymptotic biases of the parametric and nonparametric estimations can be derived and compared. The identification conditions to derive the asymptotic biases using this framework are also presented. Apart from these, we explore the asymptotic bias of the estimated partial linear model with omitted variables and compare it with the linear models using this framework. In section 3, the exact asymptotic biases of the Nadaraya-Watson's estimation, Gasser- Müller's estimation and local linear estimation are derived and compared respectively. In section 4, we consider the estimation performances of the parametric, nonparametric and partial linear regression models where several variables are omitted and where there are location shifts in the endogenous-included regressors. Section 5 illustrates the Monte Carlo simulation studies and section 6 concludes the research.

2. A UNIFIED ASYMPTOTIC BIASES COMPARISON FRAMEWORK

According to the existing econometrical theories and literatures, we only know a rough concept that the estimators are biased and inconsistent due to omitted variables, and there are no exact expansions or formulas for these inconsistencies. In this section, we first introduce a data generating process for the omitted variables model and set up a unified framework under which the asymptotic biases of the parametric and nonparametric multivariate estimations with several variables omitted can be exactly derived and compared. The identification condition for these asymptotic biases is given then and the exact asymptotic biases are derived and compared. Finally, a partial linear model's asymptotic bias due to omitted variables are also derived and compared to linear models.

2.1. The data generating process for omitted variables model

Following Hall's et al. (2007) notations, suppose there are q_1 variables included in the regressions and $q - q_1$ variables omitted in an empirical study, where there should be actually q explanatory variables included totally in fact. Let $\bar{x} = \{x_1, x_2, \dots, x_{q_1}\}$ represent the explanatory variables included in the regression models, $\tilde{x} = \{x_{q_1+1}, x_{q_1+2}, \dots, x_q\}$ represent the explanatory variables omitted from the regression models and $x = (\bar{x}, \tilde{x})$ represent the variables that actually should be included in. When some of the variables are omitted from regression models, according to the usual econometric theories, the omitted variables can be seen enter the regression error term (see Wooldridge (2010), Baltagi (2014), Arellano (2014) etc.). Let us consider the response y is structurally generated as

$$y = g(\bar{x}, \tilde{x}, \varepsilon), \tag{2.1}$$

where g is an unknown measurable function, and $\bar{x}$ is of dimension $0 \le q_1 \le \infty$, $\tilde{x}$ is of dimension $0 \le q - q_1 \le \infty$. $y, \bar{x}, \tilde{x}$ are observable, and ε may be unobservable, $y, \bar{x}, \tilde{x}, \varepsilon$ are realizations of $Y, \bar{X}, \tilde{X}, E$ respectively. Usually, we consider variables $\tilde{x}$ are omitted from the model (2.1). Several assumptions are made then.

ASSUMPTION 2.1. *There exists a smooth, continuous and 3-order derivable product measurable function $U: \tilde{X} \times E \to R$ on the product measurable space $(\tilde{X} \times E, \sigma(\tilde{X}) \otimes \sigma(E))$, and $U(\emptyset) = 0$.*

ASSUMPTION 2.2. $U(\tilde{x}, \cdot)$ *is strictly increasing for all $\tilde{x} \in \tilde{\mathbb{X}}$, where $\tilde{\mathbb{X}}$ is the support of $\tilde{X}$, and that the conditional CDF of* $g(\varepsilon)$ *given $\bar{X} = \bar{x}$ is strictly increasing for all $\bar{x} \in \bar{\mathbb{X}}$, where $\bar{\mathbb{X}}$ is the support of $\bar{X}$.*

ASSUMPTION 2.3. $\bar{x} \perp \varepsilon \mid \tilde{x}$ and $\tilde{x} \perp \varepsilon \mid \bar{x}$.

ASSUMPTION 2.4. *The usual assumptions imposed on nonparametric kernels $\mathcal{K}(\cdot)$.*

These assumptions are similar to the ones made by Liangjun Su et.al (2015), Matzkin (2003) and Hoderlein and White (2012). Especially, Assumption 2.3 is the same as the Assumption A.2 proposed by Lu and White (2014), and Assumption 2.2 comes from their proposition 1 and proposition 2. When $\bar{x}$ is binary, Assumption 2.3 is equivalent to the unconfoundness assumption in the treatment effect literature, which plays a key role in identifying ATE and ATT. There are several cases in which Assumption 2.3 is plausible, and the omitted variables' case is the most common one in empirical studies. For example, let $\bar{x}$ be years of education and y be wages, ε represents other drives of wages, such as ability, which are unobservable. Education is endogenous in this case, let $\tilde{x}$ be an IQ test score, then it is plausible that conditioning on the IQ score, education and ability are independent of each other (Lu and White, 2014). The function defined in Assumption 2.2 is the regression error term for the omitted variables' regression models in fact, and intuition tells us that the value of the regression error term depends on the value of the omitted variables, that is the bigger values the omitted variables get, the larger the regression's error term will be.

THEOREM 2.1. *If Assumptions 2.1-2.4 hold true, then*

$$y = g_1(\bar{x}) + g_2(\tilde{x}, \varepsilon) \tag{2.2}$$

holds if and only if (a) $(\bar{x}, \tilde{x})$ *and* ε *are separable under (2.1) and (b)* $\bar{x}$ *and* $\tilde{x}$ *are also separable under (2.1).* $g_1(\cdot)$, $g_2(\cdot)$ *are continuous measurable unknown functions.*

Theorem 2.1 relates the nonseparable model (2.1) with omitted variables' regressions, showing that when the explanatory variables $\tilde{x}$ are omitted from regressions, under a completely separable condition or assumption, the correctly specified model (2.1) can be turned into a regression model (2.2) that only includes a response variable y, observable explanatory variables $\bar{x}$ and an error term where the omitted variables $\tilde{x}$ and other true exogenous shocks ε enter. The parametric and nonparametric models considered in this paper when several variables are omitted are then as follows:

$$y = x\beta + u(\tilde{x},\varepsilon), \tag{2.3}$$

$$y = g(x) + u(\tilde{x},\varepsilon), \tag{2.4}$$

where $x = (\bar{x},0)$, $\bar{x}$ are the only variables included in the regressions, variables $\tilde{x}$ are omitted and enter the regression error term according to Theorem 2.1. β and $g(x)$ are the unknown parameters and nonparameter function to estimate respectively. In these models, only y, $\bar{x}$ are observable variables, $u(\cdot,\cdot)$ is the regression error term and we suppose that $\mathrm{Var}(u(\cdot,\cdot)) = \sigma^2 < \infty$. What should be noticed is that, under Assumptions 2.1 to 2.3, model (2.2) is completely separable, and the error term of model (2.3), (2.4) are also separable; if the Assumptions 2.1 to 2.3 do not hold, then model (2.2) may be partially separable or nonseparable, and the regression error term of model (2.3), (2.4) are not separable. In this paper, we mainly consider the former situation.

2.2. Identification through conditional independence

Consider the nonadditive omitted variables' regression model derived from (2.1)

$$y = g_1(\bar{x},\varepsilon_1), \tag{2.5}$$

where $\tilde{x}$ are omitted, the included variables $\bar{x}$ and the regression error ε_1 are not independently distributed and g_1 is strictly increasing in ε_1. A standard example is where y denotes years of education, $\tilde{x}$ denote abilities, ε_1 denote the effects of other unobservable variables (see Chesher (2003), Imbens and Newey (2009)). Without additional conditions, identifying the causal effect of the observable variables $\bar{x}$ on the outcome variable y is typically not possible in such a situation. However, from Theorem 2.1, this identification of unknown functions and distributions can be achieved by the conditional independence method developed by Matzkin (2003,2007, 2008), Chesher (2003) and Imbens and Newey (2009). Notice that from Theorem 2.1, if the model satisfies the assumptions, we can actually treat the omitted variables separablly into the error term of (2.5), then the error term of the omitted variables' regression becomes a function of variables $\tilde{x}$ and the true exogenous shocks ε

$$\varepsilon_1 = g_2(\tilde{x},\varepsilon), \tag{2.6}$$

ε are the true unobservable exogenous shocks from model (2.1). Then the system of the two above equations (2.5), (2.6) makes up a triangular system (Hausman, 1983). From Assumption 2.3, we can see that ε is independent of $(\bar{x},\tilde{x})$, then each coordinate of g_1 will achieve values independent of the other coordinates when conditioning on at least one value of $\tilde{x}$. By this conditional independence, we are able to show that g_1 and the distribution of $(\bar{x},\varepsilon_1)$ can both be identified.

THEOREM 2.2. *Under Assumptions 2.1-2.4, if* g_1 *is strictly increasing in* ε_1, $\mathcal{F}_{\varepsilon_1,\bar{x}|\tilde{X}=\tilde{x}}$ *is strictly increasing in* $(\bar{x},\varepsilon_1)$ *for each* $\bar{x}$, $\mathcal{F}_{\varepsilon_1|(\bar{X},\tilde{X})=(\bar{x},\tilde{x})}$ *is strictly increasing in* ε_1, *and* g_2 *is also strictly increasing in* ε *given* $\tilde{X} = \tilde{x}$. *Then for all* $\bar{x}$, $\tilde{x}$, e_1 *and* e

$$g_1(\bar{x},e_1) = \mathcal{F}^{-1}_{y|\bar{X}=\bar{x},\tilde{X}=\tilde{x}}(\mathcal{F}_{\varepsilon|\tilde{X}=\tilde{x}}(e)) \text{ and}$$

$$\mathcal{F}_{\varepsilon_1|\bar{X}=\bar{x}}(e_1) = \mathcal{F}_{y|\bar{X}=\bar{x}}(\mathcal{F}^{-1}_{y|\bar{X}=\bar{x},\tilde{X}=\tilde{x}}(\mathcal{F}_{\varepsilon|\tilde{X}=\tilde{x}}(e))),$$

where $\mathcal{F}_{\varepsilon|\tilde{X}=\tilde{x}}(e)$ denotes the CDF of e given $\tilde{X} = \tilde{x}$.

This theorem shows that by conditional independence, g_1 and the distribution of $(\bar{x},\varepsilon_1)$ can be identified uniquely, up to a normalization on the distribution of ε_1 given $\widetilde{X} = \tilde{x}$. In this sense, one can show that they are consistent estimators by nonparametric estimations of the distributions. This result is just equivalent to the identifications considered in Matzkin (2007,2008) due to the equivalent theorem therein.

2.3. Asymptotic biases of the parametric and nonparametric regressions

We first derive a different version for the asymptotic expansion of the nonparametric local constant regression model (2.1) from Li and Racine (2007). By the law of large numbers, when there are no variables omitted from regression and the sample size growths to infinite, we have

$$\begin{aligned}
\hat{g}(x) &= \frac{\sum_{i=1}^{q} y_i \mathcal{K}\left(\frac{X_i - x}{h}\right)}{\sum_{i=1}^{q} \mathcal{K}\left(\frac{X_i - x}{h}\right)} \\
&= \frac{\sum_{i=1}^{q} (g(x_i) + u_i)\mathcal{K}\left(\frac{X_i - x}{h}\right)}{\sum_{i=1}^{q} \mathcal{K}\left(\frac{X_i - x}{h}\right)} \\
&= \frac{\frac{1}{n}\sum_{i=1}^{q} g(x_i)\mathcal{K}\left(\frac{X_i - x}{h}\right) + \frac{1}{n}\sum_{i=1}^{q} u_i\mathcal{K}\left(\frac{X_i - x}{h}\right)}{\frac{1}{n}\sum_{i=1}^{q} \mathcal{K}\left(\frac{X_i - x}{h}\right)} \qquad (2.7) \\
&\longrightarrow_p \frac{\mathbb{E}\left(g(x)\mathcal{K}\left(\frac{X - x}{h}\right)\right) + \mathbb{E}\left(uK\left(\frac{X - x}{h}\right)\right)}{\mathbb{E}\left(\mathcal{K}\left(\frac{X - x}{h}\right)\right)},
\end{aligned}$$

where $u(\cdot)$ is the nonparametric regression's error term, and $\mathcal{K}\left(\frac{X_i - x}{h}\right) = \prod_{s=1}^{q} \mathcal{K}\left(\frac{X_{is} - x}{h}\right)$. According to Assumption 2.4, when there are no variables omitted and the error term is exogenous, by Taylor expansions we are able to show that $\hat{g}(x) \longrightarrow_p g(x)$, the above Nadaraya-Watson estimator is consistent (see the proof of Theorem 2.3). When there are variables omitted, the nonparametric estimator is no longer consistent, the next theorem gives the explicit asymptotic biases expansions for the nonparametric and parametric regression estimators of model (2.3), (2.4) using the framework set up in section 2.1 and 2.2

THEOREM 2.3. *Under Assumptions 2.1-2.4, as* $n \longrightarrow \infty$, *we have*

(i) $\hat{g}(x) \longrightarrow_p g(x) + \Omega(\tilde{x}) + \Phi(\tilde{x})$ *and*

(ii) $\hat{\beta} \longrightarrow_p \beta + C\mathbb{E}(x\Phi(\tilde{x}))$,

where $\mathbb{E}(x\Phi(\tilde{x})) = (\mathbb{E}x_1\Phi(\tilde{x}), \mathbb{E}x_2\Phi(\tilde{x}), \ldots, \mathbb{E}x_{q_1}\Phi(\tilde{x}), 0, \ldots, 0)'$ *,and* $\mathcal{C}$ *is the inverse matrix for* $Plim_{n\to\infty}(x'x/n)$. $\Omega(\cdot)$, $\Phi(\cdot)$ *are strictly increasing and continuous nonnegative functions of the omitted variables* $\tilde{x}$, *whose analytic expressions are provided in the Appendix.*

We get the explicit general asymptotic biases of parametric and nonparametric regressions when there exist omitted variables by imposing the same nonparametric error structure and a nonparametric estimation of the error terms. In this theorem, when there are no variables omitted, $\Omega(\tilde{x}) = \Phi(\tilde{x}) = 0$, the parametric and nonparametric are all consistent; when there are several variables omitted, the estimators are asymptotically biased and the biases are functions of the omitted variables $\tilde{x}$. The asymptotic bias of the ordinary least squared estimation of the linear model (2.3) is $\mathcal{C}_{i1}\mathbb{E}x_1\Phi(\tilde{x}) + \mathcal{C}_{i2}\mathbb{E}x_2\Phi(\tilde{x}) + \dots + \mathcal{C}_{iq_1}\mathbb{E}x_{q_1}\Phi(\tilde{x})$, and can be seen as a sum of every little bias part. So we call the asymptotic bias of the parametric linear model as a kind of nonlinear additive error structure and the asymptotic bias of the nonparametric model as a kind of nonadditive error structure when some of the variables are omitted from regressions. The reason for the additivity for the parametric linear model's asymptotic bias is that the model is liner and additive, while the nonparametric model is nonlinear. The situation where there are no variables omitted can also be derived from this theorem, so the unified framework and method we use can be seen as a kind of generalized asymptotic expansions for the parametric and nonparametric estimations.

REMARK 2.1. From Theorem 2.3, it is easy to see that : (i) the more explanatory variables are omitted from model (2.3) and (2.4), the larger asymptotic biases of the parametric and nonparametric regressions' estimations will have; and (ii) the greater correlations among the omitted explanatory variables and the included explanatory variables, the larger asymptotic biases of the parametric and nonparametric regressions' estimations will have.

In Theorem 2.3, $\Omega(\cdot)$, $\Phi(\cdot)$ are all strictly increasing functions of the omitted variables $\tilde{x}$, so the asymptotic biases will become larger if more variables are omitted from regressions. Similar to Florence et al. (2012), we use the degrees of the correlations among the omitted variables (enter the error term) and the included explanatory variables to value the degree of endogeneity. The greater correlations among the variables, the larger degree of the endogeneity will be. From the proof of Theorem 2.3, it can be seen that $\mathbb{E}(x_i\Phi(\tilde{x})) = \mathrm{Cov}(x_i,\Phi(\tilde{x})) + \mathbb{E}(x_i)\mathbb{E}(\Phi(\tilde{x}))$ for $i = 1,2,\dots,q_1$. So when the values of $\mathrm{Cov}(x_i,\Phi(\tilde{x}))$ get bigger, the asymptotic bias of the parametric linear regression will become larger. Similarly, it also can be shown that $\mathbb{E}\left(u(\tilde{x})\mathcal{K}\left(\frac{X-x}{h}\right)\right) = \mathrm{Cov}\left(u(\tilde{x}),\mathcal{K}\left(\frac{X-x}{h}\right)\right) + \mathbb{E}(u(\tilde{x}))\mathbb{E}\left(\mathcal{K}\left(\frac{X-x}{h}\right)\right)$ for $i = 1,2,\dots,q_1$. So when the value of $\mathrm{Cov}\left(u(\tilde{x}),\mathcal{K}\left(\frac{X-x}{h}\right)\right)$ gets bigger, the asymptotic bias of the nonparametric regression will become larger.

ASSUMPTION 2.5. $x_i \geq 1$ *for all* $\bar{x}$, $i = 1,2,\dots,q_1$.

Bounded random variables are widely used in econometric and statistical literatures; Assumption 2.5 considers the situation where every included explanatory variable in the regression models has a lower bound. This assumption can be satisfied in most economic and social empirical studies, and if not, easy data transformation or changing measurement methods can achieve this assumption. Then, we have a theorem to express the relationships between these two asymptotic biases when several variables are omitted

THEOREM 2.4. *Under assumptions 2.1-2.5, as* $n \longrightarrow \infty$, *we have*

$$\hat{g}(x) - g(x) = \mathcal{O}_{\mathcal{A}}(\hat{\beta} - \beta)$$

holds on at least one of the sets $\mathcal{A} \in \{ \bigcap_{i=1}^{q_1} A_i \}$, *where* $A_i = x_i \cap \tilde{x}$ *for* $i = 1,2,\dots,q_1$.

Theorem 2.4 shows that for the true data generating process $(y, \bar{x}, \tilde{x}) \in \mathfrak{R}^{q+1}$ mentioned in section 2.1, we are always able to find at least an intersection set $\mathcal{A}$ of the included variables $\bar{x}$ and the omitted variables $\tilde{x}$ on which the asymptotic bias of the nonparametric regression is smaller than the asymptotic bias of the parametric linear regression. We also find that this result is robust to the value of

$\Omega(\bar{x})$, whether $\Omega(\bar{x}) > 1$ or $\Omega(\bar{x}) \leq 1$. The primary reasons for this theorem are the different error structures and estimation methods of these two different models. As for the parametric linear model (2.3), the asymptotic bias of the least squared estimation $\mathcal{CE}(x\Phi(\tilde{x}))$ satisfies the principles of superposition, and the essence of the additive nonlinear error structure is that the total relationship between the included variables $\bar{x}$ and the omitted variables $\tilde{x}$ equals to the sum of the relationships between every included variable x_i and all the omitted variables $\tilde{x}$ for $i = 1,2,...,q_1$. There is a linear relationship between the degree of endogeneity, the numbers of variables omitted and the asymptotic bias of the parametric linear model, implicating that the asymptotic bias has a monotone increasing trend with a larger degree of endogeneity and more variables omitted from the model. On the other hand, as for the nonparametric regression model (2.4), the asymptotic bias of the nonparametric kernel estimation $\Omega(\bar{x}) + \Phi(\tilde{x})$ does not satisfy the principles of superposition, and the essence of the nonadditive error structure is that the total relationship between the included variables $\bar{x}$ and the omitted variables $\tilde{x}$ is not equal to the sums of every little part as shown in the linear model. There is a nonlinear relationship between the degree of endogeneity, numbers of variables omitted and the size of the asymptotic bias. And just because of the linear super-positional property of the parametric linear model's additive error structure and the nonlinear non-super-positional property of the nonparametric model's nonadditive error structure, we are able to find a region $\mathcal{A}$ where the asymptotic bias of the nonparametric model is smaller than the parametric linear model when the degree of model endogeneity is increasing. The result of Theorem 2.4 also holds for the operator $\mathcal{O}_P(\cdot)$ as shown in Theorem 2.3.

LEMMA 2.1. *Under Assumptions 2.4-2.5, let* $\mathcal{C}_i x_i \Phi(\tilde{x})$ *be a real random variable satisfying* $0 \leq \Phi(\tilde{x}) \leq x_i\Phi(\tilde{x}) \leq \delta_i$ *,where* $\delta_i \in \mathfrak{R}^+$. *Then* $\mathrm{Var}(\mathcal{C}_i x_i \Phi(\tilde{x})) \leq \mathcal{C}_i(\delta_i - \Phi(\tilde{x}))^2/4$ *for* $i = 1,2,...,q_1$.

LEMMA 2.2. *If* $\mathbb{P}\{|\mathcal{C}_i x_i \Phi(\tilde{x})| \leq \delta_i\} = 1$, $\mathbb{P}\{|\mathcal{C}_j x_j \Phi(\tilde{x})| \leq \delta_j\} = 1$ *hold true for some constants* δ_i , $\delta_j \in \mathfrak{R}^+$ *respectively, then we have*

$$|\mathrm{Cov}(\mathcal{C}_i x_i \Phi(\tilde{x}), \mathcal{C}_j x_j \Phi(\tilde{x}))| \leq 4\alpha_{ij}\delta_i\delta_j, \qquad \textbf{(2.8)}$$

where $\alpha_{ij} = Sup_{\mathcal{A}\in\sigma(\mathcal{C}_i x_i \Phi(\tilde{x})),\mathcal{B}\in\sigma(\mathcal{C}_j x_j \Phi(\tilde{x}))}|\mathbb{P}(\mathcal{A})\mathbb{P}(\mathcal{B}) - \mathbb{P}(\mathcal{AB})|$ *, and* $i \neq j \in \{1,2,...,q_1\}$.

If we treat all the regressors in the regression models as random variables, then $x_i\Phi(\tilde{x})$ is a Borel measurable function of the included random variables $x_1,x_2,...,x_{q_1}$ and the omitted random variables $x_{q_1+1},x_{q_1+2},...,x_q$. So it is a random variable since the composition of measurable functions is also measurable. By Assumption 2.5, $x_i\Phi(\tilde{x})$ is a bounded variable. So its variance should also be bounded (Audenaert, 2010). Lemma 2.1 gives out an upper bound for its variance while other kinds of more explicit upper and lower bounds can be achieved. In regression functions, according to Assumption 2.3, we suppose that $\bar{x}, \tilde{x} \perp \varepsilon$, but there actually should be correlations (not completely multicollinearity) among variables $\bar{x}$ and variables $\tilde{x}$, $\mathrm{Cov}(x_i\Phi(\tilde{x}),x_j\Phi(\tilde{x})) \neq 0, i \neq j \in \{1,2,...,q_1\}$. As for these weakly dependent variables, we introduce the $\alpha(n)-$index to value the dependencies among $\mathcal{C}_i x_i \Phi(\tilde{x})$, and Lemma 2.2 gives out the upper bound of the covariance of the random variables $\mathcal{C}_i x_i \Phi(\tilde{x})$, $i = 1,2,...,q_1$.

THEOREM 2.5. *Let* $x_1\Phi(\tilde{x}), x_2\Phi(\tilde{x}), \ldots, x_q\Phi(\tilde{x})$ *be a sequence of real random variables where* $\Phi(\tilde{x}) \leq x_i\Phi(\tilde{x}) \leq \delta_i$ *,and* $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_q$ *be a sequence of real numbers. Set* $\mathcal{S}_n = \sum_{i=1}^{q} \mathcal{C}_i x_i \Phi(\tilde{x})$ *,suppose that* $\mathbb{E}(\mathcal{S}_n)$ *and* $\mathrm{Var}(\mathcal{S}_n) < \infty$ *all exist ,then it holds that*

$$\mathbb{P}\{|\mathcal{S}_n - \mathbb{E}(\mathcal{S}_n)| \geq \lambda\} \leq \frac{\sum_{i=1}^{q}(\mathcal{C}_i\delta_i - \mathcal{C}_i\Phi(\tilde{x}))^2/4 + 4\sum_{i\neq j=1}^{q}\alpha_{ij}\delta_i\delta_j}{\lambda^2}, \qquad \textbf{(2.9)}$$

where $\lambda = Sup_{\mathcal{A}}(\Omega(\bar{x}) + \Phi(\tilde{x}))$ is a real number.

Assumption 2.5 and Theorem 2.5 imply that the probability $\mathbb{P}\{\mathbb{E}(\mathcal{S}_n) \le \mathcal{S}_n - \lambda\}$ or $\mathbb{P}\{\mathbb{E}(\mathcal{S}_n) \ge \mathcal{S}_n + \lambda \ge \lambda\}$ is decided by the correlations $\alpha_{ij}(n)$ between the variables $\mathcal{C}_i x_i \Phi(\tilde{x})$ and $\mathcal{C}_j x_j \Phi(\tilde{x})$. When other factors remain unchanged, the larger degree of the models' endogeneity caused by omitting variables is, the bigger upper bound of the probability $\mathbb{P}\{\mathbb{E}(\mathcal{S}_n) \ge \lambda\}$ will have. Hence, this theorem shows that if the degree of the models' endogeneity gets larger due to omitted variables, then the greater the probability of the asymptotic bias of the parametric linear regression is bigger than the asymptotic bias of the nonparametric regression. This means that, when the sample size is large enough and the number of included regressors is finite, for a strong endogeneity model whose several explanatory variables are omitted, the nonparametric estimation will provide a much smaller asymptotic bias estimator compared with the parametric OLS. The nonparametric model (2.4) considered in this paper outperforms the parametric linear model (2.3) due to omitted variables. In addition, what should be noticed is that the upper bound in Theorem 2.5 can be extended to other bounds, such as exponential upper bounds used in concentration inequalities of weighted and weakly dependent variables.

The rest part of this section will consider a semi-parametric partial linear model

$$y = \bar{x}_1\beta + g(\bar{x}_2) + u(\tilde{x},\varepsilon) \quad . \tag{2.10}$$

Comparing with the parametric linear model (2.3), we divide the included explanatory variables $\bar{x}$ into two parts $\bar{x} = (\bar{x}_1,\bar{x}_2)$, where variables $\bar{x}_1$ enter the parametric part and variables $\bar{x}_2$ enter the nonparametric part of the model (2.10). Variables $\tilde{x}$ are omitted and enter the regression error term, ε is the true exogenous unobservable shocks. What we are interested in now is the size relationship between the asymptotic bias of the partial linear model's estimation and the asymptotic bias of the linear model's estimation of the unknown parameter β when several explanatory variables are omitted. By Robinson (1988), we can get that

$$\begin{aligned}
\hat{\beta}_{PLM} &\triangleq \mathcal{H}^{-1}\mathcal{L}'(y - E(y|\bar{x}_2)) \\
&= \mathcal{H}^{-1}\mathcal{L}'(\bar{x}_1\beta + g(\bar{x}_2) + u(\tilde{x},\varepsilon) - \beta E(\bar{x}_1|\bar{x}_2) - \mathbb{E}(g(\bar{x}_2)) - \mathbb{E}(u(\tilde{x},\varepsilon)|\bar{x}_2)) \\
&= \beta + \mathcal{H}^{-1}\mathcal{L}'(u(\tilde{x},\varepsilon) - \mathbb{E}(u(\tilde{x},\varepsilon)|\bar{x}_2)) \\
&= \beta + (\mathcal{H}/n)^{-1}(\mathcal{L}'(g(\bar{x}_2) - \mathbb{E}(g(\bar{x}_2)))) + (\mathcal{H}/n)^{-1}(\mathcal{L}'(u(\tilde{x},\varepsilon) - \mathbb{E}(u(\tilde{x},\varepsilon)|\bar{x}_2))/n).
\end{aligned}$$

Using the law of large numbers, we then get

$$\begin{aligned}
\hat{\beta}_{PLM} &\longrightarrow_p \beta + \left(\frac{(\bar{x}_1 - \mathbb{E}(\bar{x}_1|\bar{x}_2))'(\bar{x}_1 - \mathbb{E}(\bar{x}_1|\bar{x}_2))}{n}\right)^{-1}\left(\frac{(\bar{x}_1 - \mathbb{E}(\bar{x}_1|\bar{x}_2))'(u(\tilde{x},\varepsilon) - \mathbb{E}(u(\tilde{x},\varepsilon)|\bar{x}_2))}{n}\right) \\
&\longrightarrow_p \beta + \\
&\frac{\mathbb{E}(\bar{x}_1 \cdot u(\tilde{x},\varepsilon)) - \mathbb{E}(\bar{x}_1 \cdot \mathbb{E}(u(\tilde{x},\varepsilon)|\bar{x}_2)) - \mathbb{E}(u(\tilde{x},\varepsilon) \cdot \mathbb{E}(\bar{x}_1|\bar{x}_2)) + \mathbb{E}(\mathbb{E}(\bar{x}_1|\bar{x}_2) \cdot \mathbb{E}(u(\tilde{x},\varepsilon)|\bar{x}_2))}{\mathbb{E}(\bar{x}_1 \cdot \bar{x}_1) - 2E(\bar{x}_1 \cdot \mathbb{E}(\bar{x}_1|\bar{x}_2)) + \mathbb{E}(\mathbb{E}(\bar{x}_1|\bar{x}_2) \cdot \mathbb{E}(\bar{x}_1|\bar{x}_2))},
\end{aligned}$$

where $\mathcal{H} = \mathcal{L}'\mathcal{L}$, $\mathcal{L} = (\bar{x}_1 - \mathbb{E}(\bar{x}_1|\bar{x}_2))$. Particularly, when there are no variables omitted, $u(\tilde{x},\varepsilon) = u(\varepsilon)$. According to the Assumption 2.3, using the law of total expectations, it is easy to see that the second term of the above equation tends to 0, which implies that $\hat{\beta}_{PLM} \longrightarrow_p \beta$ when no variables are omitted. If there are some explanatory variables omitted, the partial linear model's estimation is not consistent anymore. We can then prove that $\hat{\beta}_{PLM} - \beta = \mathcal{O}_{\mathcal{A}_0}(\hat{\beta}_{LM} - \beta)$ holds true where $\mathcal{A}_0$ is an intersection set of the variables $\bar{x}_1, \bar{x}_2$ and $\tilde{x}$ as shown in Theorem 2.4. The result of Theorem 2.4 is also valid for the relationship between the asymptotic bias of the semi-parametric partial linear model and the asymptotic bias of the parametric linear model, where the asymptotic bias of the partial linear model is smaller than that of the linear model when there are variables omitted.

3. ASYMPTOTIC EXPANSIONS FOR DIFFERENT NONPARAMETRIC METHODS

3.1. Generalized asymptotic bias of the Gasser-Müller's estimator

Suppose we have $n \times q$ measurements taken at the points $t_{k1}, t_{k2}, \ldots, t_{kn}$ $(0 \le t_{k1} \le t_{k2} \le \ldots \le t_{kn} \le 1)$, $k = 1,2,\ldots,q$ and $t_{ki} = (t_{1i}, t_{2i}, \ldots, t_{q_1 i}, t_{q_1+1i}, t_{q_1+2i}, \ldots, t_{qi}) = (\bar{t}_{ki}, \tilde{t}_{ki})$, $i = 1,2,\ldots,n$. $\bar{t}$ represents all the explanatory variables included in the regression while $\tilde{t}$ represents the omitted ones. The nonparametric model of omitted variables considered is the following

$$X(t_i) = \mathcal{U}(t_{ki}) + \mathcal{E}(\tilde{t}_{ki}, e),\ i = 1,2,\ldots,n;\ k = 1,2,\ldots,q, \tag{3.1}$$

where $t_{ki} = (\bar{t}_{ki}, 0)$, e are the regression's true exogenous unobservable shocks. The Gasser-Müller's estimator (1979) is defined then as a q-dimensional integration

$$\widehat{\mathcal{U}}_n(t) = \frac{1}{h_1 h_2 \cdots h_q} \sum_{j=1}^{q} \int_{s_{j-1}}^{s_j} \mathcal{K}\left(\frac{t-\mu}{h}\right) d\mu \cdot X(t_j), \tag{3.2}$$

$h_1, h_2, \cdots, h_q$ are the bandwidths for corresponding explanatory variables, $\mathcal{K}\left(\frac{t-\mu}{h}\right) = \prod_{i=1}^{q} \mathcal{K}\left(\frac{t_i-\mu}{h}\right)$, $s_j = \bigcup_{k=1}^{q} s_{kj}$, and $0 \le s_{k0} \le s_{k1} \le \ldots \le s_{kn}$, $t_{kj} \le s_{kj} \le t_{kj+1}$, $j = 1,2,\ldots,n-1; k = 1,2,\ldots,q$.

THEOREM 3.1. *Let $\mathcal{U}$ be ϑ times differentiable, if $\vartheta = 0$ then $\mathcal{U}$ is continuous, and $\mathcal{K}_\vartheta$ be a bounded function. Then, if (a) $\mathcal{U}^{(\vartheta)}$ is Lipschitz continuous at* t*; and (b)* $nh_1h_2 \cdots h_q \longrightarrow \infty$*, and* $h_k \longrightarrow 0$ *as* $n \longrightarrow \infty$ *for* $k = 1,2,\ldots,q$. *We then have*

$$\widehat{\mathcal{U}}_n(t) \longrightarrow_p \mathcal{U}(t) + \mathcal{E}(\tilde{t}) + \frac{1}{2}\mathcal{K}_2 \mathcal{E}_{ss}(\tilde{t}) + \mathcal{O}\left(h_{q_1+1}^2 + h_{q_1+2}^2 + \ldots + h_q^2 + 2/(n^2 h_{q_1+1} \cdots h_q)\right) + o_p(1),$$

where $\mathcal{E}_{ss}(\tilde{t}) = \sum_{s \neq s' = q_1+1}^{q} h_s^2 \frac{\partial^2 \mathcal{E}(\tilde{t})}{\partial t_s \partial t_{s'}}$, $\mathcal{K}_2 = \int_{-\infty}^{+\infty} \mathcal{K}(\mu)\mu^2 d\mu$.

Theorem 3.1 gives out the generalized asymptotic expansions for the Gasser-Müller's estimator of the nonparametric regression model (3.1), which is equivalent to model (2.4), using the framework set up in section 2. $\mathcal{U}(\cdot)$ in (3.1), denoting the unknown regression function, is equivalent to $g(\cdot)$ in (2.4); and X in (3.1), denoting the response variable, is equivalent to y. When there are no variables omitted, $\mathcal{E}(\tilde{t}) = 0$. So the asymptotic bias term of $\widehat{\mathcal{U}}_n(t)$ tends to be 0, which means $\widehat{\mathcal{U}}_n(t) \longrightarrow_p \mathcal{U}(t)$. Hence the Gasser-Müller's estimator is consistent. However, when there are variables omitted, then the Gasser-Müller's estimator is inconsistent, and the asymptotic bias is derived in Theorem 3.1. As shown in Theorem 2.3, 3.1, whether there are variables omitted or not, both the asymptotic expansions can be derived from the framework we set up.

THEOREM 3.2. *Under Assumptions 2.2-2.3, we have*

$$\hat{g}(x)_{GM} - g(x) = \mathcal{O}(\hat{g}(x)_{NW} - g(x))$$

as $n \longrightarrow \infty$, $h_k \longrightarrow 0$ *and* $nh_1h_2 \cdots h_q \longrightarrow \infty$ *for* $k = 1,2,\ldots,q$.

Similar to Theorem 2.4, we compare the asymptotic bias of the Nadaraya-Watson's estimator with that of the Gasser-Müller's estimator, and find that the asymptotic bias of the Gasser-Müller's estimator is much smaller than the asymptotic bias of the Nadaraya-Watson's estimator. The difference between Theorem 2.4 and Theorem 3.2 is that, when comparing the asymptotic biases of the parametric and nonparametric estimators, we find there exists at least an intersection set of the included variables and the omitted variables on which the asymptotic bias of the nonparametric Nadaraya-Watson's estimator is smaller than the asymptotic bias of OLS with the growing sample size; however, in

Theorem 3.2, we find that the asymptotic bias of the Gasser-Müller's estimator is always smaller than the asymptotic bias of the Nadaraya-Watson's estimator for all of the values the included variables have taken. This means that the condition for Theorem 3.2 to hold is weaker than that of Theorem 2.4. Actually the asymptotic bias of the Nadaraya-Watson's estimator has two more positive terms than the asymptotic bias of the Gasser-Müller's estimator, causing the bias of the Nadaraya-Watson's estimator to be larger than that of the Gasser-Müller (see the proof of Theorem 3.2 in the Appendix).

3.2. Generalized asymptotic bias of the local linear estimator

Since the sparsity of data in higher dimensions becomes more of a problem for higher order polynomials (*curse of dimensionality*), we only consider multivariate local linear regression here when there are several explanatory variables omitted (Jiangqin Fan et al., 1997). For the model (2.4), the local linear estimator of $g(x)$ is defined as

$$\hat{g}(x)_{LL} = \sum_{i=1}^{n} y_i \frac{\mathcal{K}_h(X_i - x)\left(\mathcal{S}_{n,2} - (X_i - x)\mathcal{S}_{n,1}\right)}{\mathcal{S}_{n,0}\mathcal{S}_{n,2} - \mathcal{S}_{n,1}^2}, \qquad \mathbf{(3.3)}$$

where $\mathcal{S}_{n,k} = \sum_{i=1}^{n}(X_i - x)^k \mathcal{K}_h(X_i - x)$, $X_i - x = \prod_{s=1}^{q}(X_{is} - x)$, $k = 0,1,2$ and $\mathcal{K}_h(X_i - x) = \prod_{s=1}^{q}\mathcal{K}_h(X_{is} - x)$, $\mathcal{K}_h(X_{is} - x) = \frac{1}{h}\mathcal{K}\left(\frac{X_{is}-x}{h}\right)$. By the law of large numbers, and multiply $\frac{1}{n^2}$ to the molecular and the denominator of the above equation, we can get

$$\hat{g}(x)_{LL} = \sum_{i=1}^{n} \frac{\frac{1}{(h_1h_2\cdots h_q)^2}\left(y_i\mathcal{K}\left(\frac{X_i-x}{h}\right)\sum_{i=1}^{n}(X_i-x)^2\mathcal{K}\left(\frac{X_i-x}{h}\right) - y_i\mathcal{K}\left(\frac{X_i-x}{h}\right)(X_i-x)\sum_{i=1}^{n}(X_i-x)K\left(\frac{X_i-x}{h}\right)\right)}{\frac{1}{(h_1h_2\cdots h_q)^2}\left(\sum_{i=1}^{n}\mathcal{K}\left(\frac{X_i-x}{h}\right)\sum_{i=1}^{n}(X_i-x)^2\mathcal{K}\left(\frac{X_i-x}{h}\right) - \left(\sum_{i=1}^{n}(X_i-x)K\left(\frac{X_i-x}{h}\right)\right)^2\right)}$$

$$\to_p \sum_{i=1}^{n} \frac{\frac{1}{n}y_i\mathcal{K}\left(\frac{X_i-x}{h}\right)\mathbb{E}\left((X-x)^2\mathcal{K}\left(\frac{X-x}{h}\right)\right) - \frac{1}{n}y_i\mathcal{K}\left(\frac{X_i-x}{h}\right)(X_i-x)\mathrm{E}\left((X-x)K\left(\frac{X-x}{h}\right)\right)}{\mathbb{E}\left(\mathcal{K}\left(\frac{X-x}{h}\right)\right)\mathbb{E}\left((X-x)^2\mathcal{K}\left(\frac{X-x}{h}\right)\right) - \mathbb{E}\left((X-x)K\left(\frac{X-x}{h}\right)\right)\mathbb{E}\left((X-x)K\left(\frac{X-x}{h}\right)\right)}$$

$$\to_p \frac{\mathbb{E}\left(yK\left(\frac{X-x}{h}\right)\right)\mathbb{E}\left((X-x)^2\mathcal{K}\left(\frac{X-x}{h}\right)\right) - \mathbb{E}\left(yK\left(\frac{X-x}{h}\right)(X-x)\right)\mathbb{E}\left((X-x)K\left(\frac{X-x}{h}\right)\right)}{\mathbb{E}\left(\mathcal{K}\left(\frac{X-x}{h}\right)\right)\mathbb{E}\left((X-x)^2\mathcal{K}\left(\frac{X-x}{h}\right)\right) - \mathbb{E}\left((X-x)K\left(\frac{X-x}{h}\right)\right)\mathbb{E}\left((X-x)K\left(\frac{X-x}{h}\right)\right)}.$$

Substitute (2.4) into the above equation, we then get

$$\hat{g}(x)_{LL} \to_p \frac{\mathbb{E}\left(g(x)K\left(\frac{X-x}{h}\right)\right)\mathbb{E}\left((X-x)^2\mathcal{K}\left(\frac{X-x}{h}\right)\right) - \mathbb{E}\left(g(x)K\left(\frac{X-x}{h}\right)(X-x)\right)\mathbb{E}\left((X-x)K\left(\frac{X-x}{h}\right)\right)}{\mathbb{E}\left(\mathcal{K}\left(\frac{X-x}{h}\right)\right)\mathbb{E}\left((X-x)^2\mathcal{K}\left(\frac{X-x}{h}\right)\right) - \mathbb{E}\left((X-x)K\left(\frac{X-x}{h}\right)\right)\mathbb{E}\left((X-x)K\left(\frac{X-x}{h}\right)\right)}$$

$$+ \frac{\mathbb{E}\left(u(\tilde{x},\varepsilon)\mathcal{K}\left(\frac{X-x}{h}\right)\right)\mathbb{E}\left((X-x)^2\mathcal{K}\left(\frac{X-x}{h}\right)\right) - \mathbb{E}\left(u(\tilde{x},\varepsilon)\mathcal{K}\left(\frac{X-x}{h}\right)(X-x)\right)\mathbb{E}\left((X-x)K\left(\frac{X-x}{h}\right)\right)}{\mathbb{E}\left(\mathcal{K}\left(\frac{X-x}{h}\right)\right)\mathbb{E}\left((X-x)^2\mathcal{K}\left(\frac{X-x}{h}\right)\right) - \mathbb{E}\left((X-x)K\left(\frac{X-x}{h}\right)\right)\mathbb{E}\left((X-x)K\left(\frac{X-x}{h}\right)\right)}.$$

To derive the asymptotic bias of the local linear estimator when there exists endogeneity in the model due to omitted variables, we first give the following assumption

ASSUMPTION 3.1. *There exists a symmetric kernel function* $\mathcal{K}(\cdot)$ *and a pdf.* $f(\cdot)$ *such that*

$\int \upsilon^k \mathcal{K}(\upsilon) f(\upsilon) d\upsilon = \int \upsilon^p \mathcal{K}(\upsilon) d\upsilon \cdot \int \upsilon^{k-p} \mathcal{K}(\upsilon) f(\upsilon) d\vartheta$ *holds true for* $k \geq p$ *where* $k,p \in \mathbb{N}^+$ *are all real positive integer numbers.*

This assumption is different from the Assumption 2.4. In this assumption, we impose a departure condition to the kernel function used in the local linear estimator's construction, while we only ask the kernel function in Assumption 2.4 to satisfy symmetry, nonnegative and other usual conditions used in local constant estimations (Li and Racine, 2007). It is obvious that Assumption 3.1 is more severe on the kernels than in Assumption 2.4, and the Assumption 3.1 will only be used in section 3.2.

THEOREM 3.3. *Under Assumptions 2.1-2.3 and Assumption 3.1, as* $n \longrightarrow \infty$, $h_k \longrightarrow 0$ *and* $nh_1h_2 \cdots h_q \longrightarrow \infty$ *for* $k = 1,2,\ldots,q$ *, we have*

$$\hat{g}(x)_{LL} = g(x) + u(\tilde{x}) + \frac{1}{2}\mathcal{K}_2 u_{ss}(\tilde{x}) + \mathcal{O}\left(h_{q_1+1}^2 + h_{q_1+2}^2 + \ldots + h_q^2\right) + o_p(1),$$

where $u_{ss}(\tilde{x}) = \sum_{s \neq s' = q_1+1}^{q} h_s^2 \frac{\partial^2 u(\tilde{x})}{\partial t_s \partial t_{s'}}$, $\mathcal{K}_2 = \int_{-\infty}^{+\infty} \mathcal{K}(\mu)\mu^2 d\mu$.

It is well known that the biases of the Gasser-Müller's estimator and the local linear estimator are all $\sum_{s=1}^{q} 1/2\, \mathcal{K}_2 h_s^2 \frac{f(x) g_{ss}(x)}{f(x)}$, which are smaller than the asymptotic bias of the Nadaraya-Watson's estimator for excluding the term $\sum_{s=1}^{q} 1/2\, \mathcal{K}_2 h_s^2 \frac{f_s(x) g_s(x)}{f(x)}$ when there are no variables omitted and the model (2.1) is correctly specified. Our study finds that the asymptotic bias of the local linear estimator is also the same as the asymptotic bias of the Gasser-Müller's estimator when there are variables omitted from model (2.1), and $\hat{g}(x)_{LL} - g(x) = \mathcal{O}(\hat{g}(x)_{NW} - g(x))$ also holds true for the comparisons between the local linear estimator and the Nadaraya-Watson estimator as shown in Theorem 3.2.

4. LIMIT THEORY UNDER LOCATION SHIFTS

Following Phillips and Liangjun Su (2011), we first introduce a parametric linear and nonparametric regression model respectively where there are continuous location shifts in the regressors whose supports are infinite, and the omitted variables enter the regression error term as shown in Theorem 2.1 and models (2.3), (2.4). The data generating processes are as follows

$$y = x\beta + u(\tilde{x},\varepsilon), \quad \textbf{(4.1)}$$

$$y = g(x) + u(\tilde{x},\varepsilon), \quad \textbf{(4.2)}$$

where $x = (\bar{x},0)$, and

$$\bar{x} = \sum_{\bar{\alpha}=-m}^{m} U_{\bar{\alpha}} \mathbb{I}\{ij \in A_{\bar{\alpha}}\} + U_{\bar{x}}, i = 1,2,\ldots,q_1; j = 1,2,\ldots,n, \quad \textbf{(4.3)}$$

$$\tilde{x} = \sum_{\tilde{\alpha}=-m}^{m} U_{\tilde{\alpha}} \mathbb{I}\{ij \in A_{\tilde{\alpha}}\} + U_{\tilde{x}}, i = q_{1+1},q_{1+2},\ldots,q; j = 1,2,\ldots,n. \quad \textbf{(4.4)}$$

$\{A_\alpha\}$ are disjoints clusters of individuals associated with locations $\{U_\alpha\}$ for the endogenous regressors $\bar{x}$ and the omitted variables $\tilde{x}$, $U_{\bar{x}}$, $U_{\tilde{x}}$ are the errors between the regressors and the location shifts respectively. We allow the included variables and the omitted variables all to have location shifts and the endogenous variables' location shifts tend to be infinite if and only if $\mathrm{n} \longrightarrow \infty$ in this paper.[1]

ASSUMPTION 4.1. *The regression errors* $(\mathrm{u}(\tilde{\mathrm{x}},\varepsilon),\mathrm{U}_{\tilde{\mathrm{x}}})$ *are independently distributed.*

ASSUMPTION 4.2. $\mathbb{E}(\mathrm{U}_{\tilde{\mathrm{x}}}) = 0$, $\mathbb{E}(\mathrm{U}_{\tilde{\mathrm{x}}}^2) = \sigma^2$, and $\mathbb{E}|\mathrm{U}_{\tilde{\mathrm{x}}}|^{2+\theta} < \infty$ for some $\theta > 0$.

ASSUMPTION 4.3. *(a) The probability density function* $\mathrm{f}(\cdot\,,\cdot)$ *of variables* $(\mathrm{u}(\tilde{\mathrm{x}},\varepsilon),\mathrm{U}_{\tilde{\mathrm{x}}})$ *has second-order partial derivative* $\mathrm{f}''(\mathrm{u},\mathrm{U}_{\tilde{\mathrm{x}}})$ *with respect to* $\mathrm{U}_{\tilde{\mathrm{x}}}$ *such that* $\mathrm{f}''(\mathrm{u},\mathrm{U}_{\tilde{\mathrm{x}}})$ *is continuous in* $\mathrm{U}_{\tilde{\mathrm{x}}}$ *and* $\int\int |\mathrm{u}(\tilde{\mathrm{x}},\varepsilon)\mathrm{f}''(\mathrm{u},\mathrm{U}_{\tilde{\mathrm{x}}})|\mathrm{dud}\,\mathrm{U}_{\tilde{\mathrm{x}}} < \infty$. *The p.d.f. of* $\mathrm{U}_{\tilde{\mathrm{x}}}$ *and* $\mathrm{U}_{\tilde{\mathrm{x}}}$*:* $\mathrm{f}_{\mathrm{U}_{\tilde{\mathrm{x}}}}(\cdot)$, $\mathrm{f}_{\mathrm{U}_{\tilde{\mathrm{x}}}}(\cdot)$ *have second-order continuous derivatives. (b) For some* $\xi > 1$ *and any sequence* $\mathbb{P}_{\mathrm{n}}$ *,*$\mathrm{f}_{\tilde{\mathrm{x}}}(\mathbb{P}_{\mathrm{n}}) = \mathcal{O}(|\mathbb{P}_{\mathrm{n}}|^{-\xi-1})$ *as* $|\mathbb{P}_{\mathrm{n}}| \longrightarrow \infty$.

ASSUMPTION 4.3. *As* $\mathrm{n} \longrightarrow \infty$,*we have* $\overline{\mathcal{L}}_{\mathrm{n}} \longrightarrow \infty$*,* $\widetilde{\mathcal{L}}_{\mathrm{n}} \longrightarrow \mathcal{L}$*,* $\mathrm{h} \longrightarrow 0$ *and* $\mathrm{nh}/\overline{\mathcal{L}}_{\mathrm{n}} \longrightarrow \infty$.

As pointed out by Phillips and Liangjun Su (2011), we can relax Assumption 4.1 to be strictly stationary and strong mixing (e.g., $\alpha -$, etc.) as mentioned in Lemma 2.2, that decay to zero at certain rates. Besides, we can get from Assumption 4.3 that $\mathrm{f}_{\tilde{\mathrm{x}}}'(\mathbb{P}_{\mathrm{n}}) = \mathcal{O}(|\mathbb{P}_{\mathrm{n}}|^{-\xi-2})$ and $\mathrm{f}_{\tilde{\mathrm{x}}}''(\mathbb{P}_{\mathrm{n}}) = \mathcal{O}(|\mathbb{P}_{\mathrm{n}}|^{-\xi-3})$, where some smoothness and tail conditions are imposed on the joint and marginal p.d.fs. What should be noticed is that in this paper we consider the situation where the location shifts of the endogenous variables tend to be infinite with the sample size and the location shifts of the omitted variables are finite. We now state the asymptotic result

THEOREM 4.1. *Under Assumptions 2.1-2.4 and Assumptions 4.1-4.3, as* $\mathrm{n} \longrightarrow \infty$ *we have*

(i) $\hat{g}(x) \longrightarrow_p g(x) + o_p(1)$ *;*

(ii) $\hat{\beta}_{LM} \longrightarrow_p \beta + o_p(1)$*; and*

(iii) $\hat{\beta}_{PLM} \longrightarrow_p \beta + o_p(1)$*,*

where $\hat{\beta}_{LM}$ *is the OLS for the parametric linear model (4.1) with the location shifts; and* $\hat{\beta}_{PLM}$ *is the Robinson estimator for the partial linear model (2.10) with location shifts.*

Specially, when the Assumption 4.3(b) does not hold true, we then have

REMARK 4.1. Under Assumptions 2.1-2.4 and Assumptions 4.1-4.3(a), as $\mathrm{n} \longrightarrow \infty$, we have $\hat{g}(x) \longrightarrow_p g(x) + \Phi(\tilde{\mathrm{x}}) + o_p(1)$.

The theorem shows that the parametric, nonparametric and semi-parametric estimations are all consistent when there are variables omitted and there are location shifts in the endogenous included variables. In this situation, we need no instrument variables to gain consistent estimation. This result is in accordance with Phillips and Liangjun Su (2011). When $\mathcal{L}_{\mathrm{n}} \longrightarrow \infty$ and the condition (b) in Assumption 4.3 does not hold, from Theorem 2.3 and Remark 4.1, we find that although the nonparametric estimation is no longer consistent at this time, its asymptotic bias is smaller than the asymptotic bias when there are no location shifts in the regressors because of the fact that $\Omega(\tilde{\mathrm{x}}) \geq 0$. Combining Theorem 4.1 and Remark 4.1, we find that, as for the omitted variables' regressions, the asymptotic bias of the nonparametric estimation when there are location shifts in the endogenous regressors is always much smaller than the one when there are no location shifts.

5. MONTE CARLO SIMULATIONS

In this section, we will illustrate the Monte Carlo simulation results of section 3 and section 4 where the simulation results of section 2 are provided in the online supplementary materials. First of all, we will illustrate the Monte Carlo simulation results for the Nadaraya-Watson's estimator, the

Gasser-Müller's estimator and the local linear estimator where we compare their asymptotic biases and estimation performances for a nonlinear DGP with several variables omitted. The DGP considered is as follows

$$y = \cos(x^4) + z^4 + q^3 + \varepsilon_4, \quad \textbf{(5.1)}$$

where x, z, q are random variables taken from different uniform distributions, and $\varepsilon_4 \sim \mathbb{N}(0,1)$. We use the distance metric defined in section 5.1 to measure the estimation biases of the nonparametric regression methods respectively when the explanatory variables z, q are omitted from (5.1). The results are given in Table 1 and Figure 1.

Table 1. Estimation biases of different nonparametric regression methods.

	$N = 50$	$N = 100$	$N = 150$
Estimators	Bias	Bias	Bias
Local linear	0.04670486	0.07417212	0.10505432
Nadaraya-Watson	0.05918871	0.07899649	0.10695637
Gasser- Müller	0.36729123	0.42263291	0.40266359

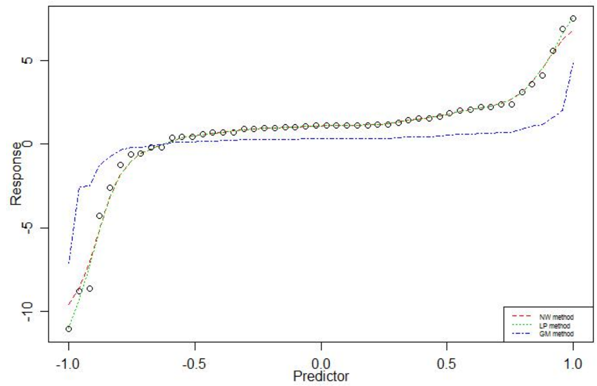


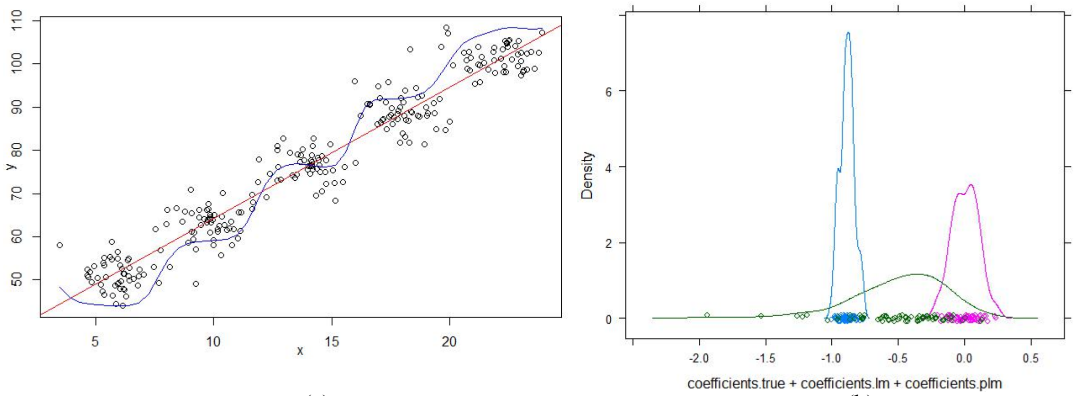
Figure 1. Predictions and fits of different nonparametric regression methods.

As shown in Table 1, the estimation bias of the local linear estimator is smaller than that of the Nadaraya-Watson's estimator, while the estimation bias of the Nadaraya-Watson's estimator is smaller than that of the Gasser- Müller's estimator. The local linear estimation performs the best of all when there is endogeneity in the regression model due to omitted variables. This can be also seen from Figure 1, the local linear method and the Nadaraya-Watson method both fit well the response variable y, while the local linear estimator fits better at the two-tails of the response data than the Nadaraya-Watson's estimator. These simulations support the results of Theorem 3.1, 3.2 and 3.3.

Secondly, we carry out a Monte Carlo study to illustrate the performances of the regressions with location shifts in the regressors. We use the DGP proposed by Phillips and Liangjun Su (2011)

$$y = \xi_i + bx_i + cz_i + da_i + \varepsilon_5, \quad \textbf{(5.2)}$$

where $x_i = \delta_i + \sum_{\alpha=1}^{k} U_\alpha \mathbb{I}\{i \in A_\alpha\} + U_{x_i}$, $z_i = \delta_i + \sum_{\alpha=1}^{k} U_\alpha \mathbb{I}\{i \in A_\alpha\} + U_{z_i}$, $a_i = \delta_i + \sum_{\alpha=1}^{k} U_\alpha \mathbb{I}\{i \in A_\alpha\} + U_{a_i}$, and $\mathrm{Cov}(U_{x_i}, U_{z_i}) = 0.7$, the errors $U = (U_{x_i}, U_{z_i}, U_{a_i})' \sim \mathbb{N}(0, \Sigma)$, $\Sigma = \begin{pmatrix} 1 & 0.5 & 0.7 \\ 0.5 & 1 & 0.7 \\ 0.7 & 0.7 & 1 \end{pmatrix}$. The locations at location α are distributed by errors and the data tend to cluster around each location point. Figure 2(a) illustrates this locational clustering phenomenon with a typical data set for $k = 5$ locations corresponding to $U_\alpha \in \{-4,0,4,8,12\}$ with $\xi_i = 20, b = -1, c = d = 1, \delta_i = 10$ and 100 observations for each α. Along the curve, we see clusters of points around explanatory variable X levels {6, 10, 14, 18, 22}. As the location shifts, the Nadaraya-Watson nonparametric regression better fits the data than the linear regression irrespective of the endogeneity due to omitted variables. Figure 2(b) shows the sampling distributions of the true model's estimator (left), the partial linear model's estimator (median) and the linear model's estimator (right) of the parameter b in (5.2) when there are explanatory variables omitted. It can be seen that approximately $\hat{b}_{\text{true}} \sim \mathbb{N}(-0.8664, 0.0031)$, $\hat{b}_{\text{LM}} \sim \mathbb{N}(-0.3983, 0.0057)$ and $\hat{b}_{\text{PLM}} \sim \mathbb{N}(-0.6398, 0.1291)$, the estimation bias of the partial linear model is smaller than that of the linear model although there are variables omitted from regressions, and all the parametric and nonparametric estimations are consistent with location shifts in the endogenous regressor when the sample size tends to be infinite. Our findings accord with Phillips and Liangjun Su (2011) and support the results of section 2 and Theorem 4.1.



(a) (b)

Figure 2. OLS, Nadaraya-Watson and Robinson's estimator performances under location shifts.
Note: (a) Least-squared parametric regression curve (red line) and the Nadaraya-Watson nonparametric regression curve (blue curve) for the five location shifts with variable z omitted; (b) the estimators' sampling distributions of b(=1) for the correctly specified model (left), the partial linear model (median) and the linear model (right) with location shifts and variable z omitted. N=500.

5. CONCLUDING REMARKS

Comparisons of the performances between parametric and nonparametric methods under different model specifications have been one of the focuses of attention for econometricians. In this paper, we have studied the asymptotic properties of the parametric and nonparametric methods when there are some explanatory variables omitted from regression models. A unified framework to derive, indentify and compare these asymptotic biases is proposed, and as far as we know, this framework has not been considered in the literature yet. We have found the conditions under which the asymptotic bias of the nonparametric method and the asymptotic bias of the semi-parametric method are smaller than that of the parametric method, and it has been shown that the nonparametric approach as well as the semi-parametric approach can dominate the parametric approach in dealing with the endogeneity problem caused by omitting variables if the sample size is large enough and the included endogenous variables are finite. Meanwhile, we have also shown that the asymptotic bias of the local linear estimator, together with the Gasser-Müller's estimator, is smaller than that of the Nadaraya-Watson's estimator when there are some explanatory variables omitted. It is still an open question to decide how to take this proposed framework into account to omitted variables testing (e.g., Fan and Li, 1996, etc.) and model separability testing (e.g., Su et al., 2015, Swofford and Whiney, 1987, 1988, Ait-Sahalia et al., 2001 and Blackorby et al., 2006, etc.). Another interesting question is whether there exist ways for us to improve the estimations and inferences when dealing with the endogeneity problem caused by omitting variables using this framework. Works along these directions are ongoing and anticipated.

NOTE

1. Monte Carlo simulation studies show that the results in section 4 also hold true if there are no location shifts in the omitted variables, and the situation where there are location shifts only in the omitted variables but not in the included variables is just the same as the situation we discussed in section 2.

REFERENCES

Aït-Sahalia, Y., Bickel, P. J. and T. M. Stoker (2001). Goodness-of-fit tests for kernel regression with an application to option implied volatilities. Journal of Econometrics 105, 363-412.

Arellano, M. (2004). Panel data econometrics. Oxford: Oxford University Press.

Audenaert, K. M. R. (2010). Variance bounds, with an application to norm bounds for commutators. Linear Algebra & Its Applications 432, 1126-1143.

Baltagi, B. H. (2013). Econometric analysis of panel data. New York, NY: John Wiley.

Blackorby, C., Primont, D. and R. R. Russell (2006). On testing separability restrictions with flexible functional forms. Journal of Econometrics 5, 195-209.

Cameron A. C. and P. K. Trivedi (2008). Microeconometrics: methods and applications. Cambridge: Cambridge University Press.

Chesher, A. (2003). Identification in nonseparable models. Econometrica 71, 1405-1441.

Dawid, A. P. (1979). Conditional independence in statistical theory. *J.R.Stat.Soc.Ser.B 41*(1), 1-31.

Fan, J. Q. and R. Gijbels (1996). Local polynomial modeling and its applications. London: Chapman & Hall.

Fan, J. and R. Gijbels (1992). Variable bandwidth and local linear regression smoothers. Annals of Statistics 20, 2008-2036.

Fan, J., Gasser, T., Gijbels, I., Brockmann, M. and J. Engel (1997). Local polynomial regression: optimal kernels and asymptotic minimax efficiency. Annals of the Institute of Statistical Mathematics 49, 79-99.

Fan, Y. and Q. Li (1996). Consistent model specification tests: omitted variables and semiparametric functional forms. Econometrica 64, 865-890.

Feldstein, M. S. (1973). Multicollinearity and the mean square error of alternative estimators. Econometrica 41, 337-346.

Florens, J. P., Johannes, J. and S. V. Bellegem (2012). Instrumental regression in partially linear models. Econometrics Journal 15, 304–324.

Gasser, T. and H. G. Müller (1979). Kernel estimation of regression functions. In Gasser. T and Rosenblatt M. (Eds.), Smoothing Techniques for Curve Estimation, 23-68. Berlin: Springer.

Gasser, T. and H. G. Müller (1984). Estimating regression functions and their derivatives by the kernel method. Scandinavian Journal of Statistics, 11, 171-185.

Godfrey, L. G. (1999). Instrument relevance in multivariate linear models. Review of Economics & Statistics 81, 550-552.

Green, W. H. (2012). Econometrics Analysis (7th ed.). New Jersey, NJ: Prentice Hall.

Hall, P., Li, Q. and J. S. Racine (2007). Nonparametric estimation of regression functions in the presence of irrelevant regressors. Review of Economics & Statistics 89, 784-789.

Hausman, J. A. (1983). Chapter 7 Specification and estimation of simultaneous equation models. Handbook of Econometrics, Volume 1, 391-448. Amsterdam: Elsevier B.V.

Hoderlein, S. and H. White (2012). Nonparametric identification in nonseparable panel data models with generalized fixed effects. Journal of Econometrics 168, 300-314.

Imbens, G. W. and W. K. Newey (2009). Identification and estimation of triangular simultaneous equations models without additivity. Econometrica 77, 1481-1512.

Kinal, T. and K. Lahiri (1981). Exact sampling distribution of the omitted variable estimator. Economics Letters 8, 121-127.

Lavergne, P. and Q. Vuong (2000). Nonparametric significance testing. Econometrics Theory 16, 576-601.

Li, Q. and J. S. Racine (2007). Nonparametric econometrics: theory and practice. New Jersey, NJ: Princeton University Press.

Lu, X. and H. White (2014). Testing for separability in structural equations. Journal of Econometrics, 182, 14-26.

Lütkepohl, H. (1982). Non-causality due to omitted variables. Journal of Econometrics 19, 367-378.

Maasoumi, E. and J. Racine (2002). Entropy and predictability of stock market returns. Journal of Econometrics 107, 291-312.

Matzkin, R. L. (2003). Nonparametric estimation of nonadditive random functions. Working paper, Universidad de San Andrés.

Matzkin, R. L. (2007). Nonparametric identification. In J. J. Heckman and E. E. Leamer (Eds.), Handbook of Econometrics, Volume 6, 1157–1167.

Matzkin, R. L. (2008). Identification in nonparametric simultaneous equations models. Econometrica 76, 945-978.

Mccallum, B. T. (1972). Relative asymptotic bias from errors of omission and measurement. Econometrica 40, 757-758.

Muilwijk, J. (1966). Note on a theorem of M. N. Murthy and V. K. Sethi. Sankhyā Ser B 28, 183-183.

Murthy, M. N. and V. K. Sethi (1965). Self-welighting design at tabulation stage. Sankhyā Ser A 27, 201-210.

Nadaraya, E. A. (1989). Nonparametric estimation of probability densities and regression curves. Dordrecht: Kluwer Academic Publishers.

Pearl, J. (2000). Causality: models, reasoning, and inference. Cambridge: Cambridge University Press.

Phillips, P. C. B. and L. J. Su (2009). Nonparametric structural estimation via continuous location shifts in an endogenous regressor. Cowles Foundation Discussion Paper 1702, Yale University.

Phillips, P. C. B. and L. J. Su (2011). Nonparametric structural estimation under location shifts. Econometrics Journal 14, 457-486.

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

Sessions, D. N. and L. K. Stevans (2006). Investigating omitted variable bias in regression parameter estimation: a genetic algorithm approach. Computational Statistics & Data Analysis 50, 2835-2854.

Stafford, J. E. and D. F. Andrews (1993). A geometric approach to assess bias due to omitted covariates in generalized linear models. Biometrika 80, 807-815.

Su, L., Y. Tu and A. Ullah (2015). Testing Additive Separability of Error Term in Nonparametric Structural Models. Econometric Reviews 34, 1056-1087.

Swofford, J. L. and G. A. Whitney (1987). Nonparametric tests of utility maximization and weak separability for consumption, leisure and money. Review of Economics & Statistics 69, 458-464.

Swofford, J. L. and G. A. Whitney (1988). A comparison of nonparametric tests of weak separability for annual and quarterly data on consumption, leisure, and money. Journal of Business & Economic Statistics 6, 241-246.

Toro-Vizcarrondo, C. and T. D. Wallace (2014). A test of the mean square error criterion for restrictions in linear regression. Publications of the American Statistical Association 63, 558-572.

Triacca, U. (1998). Non-causality: the role of the omitted variables. Economics Letters 60, 317-320.

Wallace, T. D. and C. E. Toro-Vizcarrondo (1969). Tables for the mean square error test for exact linear restrictions in regression. Publications of the American Statistical Association 64, 1649-1663.

Wang, Q. and P. C. B. Phillips (2009a). Asymptotic theory for local time density estimation and nonparametric cointegrating regression. Econometric Theory 25, 710-738.

Wang, Q. and P. C. B. Phillips (2009b). Structural nonparametric cointegrating regression. Econometrica 77, 6, 1901-1948.

Wang, Q. and P. C. B. Phillips (2011). Asymptotic theory for zero energy functionals with nonparametric regression applications. Econometric Theory 27, 235-259.

Wooldrige, J. M. (2010). Econometric analysis of cross section and panel data. Massachusetts, MA: MIT Press.

APPENDIX A: Proofs

In the appendix A, we give out the proofs of the results in section 2, and the proofs of section 3 and section 4 are provided in the online appendix.

Proof of Theorem 2.1: We divide this theorem into two parts: the "if" part and the "only if" part. (i) For the "if" part, using (a) $(\bar{x},\tilde{x})$ and ε are separable under model (2.1), we can get $y = g(\bar{x},\tilde{x},\varepsilon) = g_1(\bar{x},\tilde{x}) + g_2(\varepsilon)$; using (b) $\bar{x}$, $\tilde{x}$ are separable, we have $y = g_1(\bar{x},\tilde{x}) + g_2(\varepsilon) = g_{11}(\bar{x}) + g_{12}(\tilde{x}) + g_2(\varepsilon)$, where $g_{11}(\cdot)$,$g_{12}(\cdot)$ and $g_2(\cdot)$ are all continuous functions and hence after; by (a) again, we then get $y = g_{11}(\bar{x}) + g_3(\tilde{x},\varepsilon)$. (ii) For the "only if" part, we set $v = g_3(\tilde{x},\varepsilon) - \mathbb{E}(g_3(\tilde{x},\varepsilon)|\bar{x},\tilde{x})$, then from Assumption 2.3, $\tilde{x} \perp \varepsilon|\bar{x}$ implies $\tilde{x} \perp (\varepsilon,\bar{x})|\bar{x}$ by Dawid (1979, Lemma 4.1), and $\tilde{x} \perp (\varepsilon,\bar{x})|\bar{x}$ implies $\tilde{x} \perp v|\bar{x}$ by Dawid (1979, Lemma 4.2). So by Assumption 2.2 and Proposition 2, 3 of Lu and White (2014), we have $\tilde{x}$ and ε are separable, then $g_3(\tilde{x},\varepsilon) = g_{31}(\tilde{x}) + g_{32}(\varepsilon)$. Notice that, by Proposition 1 of Lu and White (2014), $g_3(\tilde{x},\varepsilon) + g_4(\bar{x}) = g_{31}(\tilde{x}) + g_{32}(\varepsilon) + g_4(\bar{x}) = g(\bar{x},\tilde{x},\varepsilon)$ holds true if and only if $\bar{x}$, ε are separable while $\bar{x}$, $\tilde{x}$ are separable. □

Proof of Theorem 2.2: The proof is quite similar to Matzkin (2004, 2007). Let $\bar{x}, \tilde{x}, e_1, e$ denote the value of $\bar{X}$, $\tilde{X}$, ε_1, ε respecively. By conditional independence and strict monotonicity

$$\mathbb{P}\{\bar{X} \le \bar{x}|\tilde{X} = \tilde{x}\} = \mathbb{P}\{\bar{X} \le \bar{x}|\tilde{X} = \tilde{x},\varepsilon = e\} \tag{A.1}$$

implies

$$\begin{aligned}&\mathbb{P}\{\varepsilon \le e|\tilde{X} = \tilde{x}\} = \mathbb{P}\{\varepsilon \le e|\bar{X} = \bar{x},\tilde{X} = \tilde{x}\}\\ &= \mathbb{P}\{g_2(\tilde{X},\varepsilon) \le g_2(\tilde{x},e)|\bar{X} = \bar{x},\tilde{X} = \tilde{x}\}\\ &= \mathbb{P}\{g_1(\bar{X},\varepsilon_1) \le g_1(\bar{x},e_1)|\bar{X} = \bar{x},\tilde{X} = \tilde{x}\}\\ &= \mathcal{F}_{y|\bar{X} = \bar{x},\tilde{X} = \tilde{x}}(g_1(\bar{x},e_1)),\end{aligned}$$

hence

$$g_1(\bar{x},e_1) = \mathcal{F}^{-1}_{y|\bar{X} = \bar{x},\tilde{X} = \tilde{x}}\left(\mathcal{F}_{e|\tilde{X} = \tilde{x}}(e)\right).$$

(A.2)

since

$$\mathcal{F}_{\varepsilon_1|\bar{X} = \bar{x}} = \mathcal{F}_{y|\bar{X} = \bar{x}}(g_1(\bar{x},e_1)),$$

(A.3)

it then follows that

$$\mathcal{F}_{\varepsilon_1|\bar{X} = \bar{x}} = \mathcal{F}_{y|\bar{X} = \bar{x}}\left(\mathcal{F}^{-1}_{y|\bar{X} = \bar{x},\tilde{X} = \tilde{x}}\left(\mathcal{F}_{\varepsilon|\tilde{X} = \tilde{x}}(e)\right)\right).$$

□

Proof of Theorem 2.3: (i) We first prove part one of this theorem: when there are variables omitted, we can carry out a direct decomposition to the kernel function used in the nonparametric estimator (2.7) in section 2.3: $\mathcal{K}\left(\frac{X_i-x}{h}\right) = \prod_{i=1}^{q_1}\mathcal{K}\left(\frac{X_i-x}{h}\right)\prod_{i=q_1+1}^{q}\mathcal{K}\left(\frac{X_i-x}{h}\right) = \mathcal{K}\left(\frac{\bar{X}_i-x}{h}\right)\mathcal{K}\left(\frac{\tilde{X}_i-x}{h}\right)$, where we divide the kernel function into two parts: the included variables' part and the omitted variables' part. Based on this decomposition, we are able to prove this theorem by Taylor expansions. First of all, as for the nonparametric regression model (2.4) where some explanatory variables are omitted, substitute (2.4) into (2.7), we then get

$$\hat{g}(x)\longrightarrow_p \frac{\mathbb{E}\left(g(x)\mathcal{K}\left(\frac{X-x}{h}\right)\right)+\mathbb{E}\left(u(\tilde{x})\mathcal{K}\left(\frac{X-x}{h}\right)\right)}{\mathbb{E}\left(\mathcal{K}\left(\frac{X-x}{h}\right)\right)}, \qquad (\mathbf{A.4})$$

in which we set $u(\tilde{x}) \triangleq u(\tilde{x},\varepsilon)$ for simple, because we only carry a Taylor expansion to $\tilde{x}$ in this proof, similarly we set $\frac{\partial^{\upsilon}u(\tilde{x})}{\partial\tilde{x}^{\upsilon}} \triangleq \frac{\partial^{\upsilon}u(\tilde{x},\varepsilon)}{\partial\tilde{x}^{\upsilon}}, \upsilon = 1,2,...$ for υ th deriviate. One of the reasons why some explanatory variables are omitted from the regressions in empirical studies may be that the researchers mistakenly think that the variables $\bar{x}, \tilde{x}$ are uncorrelated and independent of each other, which means $f(\bar{x},\tilde{x}) = f(\bar{x})f(\tilde{x})$, where $f(\cdot)$ is the p.d.f. Carrying out a Taylor expansion to the numerator part of (A.4), as $n \longrightarrow \infty$, $h_s \longrightarrow 0$, $s = 1,2,...,q$, we can get

$$\mathbb{E}\left(g(X)K\left(\frac{X-x}{h}\right)\right) = \mathbb{E}\left(g(X)K\left(\frac{\bar{X}-x}{h}\right)\mathcal{K}\left(\frac{\tilde{X}-x}{h}\right)\right)$$

$$= \int\int\left(g(\bar{X},\tilde{X})\mathcal{K}\left(\frac{\bar{X}-x}{h}\right)\mathcal{K}\left(\frac{\tilde{X}-x}{h}\right)f(\bar{X},\tilde{X})\right)d\bar{X}d\tilde{X}$$

$$= \int\int (g(\bar{x}+h\bar{\upsilon},\tilde{x}+h\tilde{\upsilon})f(\bar{x}+h\bar{\upsilon},\tilde{x}+h\tilde{\upsilon})\mathcal{K}(\bar{\upsilon})\mathcal{K}(\tilde{\upsilon}))\,d\bar{\upsilon}d\tilde{\upsilon}$$

$$= \int\int\left[g(x) + \bar{\upsilon}\sum_{s=1}^{q_1}\frac{\partial g(x)}{\partial x_s}h_s + \tilde{\upsilon}\sum_{q_1+1}^{q}\frac{\partial g(x)}{\partial x_s}h_s + \frac{\bar{\upsilon}^2}{2}\sum_{s\neq s'=1}^{q_1}\frac{\partial^2 g(x)}{\partial x_s\partial x_{s'}}h_s^2 + \frac{\tilde{\upsilon}^2}{2}\sum_{s\neq s'=q_1}^{q}\frac{\partial^2 g(x)}{\partial x_s\partial x_{s'}}h_s^2 + \mathcal{O}(\sum_{s=1}^{q_1}h_s^2\right.$$

$$\left.+\sum_{s\neq s'=q_1}^{q}h_s^2)\right]\left[f(\bar{x})\bar{\upsilon}\sum_{s=1}^{q_1}\frac{\partial f(\bar{x})}{\partial x_s}h_s + \frac{\bar{\upsilon}^2}{2}\sum_{s\neq s'=1}^{q_1}\frac{\partial^2 f(\bar{x})}{\partial x_s\partial x_{s'}}h_s^2\right.$$

$$\left.+\mathcal{O}(\sum_{s=1}^{q_1}h_s^2)\right]\left[f(\tilde{x})\tilde{\upsilon}\sum_{s=q_1+1}^{q}\frac{\partial f(\tilde{x})}{\partial x_s}h_s + \frac{\tilde{\upsilon}^2}{2}\sum_{s\neq s'=q_1+1}^{q}\frac{\partial^2 f(\tilde{x})}{\partial x_s\partial x_{s'}}h_s^2\right.$$

$$\left.+\mathcal{O}(\sum_{s=q_1+1}^{q}h_s^2)\right]\mathcal{K}(\bar{\upsilon})\mathcal{K}(\tilde{\upsilon})d\bar{\upsilon}d\tilde{\upsilon}$$

$$= g(x)f(\bar{x})f(\tilde{x}) + g(x)f(\bar{x})\frac{\tilde{\mathcal{K}}_2}{2}\sum_{s\neq s'=q_1+1}^{q}\frac{\partial^2 f(\tilde{x})}{\partial x_s\partial x_{s'}}h_s^2 + g(x)f(\bar{x})\mathcal{O}\left(\sum_{s=q_1+1}^{q}h_s^2\right)$$

$$+ f(\bar{x})\tilde{\mathcal{K}}_2\sum_{s=q_1+1}^{q}\frac{\partial f(\tilde{x})}{\partial x_s}h_s\sum_{q_1+1}^{q}\frac{\partial g(x)}{\partial x_s}h_s + f(\bar{x})\frac{\tilde{\mathcal{K}}_3}{2}\sum_{q_1+1}^{q}\frac{\partial g(x)}{\partial x_s}h_s\sum_{s\neq s'=q_1+1}^{q}\frac{\partial^2 f(\tilde{x})}{\partial x_s\partial x_{s'}}h_s^2$$

$$+ f(\bar{x})f(\tilde{x})\frac{\tilde{\mathcal{K}}_2}{2}\sum_{s\neq s'=1}^{q_1}\frac{\partial^2 g(x)}{\partial x_s\partial x_{s'}}h_s^2 + f(\bar{x})\frac{\tilde{\mathcal{K}}_3}{2}\sum_{s\neq s'=1}^{q_1}\frac{\partial^2 g(x)}{\partial x_s\partial x_{s'}}h_s^2\sum_{s=q_1+1}^{q}\frac{\partial f(\tilde{x})}{\partial x_s}h_s$$

$$+ f(\bar{x})\frac{\tilde{\mathcal{K}}_4}{4}\sum_{s\neq s'=1}^{q_1}\frac{\partial^2 g(x)}{\partial x_s\partial x_{s'}}h_s^2\sum_{s\neq s'=q_1+1}^{q}\frac{\partial^2 f(\tilde{x})}{\partial x_s\partial x_{s'}}h_s^2 + f(\bar{x})\frac{\tilde{\mathcal{K}}_2}{2}\sum_{s\neq s'=1}^{q_1}\frac{\partial^2 g(x)}{\partial x_s\partial x_{s'}}h_s^2\,\mathcal{O}\left(\sum_{s=q_1+1}^{q}h_s^2\right)$$

$$+ f(\bar{x})f(\tilde{x})\mathcal{O}\left(\sum_{s=q_1+1}^{q}h_s^2\right) + f(\bar{x})\frac{\tilde{\mathcal{K}}_2}{2}\sum_{s\neq s'=q_1+1}^{q}\frac{\partial^2 f(\tilde{x})}{\partial x_s\partial x_{s'}}h_s^2\,\mathcal{O}\left(\sum_{s=q_1+1}^{q}h_s^2\right) + f(\bar{x})\mathcal{O}\left(2\sum_{s=q_1+1}^{q}h_s^2\right),$$

where $\widetilde{\mathcal{K}}_j = \int \tilde{\mathcal{v}}^j \mathcal{K}(\tilde{\mathcal{v}}) d\tilde{\mathcal{v}}$ and $\bar{\mathcal{K}}_j = \int \bar{\mathcal{v}}^j \mathcal{K}(\bar{\mathcal{v}}) d\bar{\mathcal{v}}$ for j = 1,2,... Similarly, it can be shown that

$$\mathbb{E}\left(u(\tilde{x})\mathcal{K}\left(\frac{X-x}{h}\right)\right) = \mathbb{E}\left(u(\tilde{x})\mathcal{K}\left(\frac{\bar{X}-x}{h}\right)\mathcal{K}\left(\frac{\tilde{X}-x}{h}\right)\right)$$

$$= \int\int\left(u(\tilde{x})\mathcal{K}\left(\frac{\bar{X}-x}{h}\right)\mathcal{K}\left(\frac{\tilde{X}-x}{h}\right)f(\bar{X},\tilde{X})\right)d\bar{X}d\,\tilde{X}$$

$$= u(\tilde{x})f(\bar{x})f(\tilde{x}) + u(\tilde{x})f(\bar{x})\frac{\widetilde{\mathcal{K}}_2}{2}\sum_{s\neq s'=q_1+1}^{q}\frac{\partial^2 f(\tilde{x})}{\partial x_s \partial x_{s'}}h_s^2 + u(\tilde{x})f(\bar{x})\mathcal{O}\left(\sum_{s=q_1+1}^{q}h_s^2\right)$$

$$+ f(\bar{x})\widetilde{\mathcal{K}}_2\sum_{s=q_1+1}^{q}\frac{\partial f(\tilde{x})}{\partial x_s}h_s\sum_{q_1+1}^{q}\frac{\partial u(\tilde{x})}{\partial x_s}h_s + f(\bar{x})\frac{\widetilde{\mathcal{K}}_3}{2}\sum_{q_1+1}^{q}\frac{\partial u(\tilde{x})}{\partial x_s}h_s\sum_{s\neq s'=q_1+1}^{q}\frac{\partial^2 f(\tilde{x})}{\partial x_s \partial x_{s'}}h_s^2$$

$$+ f(\bar{x})f(\tilde{x})\frac{\widetilde{\mathcal{K}}_2}{2}\sum_{s\neq s'=1}^{q_1}\frac{\partial^2 u(\tilde{x})}{\partial x_s \partial x_{s'}}h_s^2 + f(\bar{x})\frac{\widetilde{\mathcal{K}}_3}{2}\sum_{s\neq s'=1}^{q_1}\frac{\partial^2 u(\tilde{x})}{\partial x_s \partial x_{s'}}h_s^2\sum_{s=q_1+1}^{q}\frac{\partial f(\tilde{x})}{\partial x_s}h_s$$

$$+ f(\bar{x})\frac{\widetilde{\mathcal{K}}_4}{4}\sum_{s\neq s'=1}^{q_1}\frac{\partial^2 u(\tilde{x})}{\partial x_s \partial x_{s'}}h_s^2\sum_{s\neq s'=q_1+1}^{q}\frac{\partial^2 f(\tilde{x})}{\partial x_s \partial x_{s'}}h_s^2 + f(\bar{x})\frac{\widetilde{\mathcal{K}}_2}{2}\sum_{s\neq s'=1}^{q_1}\frac{\partial^2 u(\tilde{x})}{\partial x_s \partial x_{s'}}h_s^2\,\mathcal{O}\left(\sum_{s=q_1+1}^{q}h_s^2\right)$$

$$+ f(\bar{x})f(\tilde{x})\mathcal{O}\left(\sum_{s=q_1+1}^{q}h_s^2\right) + f(\bar{x})\frac{\widetilde{\mathcal{K}}_2}{2}\sum_{s\neq s'=q_1+1}^{q}\frac{\partial^2 f(\tilde{x})}{\partial x_s \partial x_{s'}}h_s^2\,\mathcal{O}\left(\sum_{s=q_1+1}^{q}h_s^2\right)$$

$$+ f(\bar{x})\mathcal{O}\left(2\sum_{s=q_1+1}^{q}h_s^2\right).$$

Carrying out a Taylor expansion to the denominator part of (A.4), we then get

$$\mathbb{E}\left(\mathcal{K}\left(\frac{X-x}{h}\right)\right) = \mathbb{E}\left(\mathcal{K}\left(\frac{\bar{X}-x}{h}\right)\mathcal{K}\left(\frac{\tilde{X}-x}{h}\right)\right)$$

$$= \int\int\mathcal{K}\left(\frac{\bar{X}-x}{h}\right)\mathcal{K}\left(\frac{\tilde{X}-x}{h}\right)f(\bar{X},\tilde{X})d\,\bar{X}d\tilde{X}$$

$$= f(\bar{x})f(\tilde{x}) + f(\bar{x})\frac{\widetilde{\mathcal{K}}_2}{2}\sum_{s\neq s'=q_1+1}^{q}\frac{\partial^2 f(\tilde{x})}{\partial x_s \partial x_{s'}}h_s^2 + f(\bar{x})\mathcal{O}\left(\sum_{s=q_1+1}^{q}h_s^2\right).$$

Substitute the above three asymptotic expansions into formula (A.4), we finally get the exact asymptotic bias expression for the nonparametric regression model (2.4) when there are several explanatory variables omitted

$$\hat{g}(x) \longrightarrow_p g(x) + \Omega(\tilde{x}) + \Phi(\tilde{x}), \qquad (\mathbf{A.5})$$

where

$$\Omega(\tilde{x}) \triangleq \Bigg(\widetilde{\mathcal{K}}_2\sum_{s=q_1+1}^{q}\frac{\partial f(\tilde{x})}{\partial x_s}h_s\sum_{q_1+1}^{q}\frac{\partial g(x)}{\partial x_s}h_s + \frac{\widetilde{\mathcal{K}}_3}{2}\sum_{q_1+1}^{q}\frac{\partial g(x)}{\partial x_s}h_s\sum_{s\neq s'=q_1+1}^{q}\frac{\partial^2 f(\tilde{x})}{\partial x_s \partial x_{s'}}h_s^2 + f(\tilde{x})\frac{\widetilde{\mathcal{K}}_2}{2}\sum_{s\neq s'=1}^{q_1}\frac{\partial^2 g(x)}{\partial x_s \partial x_{s'}}h_s^2$$

$$+ \frac{\widetilde{\mathcal{K}}_3}{2}\sum_{s\neq s'=1}^{q_1}\frac{\partial^2 g(x)}{\partial x_s \partial x_{s'}}h_s^2\sum_{s=q_1+1}^{q}\frac{\partial f(\tilde{x})}{\partial x_s}h_s + \frac{\widetilde{\mathcal{K}}_4}{4}\sum_{s\neq s'=1}^{q_1}\frac{\partial^2 g(x)}{\partial x_s \partial x_{s'}}h_s^2\sum_{s\neq s'=q_1+1}^{q}\frac{\partial^2 f(\tilde{x})}{\partial x_s \partial x_{s'}}h_s^2$$

$$+ \frac{\widetilde{\mathcal{K}}_2}{2}\sum_{s\neq s'=1}^{q_1}\frac{\partial^2 g(x)}{\partial x_s \partial x_{s'}}h_s^2\,\mathcal{O}\left(\sum_{s=q_1+1}^{q}h_s^2\right) + f(\tilde{x})\mathcal{O}\left(\sum_{s=q_1+1}^{q}h_s^2\right) +$$

$$\frac{\widetilde{\mathcal{K}}_2}{2}\sum_{s\neq s'=q_1+1}^{q}\frac{\partial^2 f(\tilde{x})}{\partial x_s\partial x_{s'}}h_s^2\,\mathcal{O}\left(\sum_{s=q_1+1}^{q}h_s^2\right)+\mathcal{O}\left(2\sum_{s=q_1+1}^{q}h_s^2\right)\Bigg)$$

$$\Bigg/\left(f(\tilde{x})+\frac{\widetilde{\mathcal{K}}_2}{2}\sum_{s\neq s'=q_1+1}^{q}\frac{\partial^2 f(\tilde{x})}{\partial x_s\partial x_{s'}}h_s^2+\mathcal{O}\left(\sum_{s=q_1+1}^{q}h_s^2\right)\right);$$

and

$$\Phi(\tilde{x})\triangleq\Bigg(u(\tilde{x})f(\tilde{x})+u(\tilde{x})\frac{\widetilde{\mathcal{K}}_2}{2}\sum_{s\neq s'=q_1+1}^{q}\frac{\partial^2 f(\tilde{x})}{\partial x_s\partial x_{s'}}h_s^2+u(\tilde{x})\mathcal{O}\left(\sum_{s=q_1+1}^{q}h_s^2\right)+\widetilde{\mathcal{K}}_2\sum_{s=q_1+1}^{q}\frac{\partial f(\tilde{x})}{\partial x_s}h_s\sum_{q_{1+1}}^{q}\frac{\partial u(\tilde{x})}{\partial x_s}h_s$$

$$+\frac{\widetilde{\mathcal{K}}_3}{2}\sum_{q_{1+1}}^{q}\frac{\partial u(\tilde{x})}{\partial x_s}h_s\sum_{s\neq s'=q_1+1}^{q}\frac{\partial^2 f(\tilde{x})}{\partial x_s\partial x_{s'}}h_s^2+f(\tilde{x})\frac{\widetilde{\mathcal{K}}_2}{2}\sum_{s\neq s'=1}^{q_1}\frac{\partial^2 u(\tilde{x})}{\partial x_s\partial x_{s'}}h_s^2$$

$$+\frac{\widetilde{\mathcal{K}}_3}{2}\sum_{s\neq s'=1}^{q_1}\frac{\partial^2 u(\tilde{x})}{\partial x_s\partial x_{s'}}h_s^2\sum_{s=q_1+1}^{q}\frac{\partial f(\tilde{x})}{\partial x_s}h_s+\frac{\widetilde{\mathcal{K}}_4}{4}\sum_{s\neq s'=1}^{q_1}\frac{\partial^2 u(\tilde{x})}{\partial x_s\partial x_{s'}}h_s^2\sum_{s\neq s'=q_1+1}^{q}\frac{\partial^2 f(\tilde{x})}{\partial x_s\partial x_{s'}}h_s^2$$

$$+\frac{\widetilde{\mathcal{K}}_2}{2}\sum_{s\neq s'=1}^{q_1}\frac{\partial^2 u(\tilde{x})}{\partial x_s\partial x_{s'}}h_s^2\,\mathcal{O}\left(\sum_{s=q_1+1}^{q}h_s^2\right)+f(\tilde{x})\mathcal{O}\left(\sum_{s=q_1+1}^{q}h_s^2\right)$$

$$+\frac{\widetilde{\mathcal{K}}_2}{2}\sum_{s\neq s'=q_1+1}^{q}\frac{\partial^2 f(\tilde{x})}{\partial x_s\partial x_{s'}}h_s^2\,\mathcal{O}\left(\sum_{s=q_1+1}^{q}h_s^2\right)+\mathcal{O}\left(2\sum_{s=q_1+1}^{q}h_s^2\right)\Bigg)$$

$$\Bigg/\left(f(\tilde{x})+\frac{\widetilde{\mathcal{K}}_2}{2}\sum_{s\neq s'=q_1+1}^{q}\frac{\partial^2 f(\tilde{x})}{\partial x_s\partial x_{s'}}h_s^2+\mathcal{O}\left(\sum_{s=q_1+1}^{q}h_s^2\right)\right).$$

If there are no variables omitted, all the variables $\bar{x},\tilde{x}$ enter the regression model, from Assumption 2.1 we can get

$$\Omega(\tilde{x})=\frac{0}{f(\tilde{x})+0+0}=0; \qquad (\mathbf{A.6})$$

$$\Phi(\tilde{x})=\frac{u(\tilde{x})f(\tilde{x})+0}{f(\tilde{x})+0}=u(\tilde{x})=u(\emptyset)=0, \qquad (\mathbf{A.7})$$

hence $\hat{g}(x)\longrightarrow_p g(x)+\Omega(\tilde{x})+\Phi(\tilde{x})=g(x)$, the nonparametric estimator is consistent.

(ii) We then prove part two of this theorem: as for the linear regression model (2.3), when there are variables omitted and as $n\longrightarrow\infty$, we can get

$$\begin{aligned}\hat{\beta}&=\left(x'x\right)^{-1}x'y\\&=\left(x'x\right)^{-1}x'(x\beta+u(\tilde{x},\varepsilon))\\&=\beta+\left(\frac{x'x}{n}\right)^{-1}\left(\frac{x'u(\tilde{x},\varepsilon)}{n}\right)\\&\longrightarrow_p\beta+\begin{pmatrix}\mathcal{C}_{11}&\cdots&\mathcal{C}_{1q}\\\vdots&\ddots&\vdots\\\mathcal{C}_{n1}&\cdots&\mathcal{C}_{nq}\end{pmatrix}\begin{pmatrix}\frac{1}{n}(x_{11}u(\tilde{x},\varepsilon)+\ldots+x_{1n}u(\tilde{x},\varepsilon))\\\vdots\\\frac{1}{n}\left(x_{q1}u(\tilde{x},\varepsilon)+\ldots+x_{qn}u(\tilde{x},\varepsilon)\right)\end{pmatrix}\end{aligned}$$

$$\longrightarrow_p \beta + \begin{pmatrix} \mathcal{C}_{11} & \cdots & \mathcal{C}_{1q} \\ \vdots & \ddots & \vdots \\ \mathcal{C}_{n1} & \cdots & \mathcal{C}_{nq} \end{pmatrix} \begin{pmatrix} \mathbb{E}(x_1 u(\tilde{x},\varepsilon)) \\ \vdots \\ \mathbb{E}(x_q u(\tilde{x},\varepsilon)) \end{pmatrix}$$

$$= \beta + \begin{pmatrix} \mathcal{C}_{11}\mathbb{E}(x_1 u(\tilde{x},\varepsilon)) + \ldots + \mathcal{C}_{1q}\mathbb{E}(x_q u(\tilde{x},\varepsilon)) \\ \vdots \\ \mathcal{C}_{n1}\mathbb{E}(x_1 u(\tilde{x},\varepsilon)) + \ldots + \mathcal{C}_{nq}\mathbb{E}(x_q u(\tilde{x},\varepsilon)) \end{pmatrix}.$$

From the proofs of part one, we then have

$$\mathbb{E}(x_i u(\tilde{x})) = \mathbb{E}\left(x_i \frac{\sum_{j=q_1+1}^{q} u_j(\tilde{x})\mathcal{K}\left(\frac{\tilde{x}_j - x}{h}\right)}{\sum_{j=q_1+1}^{q} \mathcal{K}\left(\frac{\tilde{x}_j - x}{h}\right)} \right)$$

$$\longrightarrow_p \mathbb{E}\left(x_i \frac{\mathbb{E}\left(u(\tilde{x})\mathcal{K}\left(\frac{\tilde{x}-x}{h}\right)\right)}{\mathbb{E}\left(\mathcal{K}\left(\frac{\tilde{x}-x}{h}\right)\right)} \right)$$

$$\longrightarrow_p \mathbb{E}(x_i \Phi(\tilde{x}))$$

for $i = 1,2,\ldots,q_1$. Substitute the above equation into the former one, we finally get

$$\hat{\beta} \longrightarrow_p \beta + \mathrm{CE}(x\Phi(\tilde{x})),$$

where $x = (\bar{x},0)$. □

Proof of Theorem 2.4: From Theorem 2.1 and Assumptions 2.1-2.2, we know that $\Phi\colon \tilde{x} \longrightarrow \mathfrak{R}$ is a strict continuous and monotone increasing function, and $\Phi(\tilde{x}) > 0$ because $u(\tilde{x}) > 0$ as long as there are variables omitted ($\tilde{x} \neq \emptyset$). We can then get $\Phi(\tilde{x})x_i \geq 0$ for $i = 1,2,\ldots,q_1$ directly from Assumption 2.5. Set $A_i \subset \mathfrak{R}^{q-q_1+1}$ as a subset of $x_i \cap \tilde{x}$, where $\sigma(A_i) = \sigma(x_i) \cap \sigma(\tilde{x})$ for $i = 1,2,\ldots,q_1$. If $0 \leq \Omega(\tilde{x}) < 1$, by the Markov inequality,

$$\begin{aligned}
\mathbb{E}(x_i\Phi(\tilde{x})) &\geq \mathbb{E}(x_i\Phi(\tilde{x})\Omega(\tilde{x})) \\
&= \int\int x_i\Phi(\tilde{x})\Omega(\tilde{x})f(x_i,\tilde{x})dx_i d\tilde{x} \\
&\geq \int\int_{A_i} x_i\Phi(\tilde{x})\Omega(\tilde{x})f(x_i,\tilde{x})d x_i d\tilde{x} \\
&\geq \inf_{A_i} x_i\Phi(\tilde{x})\Omega(\tilde{x}) \int\int_{A_i} f(x_i,\tilde{x})d x_i d\tilde{x} \\
&\geq \inf_{A_i} \Phi(\tilde{x})\Omega(\tilde{x})\mathcal{F}_{A_i}(x_i,\tilde{x})
\end{aligned}$$

holds true, and $\mathcal{F}_{A_i}(x_i,\tilde{x})$ is the joint c.d.f. of the random variables x_i and $\tilde{x}$, $i = 1,2,\ldots,q_1$. From Theorem 2.3, there exists at least one set $\mathcal{A} = A_1 \cap A_2 \cap \ldots \cap A_q$ such that

$$\begin{aligned}
\hat{\beta}_i - \beta_i &= \mathcal{C}_{i1}\mathbb{E}(x_1\Phi(\tilde{x})) + \mathcal{C}_{i2}\mathbb{E}(x_2\Phi(\tilde{x})) + \ldots + \mathcal{C}_{iq}\mathbb{E}(x_q\Phi(\tilde{x})) \\
&\geq \inf_{A_i}\Phi(\tilde{x})\Omega(\tilde{x})\left(\mathcal{C}_{i1}\mathcal{F}_{\mathcal{A}}(x_1,\tilde{x}) + \mathcal{C}_{i2}\mathcal{F}_{\mathcal{A}}(x_2,\tilde{x}) + \ldots + \mathcal{C}_{iq}\mathcal{F}_{\mathcal{A}}(x_q,\tilde{x})\right) \\
&\geq \sup_{\mathcal{A}}(\Phi(\tilde{x}) + \Omega(\tilde{x})) \\
&= \sup_{\mathcal{A}}(\hat{g}(x) - g(x))
\end{aligned}$$

for $i = 1,2,\ldots,q_1$. If $\Omega(\tilde{x}) \geq 1$, by the Markov inequality,

$$\begin{aligned}
\mathbb{E}(x_i\Phi(\tilde{x})\Omega(\tilde{x})) &\geq \mathbb{E}(x_i\Phi(\tilde{x})) \\
&= \int\int x_i\Phi(\tilde{x})f(x_i,\tilde{x})dx_i d\tilde{x} \\
&\geq \int\int_{A_i} x_i\Phi(\tilde{x})f(x_i,\tilde{x})d x_i d\tilde{x} \\
&\geq \inf_{A_i} x_i\Phi(\tilde{x}) \int\int_{A_i} f(x_i,\tilde{x})d x_i d\tilde{x}
\end{aligned}$$

$$\geq \inf_{A_i} \Phi(\tilde{x}) \mathcal{F}_{A_i}(x_i, \tilde{x})$$

holds true for $i = 1,2,\ldots,q_1$ and where A_i is an intersection set of the variables $x_i, \tilde{x}$. Similarly, there exists at least one set $\mathcal{A}' = A_1 \cap A_2 \cap \ldots \cap A_q$ such that

$$\begin{aligned}
\hat{\beta}_i - \beta_i &= \mathcal{C}_{i1}\mathbb{E}(x_1\Phi(\tilde{x})) + \mathcal{C}_{i2}\mathbb{E}(x_2\Phi(\tilde{x})) + \ldots + \mathcal{C}_{iq}\mathbb{E}\big(x_q\Phi(\tilde{x})\big) \\
&\geq \inf_{A_i} \Phi(\tilde{x})\big(\mathcal{C}_{i1}\mathcal{F}_{\mathcal{A}'}(x_1,\tilde{x}) + \mathcal{C}_{i2}\mathcal{F}_{\mathcal{A}'}(x_2,\tilde{x}) + \ldots + \mathcal{C}_{iq}\mathcal{F}_{\mathcal{A}'}(x_q,\tilde{x})\big) \\
&\geq \sup_{\mathcal{A}'}(\Phi(\tilde{x}) + \Omega(\tilde{x})) \\
&= \sup_{\mathcal{A}'}(\hat{g}(x) - g(x))
\end{aligned}$$

holds true, and notice that usually $\mathcal{A} \neq \mathcal{A}'$. □

Proof of Lemma 2.1: First of all, let's state a lemma from Audenaert (2010), for any real random variable x, we have $\mathrm{Var}(x) = \min_z \mathbb{E}(x - z)^2$. Following Murthy and Sethi (1965, 1966), by the lemma proposed by Audenaert (2010), we then have $\mathrm{Var}(\mathcal{C}_i x_i \Phi(\tilde{x})) \leq \mathbb{E}(\mathcal{C}_i x_i \Phi(\tilde{x}) - c)^2$, where $c = (\mathcal{C}_i\delta_i + \mathcal{C}_i\Phi(\tilde{x}))/2$ for $i = 1,2,\ldots,q_1$. We also have the inequality $(\mathcal{C}_i x_i \Phi(\tilde{x}) - c)^2 \leq r^2$ with $r = (\mathcal{C}_i\delta_i - \mathcal{C}_i\Phi(\tilde{x}))/2$ as well. Therefore, $\mathbb{E}(\mathcal{C}_i x_i \Phi(\tilde{x}) - c)^2 \leq r^2$ and the bounds for Lemma 1 hence follows. The inequality is sharp as equality is achieved for a distribution where $\mathcal{C}_i x_i \Phi(\tilde{x})$ is either δ_i or $\mathcal{C}_i\Phi(\tilde{x})$ with probability 1/2. □

Proof of Lemma 2.2: The proof directly follows Doukhan (1994) , Jiangqing Fan and Qiwei Yao (2003). □

Proof of Theorem 2.5: By Theorem 2.2 and Lemma 2.1, 2.2, we have

$$\mathrm{Var}(\mathbb{S}_n) = \sum_{i=1}^{q} \mathrm{Var}(\mathcal{C}_i x_i \Phi(\tilde{x})) + \sum_{i\neq j=1}^{q} \mathrm{Cov}\big(\mathcal{C}_i x_i \Phi(\tilde{x}), \mathcal{C}_j x_j \Phi(\tilde{x})\big)$$

$$\leq \sum_{i=1}^{q} \frac{(\mathcal{C}_i\delta_i - \mathcal{C}_i\Phi(\tilde{x}))^2}{4} + 4\sum_{i\neq j=1}^{q} \alpha_{ij}\delta_i\delta_j\,.$$

Then the theorem follows directly from Bienaymé-Chebyshev inequality where $\mathcal{A} = \cap_{i=1}^{q} A_i$, $A_i = x_i \cap \tilde{x}$ for all $i = 1,2,\ldots,q_1$. □

Proof of Theorem 3.1: Following Gasser-Müller (1984), based on the mean value theorem of integration, using the compactness of support of $\mathcal{K}_\vartheta$ and the Lipschitz continuity of $\mathcal{U}$, we have

$$\left|\frac{1}{h_1\cdots h_q}\sum_{j=1}^{q}\int_{S_{j-1}}^{S_j}\mathcal{K}\left(\frac{t-\mu}{h}\right)d\mu\,\varepsilon(t_{kj}) - \frac{1}{h_1\cdots h_q}\int_0^1\mathcal{K}\left(\frac{t-\mu}{h}\right)\varepsilon(\mu)d\mu\right|$$

$$\leq \left|\frac{n}{h_1\cdots h_q}\sum_{j=1}^{q}\int_{S_{j-1}}^{S_j}\mathcal{K}\left(\frac{t-\mu}{h}\right)d\mu\,\varepsilon(t_{kj}) - \frac{n}{h_1\cdots h_q}\sum_{j=1}^{q}\int_{S_{j-1}}^{S_j}\mathcal{K}\left(\frac{t-\mu}{h}\right)\varepsilon(\mu)\,d\mu\right|$$

$$\leq \left|\frac{n}{h_1\cdots h_q}\sum_{j=1}^{q}|S_j - S_{j-1}|\mathcal{K}(\zeta_j)\,\varepsilon(t_{kj}) - \frac{n}{h_1\cdots h_q}\sum_{j=1}^{q}|S_j - S_{j-1}|\mathcal{K}(\zeta_j)\,\varepsilon(\zeta_j)\right|$$

$$\leq \frac{n\cdot\max K(\zeta_j)}{h_1\cdots h_q}\sum_{j\in f}|S_j - S_{j-1}|\big|\varepsilon(t_{kj}) - \varepsilon(\zeta_j)\big|$$

$$\leq \frac{n\cdot\max K(\zeta_j)}{h_1\cdots h_q}|f|\left(\frac{c}{n}\right)^2 = \mathcal{O}\left(\frac{1}{n^2 h_{q_1+1}\cdots h_q}\right),$$

where ζ_j are suitable mean values and f denotes the set of observations with non-zero weights, so $|f| =$

$\mathcal{O}(\mathrm{n}^{-1}\mathrm{h}_1\cdots\mathrm{h}_\mathrm{q})$, see Gasser-Müller (1984). $|\cdot|$ is the cardinality, and c is a real number. If n is big enough, we can get

$$\frac{1}{\mathrm{h}_1\cdots\mathrm{h}_\mathrm{q}}\sum_{\mathrm{j}=\mathrm{q}_1+1}^{\mathrm{q}}\int_{\mathrm{S}_{\mathrm{j}-1}}^{\mathrm{S}_\mathrm{j}}\mathcal{K}\left(\frac{\mathrm{t}-\mu}{\mathrm{h}}\right)\mathrm{d}\mu\varepsilon(\tilde{\mathrm{t}}_{\mathrm{kj}})=\int_0^1\mathcal{K}\left(\frac{\mathrm{t}-\mu}{\mathrm{h}}\right)\varepsilon(\mu)\mathrm{d}\mu+\mathcal{O}\left(\frac{1}{\mathrm{n}^2\mathrm{h}_{\mathrm{q}_1+1}\cdots\mathrm{h}_\mathrm{q}}\right). \qquad \textbf{(A.8)}$$

Similarly,

$$\frac{1}{\mathrm{h}_1\cdots\mathrm{h}_\mathrm{q}}\sum_{\mathrm{j}=\mathrm{q}_1+1}^{\mathrm{q}}\int_{\mathrm{S}_{\mathrm{j}-1}}^{\mathrm{S}_\mathrm{j}}\mathcal{K}\left(\frac{\mathrm{t}-\mu}{\mathrm{h}}\right)\mathrm{d}\mu\cdot\mathcal{U}(\mathrm{t}_{\mathrm{kj}})=\int_0^1\mathcal{K}\left(\frac{\mathrm{t}-\mu}{\mathrm{h}}\right)\varepsilon(\mu)\mathcal{U}(\mu)+\mathcal{O}\left(\frac{1}{\mathrm{n}^2\mathrm{h}_{\mathrm{q}_1+1}\cdots\mathrm{h}_\mathrm{q}}\right). \qquad \textbf{(A.9)}$$

Substitute (A.8), (A.9) into formula (3.2) of section 3, we then get

$$\begin{aligned}\widehat{\mathcal{U}}_\mathrm{n}(\mathrm{t})&=\frac{1}{\mathrm{h}_1\mathrm{h}_2\cdots\mathrm{h}_\mathrm{q}}\sum_{\mathrm{j}=1}^{\mathrm{q}}\int_{\mathrm{s}_{\mathrm{j}-1}}^{\mathrm{s}_\mathrm{j}}\mathcal{K}\left(\frac{\mathrm{t}-\mu}{\mathrm{h}}\right)\mathrm{d}\mu\cdot(\mathcal{U}(\mathrm{t}_{\mathrm{ki}})+\mathcal{E}(\tilde{\mathrm{t}}_{\mathrm{ki}}))\\&=\int_0^1\mathcal{K}\left(\frac{\mathrm{t}-\mu}{\mathrm{h}}\right)\mathcal{U}(\mu)\mathrm{d}\,\mu+\int_0^1\mathcal{K}\left(\frac{\mathrm{t}-\mu}{\mathrm{h}}\right)\varepsilon(\mu)\mathrm{d}\mu+\mathcal{O}\left(\frac{2}{\mathrm{n}^2\mathrm{h}_{\mathrm{q}_1+1}\cdots\mathrm{h}_\mathrm{q}}\right).\end{aligned}$$

By change of variables, $\mu=\mathrm{t}-\mathrm{hv}$ and note that $\mathrm{h}\longrightarrow 0$ as $\mathrm{n}\longrightarrow\infty$. Carrying out Taylor expansions, we get

$$\begin{aligned}\widehat{\mathcal{U}}_\mathrm{n}(\mathrm{t})&=\int_{-\infty}^{+\infty}\mathcal{K}(v)\mathcal{U}(\mathrm{t}-\mathrm{hv})\mathrm{dv}+\int_{-\infty}^{+\infty}\mathcal{K}(v)\varepsilon(\mathrm{t}-\mathrm{hv})\mathrm{dv}+\mathcal{O}\left(\frac{2}{\mathrm{n}^2\mathrm{h}_{\mathrm{q}_1+1}\cdots\mathrm{h}_\mathrm{q}}\right)\\&=\int_{-\infty}^{+\infty}\mathcal{K}(v)\left[\mathcal{U}(\mathrm{t})+v\sum_{\mathrm{s}=1}^{\mathrm{q}}\frac{\partial\mathrm{U}(\mathrm{t})}{\partial\mathrm{t}_\mathrm{s}}\mathrm{h}_\mathrm{s}+\frac{1}{2}v^2\sum_{\mathrm{s}\neq\mathrm{s}'=1}^{\mathrm{q}}\frac{\partial^2\mathcal{U}(\mathrm{t})}{\partial\mathrm{t}_\mathrm{s}\partial\mathrm{t}_{\mathrm{s}'}}\mathrm{h}_\mathrm{s}^2+\mathcal{O}\left(\sum_{\mathrm{s}=1}^{\mathrm{q}}\mathrm{h}_\mathrm{s}^2\right)\right]\mathrm{dv}\\&\quad+\int_{-\infty}^{+\infty}\mathcal{K}(v)\left[\varepsilon(\mathrm{t})+v\sum_{\mathrm{s}=1}^{\mathrm{q}}\frac{\partial\varepsilon(\mathrm{t})}{\partial\mathrm{t}_\mathrm{s}}\mathrm{h}_\mathrm{s}+\frac{1}{2}v^2\sum_{\mathrm{s}\neq\mathrm{s}'=1}^{\mathrm{q}}\frac{\partial^2\varepsilon(\mathrm{t})}{\partial\mathrm{t}_\mathrm{s}\partial\mathrm{t}_{\mathrm{s}'}}\mathrm{h}_\mathrm{s}^2+\mathcal{O}\left(\sum_{\mathrm{s}=1}^{\mathrm{q}}\mathrm{h}_\mathrm{s}^2\right)\right]\mathrm{dv}\\&\quad+\mathrm{O}\left(\frac{2}{\mathrm{n}^2\mathrm{h}_{\mathrm{q}_1+1}\cdots\mathrm{h}_\mathrm{q}}\right)\\&=\mathcal{U}(\mathrm{t})+\varepsilon(\mathrm{t})+\frac{1}{2}\mathcal{K}_2\sum_{\mathrm{s}\neq\mathrm{s}'=1}^{\mathrm{q}}\frac{\partial^2\mathcal{U}(\mathrm{t})}{\partial\mathrm{t}_\mathrm{s}\partial\mathrm{t}_{\mathrm{s}'}}\mathrm{h}_\mathrm{s}^2+\frac{1}{2}\mathcal{K}_2\sum_{\mathrm{s}\neq\mathrm{s}'=1}^{\mathrm{q}}\frac{\partial^2\varepsilon(\mathrm{t})}{\partial\mathrm{t}_\mathrm{s}\partial\mathrm{t}_{\mathrm{s}'}}\mathrm{h}_\mathrm{s}^2+\mathcal{O}\left(\sum_{\mathrm{s}=1}^{\mathrm{q}}\mathrm{h}_\mathrm{s}^2+\frac{2}{\mathrm{n}^2\mathrm{h}_{\mathrm{q}_1+1}\cdots\mathrm{h}_\mathrm{q}}\right),\end{aligned}$$

where $\mathrm{h}_\mathrm{s}\longrightarrow 0$ as $\mathrm{n}\longrightarrow\infty$ for $\mathrm{s}=1,2,\ldots,\mathrm{q}_1$. □

Lemma A.1 (Gasser-Müller , 1984**)** For constants $\beta,\tau\in\mathrm{R}$, $0<\tau<\infty$, $k\in\mathbb{N}$, $\mathrm{k}\geq\vartheta+2$, we have:
(i) let k be a function with support $\mathrm{k}\subset[-\tau,\tau]$ which satisfies $\mathrm{k}^{(\mathrm{j})}(\tau)=\mathrm{k}^{(\mathrm{j})}(-\tau)=0$ for $\mathrm{j}=1,2,\ldots,\vartheta-1$ and the relation (a) below

$$\int_{-\tau}^{\tau}\mathrm{k}(\mathrm{x})\mathrm{x}^\mathrm{j}\mathrm{dx}=\begin{cases}1 & \mathrm{j}=0\\0 & \mathrm{j}=1,\ldots,\mathrm{k}-\vartheta-1,\\(-1)^\vartheta\beta\dfrac{(\mathrm{k}-\vartheta)!}{\mathrm{k}!} & \mathrm{j}=\mathrm{k}-\vartheta\end{cases} \qquad \textbf{(A.10)}$$

then the ϑth derivative of k, denoted by $\mathrm{k}_\vartheta=\mathrm{k}^{(\vartheta)}$, satisfies (b)

$$\int_{-\tau}^{\tau}\mathrm{k}_\vartheta(\mathrm{x})\mathrm{x}^\mathrm{j}\mathrm{dx}=\begin{cases}0 & \mathrm{j}=0,\ldots,\vartheta-1,\vartheta+1,\ldots,\mathrm{k}-1\\(-1)^\vartheta\vartheta! & \mathrm{j}=\vartheta\\\beta & \mathrm{j}=\mathrm{k}\end{cases} \qquad \textbf{(A.11)}$$

with support $\mathrm{k}_\vartheta\subset[-\tau,\tau]$.
(ii) if k_ϑ is a function satisfying (b), then there is a ϑ times differentiable function k satisfying (a).

Remark A.1 From Lemma A.1, we immediately get $\frac{1}{4}\int_{-\infty}^{+\infty}\mathcal{K}(\mathrm{u})\mathrm{u}^4\mathrm{du}>\frac{1}{2}\int_{-\infty}^{+\infty}\mathcal{K}(\mathrm{u})\mathrm{u}^2\mathrm{du}\cdot\int_{-\infty}^{+\infty}\mathcal{K}(\mathrm{u})\mathrm{u}^2\mathrm{du}$ for uniform kernels satisfying Lemma A.1. To illustrate this, we compute the explicit values for different kernel functions:

Kernels	$\frac{1}{4}\int_{-\infty}^{+\infty}\mathcal{K}(u)u^4du$	$\frac{1}{2}\left(\int_{-\infty}^{+\infty}\mathcal{K}(u)u^2du\right)^2$
Uniform kernel	0.05	0.045
Gauss kernel	0.75	0.5
Epanechnikov kernel	0.0214285	0.125
Quadratic kernel	0.0119	0.0102102
Logistic kernel	11.3644	5.41162
Sigmoid kernel	7.610075	3.04403
Cosin kernel	0.0196801	0.017941
Triangular kernel	0.0166675	0.0138894
Tricube kernel	0.01136375	0.0103723
Triweight kernel	0.01190475	0.01020449
Silverman kernel	-6	0

As we can see from the table, most kernels satisfy the inequality expect for the Epanechnikov kernel and the Silverman kernel.

Proof of Theorem 3.2: We carry out a Gasser-Müller's estimation to estimate the error term in model (3.1) of section 3. From (3.1), we have $\varepsilon_i = \varepsilon(\tilde{t}_{ki}) + \eta_i$, η_i are i.i.d. with $\mathbb{E}(\eta_i) = 0$, $\mathrm{Var}(\eta_i) = \sigma_\eta^2 < \infty$ and $\tilde{t}_{ki} \perp \eta_i$ for $i = 1,2,...,n; j = 1,2,...,q$. As shown in Theorem 3.1, we then get

$$\hat{\varepsilon}(\tilde{t}) = \frac{1}{h_{q_1+1}\cdots h_q}\sum_{j=1}^{q}\int_{s_{j-1}}^{s_j}\mathcal{K}\left(\frac{\tilde{t}-\mu}{h}\right)d\mu\cdot\varepsilon_i$$

$$= \frac{1}{h_{q_1+1}\cdots h_q}\sum_{j=1}^{q}\int_{s_{j-1}}^{s_j}\mathcal{K}\left(\frac{\tilde{t}-\mu}{h}\right)d\mu\cdot\varepsilon(\tilde{t}_{ki}) + \frac{n}{h_{q_1+1}\cdots h_q}\frac{1}{n}\sum_{j=1}^{q}\int_{s_{j-1}}^{s_j}\mathcal{K}\left(\frac{\tilde{t}-\mu}{h}\right)d\mu\cdot\eta_i$$

$$\longrightarrow_p \frac{1}{h_{q_1+1}\cdots h_q}\sum_{j=1}^{q}\int_{s_{j-1}}^{s_j}\mathcal{K}\left(\frac{\tilde{t}-\mu}{h}\right)d\mu\cdot\varepsilon(\tilde{t}_{kj}) + \frac{n}{h_{q_1+1}\cdots h_q}\mathbb{E}\left(\int\mathcal{K}\left(\frac{\tilde{t}-\mu}{h}\right)d\mu\cdot\eta\right)$$

$$= \int_{-\infty}^{+\infty}\mathcal{K}(\tilde{v})\varepsilon(\tilde{t}-\tilde{h}\tilde{v})d\tilde{v} + \mathcal{O}\left(\frac{1}{n^2}\right) + 0$$

$$= \int_{-\infty}^{+\infty}\mathcal{K}(\tilde{v})\left[\varepsilon(\tilde{t}) + \tilde{v}\sum_{s=q_1+1}^{q}\frac{\partial\varepsilon(\tilde{t})}{\partial\tilde{t}_s}\tilde{h}_s + \frac{\tilde{v}^2}{2}\sum_{s\neq s'=q_1+1}^{q}\frac{\partial^2\varepsilon(\tilde{t})}{\partial\tilde{t}_s\partial\tilde{t}_{s'}}\tilde{h}_s^2 + \mathcal{O}\left(\sum_{s=q_1+1}^{q}\tilde{h}_s^2\right)\right]d\tilde{v} + \mathcal{O}\left(\frac{1}{n^2}\right)$$

$$= \varepsilon(\tilde{t}) + \frac{1}{2}\widetilde{\mathcal{K}}_2\sum_{s\neq s'=q_1+1}^{q}\frac{\partial^2\varepsilon(\tilde{t})}{\partial\tilde{t}_s\partial\tilde{t}_{s'}}\tilde{h}_s^2 + \mathcal{O}\left(\sum_{s=q_1+1}^{q}\tilde{h}_s^2\right) + o_p(1).$$

According to Gasser and Müller (1984), we can get the ϑth derivative of $\varepsilon(\tilde{t})$ as

$$\varepsilon''(\tilde{t}) = \frac{1}{\left(h_{q_1+1}\cdots h_q\right)^3}\sum_{j=1}^{q}\int_{s_{j-1}}^{s_j}\mathcal{K}''\left(\frac{\tilde{t}-\mu}{\tilde{h}}\right)d\mu\cdot\varepsilon_j$$

$$= \frac{1}{(h_{q_1+1}\cdots h_q)^3}\sum_{j=1}^{q}\int_{s_{j-1}}^{s_j}\mathcal{K}''\left(\frac{\tilde{t}-\mu}{\tilde{h}}\right)d\mu\cdot\varepsilon(\tilde{t}_{kj}) + 0$$

$$= \int_0^1 \mathcal{K}''\left(\frac{\tilde{t}-\mu}{\tilde{h}}\right)\varepsilon(\mu)d\mu + O\left(\frac{1}{n^2}\right)$$

$$= \int_{-\infty}^{+\infty}\mathcal{K}''(\tilde{v})\varepsilon(\tilde{t}-\tilde{h}\tilde{v})d\tilde{v} + \mathcal{O}\left(\frac{1}{n^2}\right)$$

$$= \varepsilon(\tilde{t})\int_{-\infty}^{+\infty}\mathcal{K}''(\tilde{v})d\tilde{v} + \sum_{s=q_1+1}^{q}\frac{\partial\varepsilon(\tilde{t})}{\partial\tilde{t}_s}\tilde{h}_s\int_{-\infty}^{+\infty}\mathcal{K}''(\tilde{v})\tilde{v}d\tilde{v} + \frac{1}{2}\sum_{s\neq s'=q_1+1}^{q}\frac{\partial^2\varepsilon(\tilde{t})}{\partial\tilde{t}_s\partial\tilde{t}_{s'}}\tilde{h}_s^2\int_{-\infty}^{+\infty}\mathcal{K}''(\tilde{v})\tilde{v}^2d\tilde{v}$$

$$+ \mathcal{O}\left(\sum_{s=q_1+1}^{q}\tilde{h}_s^2\right)\int_{-\infty}^{+\infty}\mathcal{K}''(\tilde{v})d\tilde{v} + \mathcal{O}\left(\frac{1}{n^2}\right).$$

Using Lemma A.1, the above equation turns out to be

$$\varepsilon''(\tilde{t}) = \sum_{s\neq s'=q_1+1}^{q}\frac{\partial^2\varepsilon(\tilde{t})}{\partial\tilde{t}_s\partial\tilde{t}_{s'}}\tilde{h}_s^2 + \mathcal{O}\left(\frac{1}{n^2}\right) = \sum_{s\neq s'=q_1+1}^{q}\frac{\partial^2\varepsilon(\tilde{t})}{\partial\tilde{t}_s\partial\tilde{t}_{s'}}\tilde{h}_s^2 + o_p(1),$$

because of the fact that

$$|\frac{1}{(h_{q_1+1}\cdots h_q)^3}\sum_{j=1}^{q}\int_{s_{j-1}}^{s_j}\mathcal{K}''\left(\frac{\tilde{t}-\mu}{\tilde{h}}\right)d\mu\cdot\varepsilon(\tilde{t}_{kj}) - \frac{1}{h_{q_1+1}\cdots h_q}\int_0^1\mathcal{K}''\left(\frac{\tilde{t}-\mu}{\tilde{h}}\right)\varepsilon(\mu)d\mu|$$

$$\leq |\frac{n}{(h_{q_1+1}\cdots h_q)^3}\sum_{j=1}^{q}\int_{s_{j-1}}^{s_j}\mathcal{K}''\left(\frac{\tilde{t}-\mu}{\tilde{h}}\right)d\mu\cdot\varepsilon(\tilde{t}_{kj}) - \frac{n}{h_{q_1+1}\cdots h_q}\sum_{j=1}^{q}\int_{s_{j-1}}^{s_j}\mathcal{K}''\left(\frac{\tilde{t}-\mu}{\tilde{h}}\right)\varepsilon(\mu)\,d\mu$$

$$\leq \frac{n\cdot\max\mathcal{K}''(\zeta_j)}{h_{q_1+1}\cdots h_q}\sum_{j\in f}|s_j - s_{j-1}||\varepsilon(\tilde{t}_{kj}) - \varepsilon(\zeta_j)|$$

$$= \mathcal{O}\left(\frac{1}{n^2}\right),$$

where ζ_j are proper mean values for the mean value theorem of integration, and $|\not{f}| = \mathcal{O}(n^{-1}h_{q_1+1}\cdots h_q)$, $|\cdot|$ denotes the cardinality. Substitute the above formula into $\widehat{\mathcal{U}}_n(t)$, as $n\longrightarrow\infty$, we then get

$$\widehat{\mathcal{U}}_n(t) = \mathcal{U}(t) + \hat{\varepsilon}(\tilde{t}) + \frac{1}{2}\mathcal{K}_2\sum_{s\neq s'=1}^{q}\frac{\partial^2\mathcal{U}(t)}{\partial t_s\partial t_{s'}}h_s^2 + \frac{1}{2}\mathcal{K}_2\sum_{s\neq s'=1}^{q}\frac{\partial^2\hat{\varepsilon}(\tilde{t})}{\partial t_s\partial t_{s'}}h_s^2 + \mathcal{O}\left(\sum_{s=1}^{q}h_s^2 + \frac{2}{n^2h_{q_1+1}\cdots h_q}\right)$$

$$= \mathcal{U}(t) + \frac{1}{2}\mathcal{K}_2\sum_{s\neq s'=1}^{q}\frac{\partial^2\mathcal{U}(t)}{\partial t_s\partial t_{s'}}h_s^2 + \varepsilon(\tilde{t}) + \frac{1}{2}\mathcal{K}_2\sum_{s\neq s'=1}^{q}\frac{\partial^2\varepsilon(\tilde{t})}{\partial t_s\partial t_{s'}}h_s^2 + \mathcal{O}\left(\sum_{s=q_1+1}^{q}h_s^2\right) + o_p(1)$$

$$+ \mathcal{O}\left(\sum_{s=1}^{q}h_s^2 + \frac{2}{n^2h_{q_1+1}\cdots h_q}\right)$$

$$= \mathcal{U}(t) + \varepsilon(\tilde{t}) + \frac{1}{2}\mathcal{K}_2\sum_{s\neq s'=1}^{q}\frac{\partial^2\mathcal{U}(t)}{\partial t_s\partial t_{s'}}h_s^2 + \frac{1}{2}\mathcal{K}_2\sum_{s\neq s'=1}^{q}\frac{\partial^2\varepsilon(\tilde{t})}{\partial t_s\partial t_{s'}}h_s^2 + \mathcal{O}\left(2\sum_{s=q_1+1}^{q}h_s^2\right) + o_p(1)$$

$$= \mathcal{U}(t) + \varepsilon(\tilde{t}) + \frac{1}{2}\mathcal{K}_2 \sum_{s \neq s' = 1}^{q} \frac{\partial^2 \varepsilon(\tilde{t})}{\partial t_s \partial t_{s'}} h_s^2 + \mathcal{O}\left(2 \sum_{s = q_1 + 1}^{q} h_s^2\right) + o_p(1).$$

As for the Nadaraya-Watson's estimator, from Theorem 2.3

$$\hat{g}(x) = g(x)$$

$$+ \frac{1}{2}\sum_{s \neq s' = 1}^{q_1} \frac{\partial^2 g(x)}{\partial x_s \partial x_{s'}} h_s^2 \left(\mathcal{K}_2 f(\tilde{x}) + \frac{\mathcal{K}_4}{2}\sum_{s \neq s' = q_1 + 1}^{q} \frac{\partial^2 f(\tilde{x})}{\partial x_s \partial x_{s'}} h_s^2 + \mathcal{K}_2 \mathcal{O}\left(\sum_{s = q_1 + 1}^{q} h_s^2\right)\right) \Big/ \left(f(\tilde{x}) + \frac{\widetilde{\mathcal{K}}_2}{2}\sum_{s \neq s' = q_1 + 1}^{q} \frac{\partial^2 f(\tilde{x})}{\partial x_s \partial x_{s'}} h_s^2 + \mathcal{O}\left(\sum_{s = q_1 + 1}^{q} h_s^2\right.\right.$$

$$+ \left(\mathcal{K}_2 \sum_{s = q_1 + 1}^{q} \frac{\partial f(\tilde{x})}{\partial x_s} h_s \sum_{q_1 + 1}^{q} \frac{\partial g(x)}{\partial x_s} h_s + \frac{\mathcal{K}_3}{2} \sum_{q_1 + 1}^{q} \frac{\partial g(x)}{\partial x_s} h_s \sum_{s \neq s' = q_1 + 1}^{q} \frac{\partial^2 f(\tilde{x})}{\partial x_s \partial x_{s'}} h_s^2\right.$$

$$+ \frac{\mathcal{K}_3}{2} \sum_{s \neq s' = 1}^{q_1} \frac{\partial^2 g(x)}{\partial x_s \partial x_{s'}} h_s^2 \sum_{s = q_1 + 1}^{q} \frac{\partial f(\tilde{x})}{\partial x_s} h_s + f(\tilde{x}) \mathcal{O}\left(\sum_{s = q_1 + 1}^{q} h_s^2\right) + \frac{\mathcal{K}_2}{2} \sum_{s \neq s' = q_1 + 1}^{q} \frac{\partial^2 f(\tilde{x})}{\partial x_s \partial x_{s'}} h_s^2 \mathcal{O}\left(\sum_{s = q_1 + 1}^{q} h_s^2\right)$$

$$\left. + \mathcal{O}\left(2 \sum_{s = q_1 + 1}^{q} h_s^2\right)\right) \Big/ \left(f(\tilde{x}) + \frac{\widetilde{\mathcal{K}}_2}{2}\sum_{s \neq s' = q_1 + 1}^{q} \frac{\partial^2 f(\tilde{x})}{\partial x_s \partial x_{s'}} h_s^2 + \mathcal{O}\left(\sum_{s = q_1 + 1}^{q} h_s^2\right)\right)$$

$$+ u(\tilde{x})$$

$$+ \frac{1}{2} \sum_{s \neq s' = 1}^{q_1} \frac{\partial^2 u(\tilde{x})}{\partial x_s \partial x_{s'}} h_s^2 \left(\mathcal{K}_2 f(\tilde{x}) + \frac{\mathcal{K}_4}{2} \sum_{s \neq s' = q_1 + 1}^{q} \frac{\partial^2 f(\tilde{x})}{\partial x_s \partial x_{s'}} h_s^2 + \mathcal{K}_2 \mathcal{O}\left(\sum_{s = q_1 + 1}^{q} h_s^2\right)\right)$$

$$\Big/ \left(f(\tilde{x}) + \frac{\mathcal{K}_2}{2} \sum_{s \neq s' = q_1 + 1}^{q} \frac{\partial^2 f(\tilde{x})}{\partial x_s \partial x_{s'}} h_s^2 + \mathcal{O}\left(\sum_{s = q_1 + 1}^{q} h_s^2\right)\right)$$

$$+ \left(\mathcal{K}_2 \sum_{s = q_1 + 1}^{q} \frac{\partial f(\tilde{x})}{\partial x_s} h_s \sum_{q_1 + 1}^{q} \frac{\partial u(\tilde{x})}{\partial x_s} h_s + \frac{\mathcal{K}_3}{2} \sum_{q_1 + 1}^{q} \frac{\partial u(\tilde{x})}{\partial x_s} h_s \sum_{s \neq s' = q_1 + 1}^{q} \frac{\partial^2 f(\tilde{x})}{\partial x_s \partial x_{s'}} h_s^2\right.$$

$$+ \frac{\mathcal{K}_3}{2} \sum_{s \neq s' = 1}^{q_1} \frac{\partial^2 u(\tilde{x})}{\partial x_s \partial x_{s'}} h_s^2 \sum_{s = q_1 + 1}^{q} \frac{\partial f(\tilde{x})}{\partial x_s} h_s + f(\tilde{x}) \mathcal{O}\left(\sum_{s = q_1 + 1}^{q} h_s^2\right)$$

$$\left. + \frac{\mathcal{K}_2}{2} \sum_{s \neq s' = q_1 + 1}^{q} \frac{\partial^2 f(\tilde{x})}{\partial x_s \partial x_{s'}} h_s^2 \mathcal{O}\left(\sum_{s = q_1 + 1}^{q} h_s^2\right) + \mathcal{O}\left(2 \sum_{s = q_1 + 1}^{q} h_s^2\right)\right)$$

$$\Big/ \left(f(\tilde{x}) + \frac{\mathcal{K}_2}{2} \sum_{s \neq s' = q_1 + 1}^{q} \frac{\partial^2 f(\tilde{x})}{\partial x_s \partial x_{s'}} h_s^2 + \mathcal{O}\left(\sum_{s = q_1 + 1}^{q} h_s^2\right)\right).$$

Denote the third and the sixth term of the above equation as χ_1, χ_2 respectively, and notice that $\chi_1, \chi_2 > 0$. Using Lemma A.1 and Remark A.1, we can get $\mathcal{K}_4 > 2\mathcal{K}_2\mathcal{K}_2 > \mathcal{K}_2\mathcal{K}_2$, hence the above equation turns to be

$$\hat{g}(x) \geq g(x) + \left(\frac{1}{2} \sum_{s \neq s' = 1}^{q_1} \frac{\partial^2 g(x)}{\partial x_s \partial x_{s'}} h_s^2 \left(\mathcal{K}_2 f(\tilde{x}) + \frac{\mathcal{K}_4}{2} \sum_{s \neq s' = q_1 + 1}^{q} \frac{\partial^2 f(\tilde{x})}{\partial x_s \partial x_{s'}} h_s^2 + \mathcal{K}_2 \mathcal{O}\left(\sum_{s = q_1 + 1}^{q} h_s^2 \right) \right) \right)$$

$$/ \left(f(\tilde{x}) + \frac{\widetilde{\mathcal{K}}_2}{2} \sum_{s \neq s' = q_1 + 1}^{q} \frac{\partial^2 f(\tilde{x})}{\partial x_s \partial x_{s'}} h_s^2 + \mathcal{O}\left(\sum_{s = q_1 + 1}^{q} h_s^2 \right) \right) + u(\tilde{x})$$

$$+ \frac{1}{2} \sum_{s \neq s' = 1}^{q_1} \frac{\partial^2 u(\tilde{x})}{\partial x_s \partial x_{s'}} h_s^2 \left(\mathcal{K}_2 f(\tilde{x}) + \frac{\mathcal{K}_4}{2} \sum_{s \neq s' = q_1 + 1}^{q} \frac{\partial^2 f(\tilde{x})}{\partial x_s \partial x_{s'}} h_s^2 + \mathcal{K}_2 \mathcal{O}\left(\sum_{s = q_1 + 1}^{q} h_s^2 \right) \right)$$

$$/ \left(f(\tilde{x}) + \frac{\mathcal{K}_2}{2} \sum_{s \neq s' = q_1 + 1}^{q} \frac{\partial^2 f(\tilde{x})}{\partial x_s \partial x_{s'}} h_s^2 + \mathcal{O}\left(\sum_{s = q_1 + 1}^{q} h_s^2 \right) \right) + \chi_1 + \chi_2$$

$$= g(x) + u(\tilde{x}) + \frac{\mathcal{K}_2}{2} \sum_{s \neq s' = 1}^{q_1} \frac{\partial^2 g(x)}{\partial x_s \partial x_{s'}} h_s^2 + \frac{\mathcal{K}_2}{2} \sum_{s \neq s' = 1}^{q_1} \frac{\partial^2 u(\tilde{x})}{\partial x_s \partial x_{s'}} h_s^2 + \chi_1 + \chi_2,$$

compare the above equation with the Gasser-Muller's asymptotic bias, we finally get

$$\hat{g}_{NW}(x) - g(x) \geq \frac{\mathcal{K}_2}{2} \sum_{s \neq s' = 1}^{q_1} \frac{\partial^2 g(x)}{\partial x_s \partial x_{s'}} h_s^2 + \frac{\mathcal{K}_2}{2} \sum_{s \neq s' = 1}^{q_1} \frac{\partial^2 u(\tilde{x})}{\partial x_s \partial x_{s'}} h_s^2 + \chi_1 + \chi_2$$

$$\geq \frac{\mathcal{K}_2}{2} \sum_{s \neq s' = 1}^{q_1} \frac{\partial^2 g(x)}{\partial x_s \partial x_{s'}} h_s^2 + \frac{\mathcal{K}_2}{2} \sum_{s \neq s' = 1}^{q_1} \frac{\partial^2 u(\tilde{x})}{\partial x_s \partial x_{s'}} h_s^2 + \mathcal{O}\left(2 \sum_{s = q_1 + 1}^{q} h_s^2 \right) + o_p(1)$$

$$= \hat{g}_{GM}(x) - g(x),$$

where $\mathcal{U}(\cdot) = g(\cdot)$ is the true unknown regression function, $\varepsilon(\cdot) = u(\cdot)$ is the true error function where some explanatory variables are omitted. Thus this theorem is proved. □

Proof of Theorem 3.3: Following equation (3.3) in section 3.2, we carry out a Taylor expansion to the numerator part

$$\mathbb{E}\left(g(X) K\left(\frac{X - x}{h} \right) \right) \mathbb{E}\left((X - x)^2 \mathcal{K}\left(\frac{X - x}{h} \right) \right)$$

$$= \mathbb{E}\big(g(x + hv) K(v)\big) \mathbb{E}\big((hv)^2 \mathcal{K}(v)\big)$$

$$= \int g(x + hv) \mathcal{K}(v) f(v) dv \int (hv)^2 \mathcal{K}(v) f(v) dv$$

$$= \int \left[g(x) + v \sum_{s=1}^{q} \frac{\partial g(x)}{\partial x_s} h_s + \frac{1}{2} v^2 \sum_{s = s' = 1}^{q} \frac{\partial^2 g(x)}{\partial x_s \partial x_{s'}} h_s^2 + \mathcal{O}\left(\sum_{s=1}^{q} h_s^2 \right) \right] \mathcal{K}(v) f(v) dv \int (hv)^2 \mathcal{K}(v) f(v) dv$$

$$= \left[g(x) \int \mathcal{K}(v) f(v) dv + \sum_{s=1}^{q} \frac{\partial g(x)}{\partial x_s} h_s \int v \mathcal{K}(v) f(v) dv \right.$$

$$\left. + \frac{1}{2} \sum_{s = s' = 1}^{q} \frac{\partial^2 g(x)}{\partial x_s \partial x_{s'}} h_s^2 \int v^2 \mathcal{K}(v) f(v) dv + \mathcal{O}\left(\sum_{s=1}^{q} h_s^2 \right) \int \mathcal{K}(v) f(v) dv \right] \int (hv)^2 \mathcal{K}(v) f(v) dv;$$

and

$$\mathbb{E}\left(g(X) \mathcal{K}\left(\frac{X - x}{h} \right) (X - x) \right) \mathbb{E}\left((X - x) \mathcal{K}\left(\frac{X - x}{h} \right) \right)$$

$$= \mathbb{E}(g(x+hv)K(\mathcal{v})hv)\mathbb{E}(hvK(\mathcal{v}))$$

$$= \int g(x+hv)\mathcal{K}(\mathcal{v})hvf(v)dv \int hvK(\mathcal{v})f(v)dv$$

$$= \int \left(g(x) + \mathcal{v}\sum_{s=1}^{q} \frac{\partial g(x)}{\partial x_s} h_s + \frac{1}{2}\mathcal{v}^2 \sum_{s=s'=1}^{q} \frac{\partial^2 g(x)}{\partial x_s \partial x_{s'}} h_s^2 + \mathcal{O}\left(\sum_{s=1}^{q} h_s^2\right)\right) \mathcal{K}(\mathcal{v})hvf(v)d\mathcal{v} \int hvK(\mathcal{v})f(v)dv$$

$$= \left(g(x)\int \mathcal{K}(\mathcal{v})hvf(\mathcal{v})d\mathcal{v} + \sum_{s=1}^{q} \frac{\partial g(x)}{\partial x_s} h_s \int \mathcal{v}^2\mathcal{K}(\mathcal{v})hf(v)\,dv + \frac{1}{2}\sum_{s=s'=1}^{q} \frac{\partial^2 g(x)}{\partial x_s \partial x_{s'}} h_s^2 \int \mathcal{v}^3\mathcal{K}(\mathcal{v})hf(v)\,dv \right.$$

$$\left. + O\left(\sum_{s=1}^{q} h_s^2\right)\int \mathcal{K}(\mathcal{v})hvf(v)dv \right) \cdot \int hvK(\mathcal{v})f(v)dv.$$

Similarly, we can get

$$\mathbb{E}\left(u(\tilde{X})\mathcal{K}\left(\frac{X-x}{h}\right)\right)\mathbb{E}\left((X-x)^2\mathcal{K}\left(\frac{X-x}{h}\right)\right)$$

$$= \left(u(\tilde{x})\int\int \mathcal{K}(\mathcal{v})f(\mathcal{v})f(\tilde{\mathcal{v}})dvd\tilde{\mathcal{v}} + \sum_{s=q_1+1}^{q} \frac{\partial u(\tilde{x})}{\partial x_s} h_s \int\int \tilde{\mathcal{v}}\,\mathcal{K}(\mathcal{v})f(\mathcal{v})f(\tilde{\mathcal{v}})dvd\tilde{\mathcal{v}} \right.$$

$$+ \frac{1}{2}\sum_{s\neq s'=q_1+1}^{q} \frac{\partial^2 u(\tilde{x})}{\partial x_s \partial x_{s'}} h_s^2 \int\int \tilde{\mathcal{v}}^2\,\mathcal{K}(\mathcal{v})f(\mathcal{v})f(\tilde{\mathcal{v}})dvd\tilde{\mathcal{v}}$$

$$\left. + \mathcal{O}\left(\sum_{s=q_1+1}^{q} h_s^2\right)\int\int \mathcal{K}(\mathcal{v})f(\mathcal{v})f(\tilde{\mathcal{v}})dvd\tilde{\mathcal{v}} \right) \cdot \int (hv)^2\mathcal{K}(\mathcal{v})f(v)dv;$$

and

$$\mathbb{E}\left(u(\tilde{X})\mathcal{K}\left(\frac{X-x}{h}\right)(X-x)\right)\mathbb{E}\left((X-x)\mathcal{K}\left(\frac{X-x}{h}\right)\right)$$

$$= \left(u(\tilde{x})\int\int hvK(\mathcal{v})f(\mathcal{v})f(\tilde{\mathcal{v}})dvd\tilde{\mathcal{v}} + \sum_{s=q_1+1}^{q} \frac{\partial u(\tilde{x})}{\partial x_s} h_s \int\int hv\tilde{\mathcal{v}}\,\mathcal{K}(\mathcal{v})f(\mathcal{v})f(\tilde{\mathcal{v}})dvd\tilde{\mathcal{v}} \right.$$

$$+ \frac{1}{2}\sum_{s\neq s'=q_1+1}^{q} \frac{\partial^2 u(\tilde{x})}{\partial x_s \partial x_{s'}} h_s^2 \int\int hv\tilde{\mathcal{v}}^2\,\mathcal{K}(\mathcal{v})f(\mathcal{v})f(\tilde{\mathcal{v}})dvd\tilde{\mathcal{v}}$$

$$\left. + \mathcal{O}\left(\sum_{s=q_1+1}^{q} h_s^2\right)\int\int hvK(\mathcal{v})f(\mathcal{v})f(\tilde{\mathcal{v}})dvd\tilde{\mathcal{v}} \right) \cdot \int hvK(\mathcal{v})f(v)dv.$$

As for the denominator part, by the Taylor method again, we can get

$$\mathbb{E}\left(\mathcal{K}\left(\frac{X-x}{h}\right)\right)\mathbb{E}\left((X-x)^2\mathcal{K}\left(\frac{X-x}{h}\right)\right)$$

$$= \int \mathcal{K}(\mathcal{v})f(\mathcal{v})dv \cdot \int (hv)^2\,\mathcal{K}(\mathcal{v})f(\mathcal{v})dv;$$

similarly,

$$\mathbb{E}\left((X-x)\mathcal{K}\left(\frac{X-x}{h}\right)\right)\mathbb{E}\left((X-x)\mathcal{K}\left(\frac{X-x}{h}\right)\right)$$

$$= \int \mathrm{h}v\mathrm{K}(v)\mathrm{f}(v)\mathrm{d}v \cdot \int \mathrm{h}v\mathrm{K}(v)\mathrm{f}(v)\mathrm{d}v\,.$$

Substitute the countdown first, second, fifth and sixth equation into (3.3) of section 3, the first term turns out to be

$$\Big(\mathrm{h}^2\mathrm{g}(\mathrm{x})\Big(\int \mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v \cdot \int v^2\,\mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v - \int v\mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v \cdot \int v\mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v\Big)$$

$$+\,\mathrm{h}^2\sum_{s=1}^{q}\frac{\partial \mathrm{g}(\mathrm{x})}{\partial \mathrm{x}_s}\mathrm{h}_s\Big(\int v\mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v \cdot \int v^2\,\mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v - \int v^2\,\mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v$$

$$\cdot \int v\mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v\Big)$$

$$+\,\mathrm{h}^2\frac{1}{2}\sum_{s=s^{'}=1}^{q}\frac{\partial^2 \mathrm{g}(\mathrm{x})}{\partial \mathrm{x}_s \partial \mathrm{x}_{s^{'}}}\mathrm{h}_s^2\Big(\int v^2\,\mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v \cdot \int v^2\,\mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v - \int v^3\,\mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v$$

$$\cdot \int v\,\mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v\Big)$$

$$+\,\mathrm{h}^2\mathcal{O}\left(\sum_{s=1}^{q}\mathrm{h}_s^2\right)\Big(\int \mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v \cdot \int v^2\,\mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v - \int v\mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v$$

$$\cdot \int v\mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v\Big)\Big)$$

$$/\Big(\mathrm{h}^2\int \mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v \cdot \int v^2\,\mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v - \mathrm{h}^2\int v\mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v \cdot \int v\mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v\Big)$$

$$= \mathrm{g}(\mathrm{x})$$

$$+\frac{1}{2}\sum_{s=s^{'}=1}^{q}\frac{\partial^2 \mathrm{g}(\mathrm{x})}{\partial \mathrm{x}_s \partial \mathrm{x}_{s^{'}}}\mathrm{h}_s^2\frac{\int v^2\,\mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v \cdot \int v^2\,\mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v - \int v^3\,\mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v \cdot \int v\,\mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v}{\int \mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v \cdot \int v^2\,\mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v - \int v\mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v \cdot \int v\mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v}$$

$$+\,\mathcal{O}\left(\sum_{s=1}^{q}\mathrm{h}_s^2\right)$$

$$= \mathrm{g}(\mathrm{x}) + \frac{1}{2}\sum_{s=s^{'}=1}^{q}\frac{\partial^2 \mathrm{g}(\mathrm{x})}{\partial \mathrm{x}_s \partial \mathrm{x}_{s^{'}}}\mathrm{h}_s^2 \cdot$$

$$\frac{\int v^2\,\mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v \cdot \int v^0\,\mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v \cdot \int v^2\,\mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v - \int v^2\,\mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v \cdot \int v\,\mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v \cdot \int v\,\mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v}{\int \mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v \cdot \int v^2\,\mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v - \int v\mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v \cdot \int v\mathcal{K}(v)\mathrm{f}(v)\mathrm{d}v}$$

$$+\,\mathcal{O}\left(\sum_{s=1}^{q}\mathrm{h}_s^2\right)$$

$$= \mathrm{g}(\mathrm{x}) + \frac{\mathcal{K}_2}{2}\sum_{s=s^{'}=1}^{q}\frac{\partial^2 \mathrm{g}(\mathrm{x})}{\partial \mathrm{x}_s \partial \mathrm{x}_{s^{'}}}\mathrm{h}_s^2 + \mathcal{O}\left(\sum_{s=1}^{q}\mathrm{h}_s^2\right).$$

Substitute the countdown first, second, third and fourth equation into (3.3) of section 3, the second term turns out to be

$$\Big(h^2u(\tilde{x})\Big(\iint \mathcal{K}(v)f(v)f(\tilde{v})dvd\tilde{v}\cdot\int v^2\mathcal{K}(v)f(v)dv-\iint v\mathcal{K}(v)f(v)f(\tilde{v})dvd\tilde{v}\cdot\int v\mathcal{K}(v)f(v)dv\Big)$$

$$+h^2\sum_{s\neq s'=q_1+1}^{q}\frac{\partial^2u(\tilde{x})}{\partial x_s\partial x_{s'}}h_s^2\Big(\iint \tilde{v}^2\mathcal{K}(v)f(v)f(\tilde{v})dvd\tilde{v}\cdot\int v^2\mathcal{K}(v)f(v)dv$$

$$-\iint \tilde{v}^2v\mathcal{K}(v)f(v)f(\tilde{v})dvd\tilde{v}\cdot\int v\mathcal{K}(v)f(v)dv\Big)$$

$$+h^2\sum_{s=q_1+1}^{q}\frac{\partial u(\tilde{x})}{\partial x_s}h_s\Big(\iint \tilde{v}\mathcal{K}(v)f(v)f(\tilde{v})dvd\tilde{v}\cdot\int v^2\mathcal{K}(v)f(v)dv$$

$$-\iint \tilde{v}v\mathcal{K}(v)f(v)f(\tilde{v})dvd\tilde{v}\cdot\int v\mathcal{K}(v)f(v)dv\Big)$$

$$+h^2\mathcal{O}\left(\sum_{s=q_1+1}^{q}h_s^2\right)\Big(\iint \mathcal{K}(v)f(v)f(\tilde{v})dvd\tilde{v}\cdot\int v^2\mathcal{K}(v)f(v)dv$$

$$-\iint v\mathcal{K}(v)f(v)f(\tilde{v})dvd\tilde{v}\cdot\int v\mathcal{K}(v)f(v)dv\Big)\Big)$$

$$/\Big(\Big(h^2\int \mathcal{K}(v)f(v)dv\cdot\int v^2\,\mathcal{K}(v)f(v)dv-h^2\int v\mathcal{K}(v)f(v)dv\cdot\int v\mathcal{K}(v)f(v)dv\Big)\Big)$$

$$=u(\tilde{x})+\sum_{s=q_1+1}^{q}\frac{\partial u(\tilde{x})}{\partial x_s}h_s\int \tilde{v}f(\tilde{v})d\tilde{v}+\frac{1}{2}\sum_{s\neq s'=q_1+1}^{q}\frac{\partial^2u(\tilde{x})}{\partial x_s\partial x_{s'}}h_s^2\int \tilde{v}^2f(\tilde{v})d\tilde{v}+\mathcal{O}\left(\sum_{s=q_1+1}^{q}h_s^2\right).$$

Notice that $\int \tilde{v}f(\tilde{v})d\tilde{v}=\mathbb{E}(\tilde{v})=\mathbb{E}\left(\frac{X-x}{h}\right)=\frac{1}{h}(\mathbb{E}(x)-x)=0$, where $\frac{1}{n}\sum_{i=1}^{n}x_i$ is an unbiased estimator of x. So the above equation turns to be

$$\hat{u}(\tilde{x})+\frac{1}{2}\sum_{s\neq s'=q_1+1}^{q}\frac{\partial^2\hat{u}(\tilde{x})}{\partial x_s\partial x_{s'}}h_s^2\int \tilde{v}^2f(\tilde{v})d\tilde{v}+\mathcal{O}\left(\sum_{s=q_1+1}^{q}h_s^2\right).$$

We then consider a nonparametric structure of the error term $\varepsilon=\varepsilon(\tilde{x})+\eta$ where $\tilde{x}\perp\eta$ as mentioned above in the proof of Theorem 3.2, and using the local linear estimation method to estimate $\hat{u}(\tilde{x})$, $\hat{u}''(\tilde{x})$ in the above equation. According to Jiangqing Fan and Gijbels (1996), the second term of the above equation does not exist, and it finally turns out to be

$$u(\tilde{x})+\frac{\mathcal{K}_2}{2}\sum_{s\neq s'=q_1+1}^{q}\frac{\partial^2u(\tilde{x})}{\partial x_s\partial x_{s'}}h_s^2+\mathcal{O}\left(\sum_{s=q_1+1}^{q}h_s^2\right).$$

Substitute the countdown first and fourth equation into (3.3) of section 3, we finally get the asymptotic expansions for the local linear nonparametric estimator when several explanatory variables are omitted

$$\hat{g}_{LL}(x)=g(x)+u(\tilde{x})+\frac{\mathcal{K}_2}{2}\sum_{s\neq s'=1}^{q}\frac{\partial^2g(x)}{\partial x_s\partial x_{s'}}h_s^2+\frac{\mathcal{K}_2}{2}\sum_{s\neq s'=1}^{q}\frac{\partial^2u(\tilde{x})}{\partial x_s\partial x_{s'}}h_s^2+\mathcal{O}\left(\sum_{s=1}^{q}h_s^2+\sum_{s=q_1+1}^{q}h_s^2\right)$$

$$=g(x)+u(\tilde{x})+\frac{\mathcal{K}_2}{2}\sum_{s\neq s'=1}^{q}\frac{\partial^2u(\tilde{x})}{\partial x_s\partial x_{s'}}h_s^2+\mathcal{O}\left(\sum_{s=q_1+1}^{q}h_s^2\right)+o_p(1)$$

as $n\longrightarrow\infty$, $h_s\longrightarrow 0$ for $s=1,2,\ldots,q_1$. □

Lemma A.2. Suppose Assumptions 2.1-2.4, Assumptions 4.1-4.3 hold true, then

$$\Phi(\tilde{x}) = o_p(1)$$

as $n \longrightarrow \infty$, $h_s \longrightarrow 0$ for $s = 1,2,\ldots,q_1$.

Proof of lemma A.2: As $n \longrightarrow \infty$, from Assumptions 4.1-4.3: $\sum_{\bar{\alpha}=-m}^{m} U_{\bar{\alpha}} \longrightarrow \infty$. Let $\mathbb{P}_n = \sum_{\bar{\alpha}=-m}^{m} U_{\bar{\alpha}} + U_{\bar{x}}$. So we can get $f_{\bar{x}}(\mathbb{P}_n) = \mathcal{O}(|\mathbb{P}_n|^{-\mathcal{V}-1}) = o_p(1)$, $f'_{\bar{x}}(\mathbb{P}_n) = \mathcal{O}(|\mathbb{P}_n|^{-\mathcal{V}-2}) = o_p(1)$ and $f''_{\bar{x}}(\mathbb{P}_n) = \mathcal{O}(|\mathbb{P}_n|^{-\mathcal{V}-3}) = o_p(1)$ while $f''_{\bar{x}}(\mathbb{P}_n) = \mathcal{O}(f'_{\bar{x}}(\mathbb{P}_n)) = \mathcal{O}(f_{\bar{x}}(\mathbb{P}_n))$. Substitute these into $\Phi(\tilde{x})$ in Theorem 2.1, we then get $\Phi(\tilde{x}) = o_p(1)$. □

Proof of theorem 4.1: We prove this theorem part by part. Suppose the researches mistakenly think that the variables $\bar{x}, \tilde{x}$ are independently distributed and the variables $\tilde{x}$ are excluded from regression models, hence variables $\tilde{x}$ are omitted. Then assume that the Assumptions 2.1-2.4, 4.1-4.3 all hold true:

(i) Following (A.4), by a change of variables, Taylor expansions and the Fubini theorem, we can get

$$\mathbb{E}\left(u(\bar{x})\mathcal{K}\left(\frac{\bar{X}-\bar{x}}{\bar{h}}\right)\mathcal{K}\left(\frac{\tilde{X}-\tilde{x}}{\tilde{h}}\right)\right)$$

$$= \mathbb{E}\left(U\left(\sum_{\tilde{\alpha}=-m}^{m} U_{\tilde{\alpha}} + U_{\tilde{x}}\right)\mathcal{K}\left(\frac{\sum_{\bar{\alpha}=-m}^{m} U_{\bar{\alpha}} + U_{\bar{x}} - \bar{x}}{\bar{h}}\right)\mathcal{K}\left(\frac{\sum_{\tilde{\alpha}=-m}^{m} U_{\tilde{\alpha}} + U_{\tilde{x}} - \tilde{x}}{\tilde{h}}\right)\right)$$

$$= \int\int U\left(\sum_{\tilde{\alpha}=-m}^{m} U_{\tilde{\alpha}} + U_{\tilde{x}}\right)\mathcal{K}\left(\frac{\sum_{\bar{\alpha}=-m}^{m} U_{\bar{\alpha}} + U_{\bar{x}} - \bar{x}}{\bar{h}}\right)\mathcal{K}\left(\frac{\sum_{\tilde{\alpha}=-m}^{m} U_{\tilde{\alpha}} + U_{\tilde{x}} - \tilde{x}}{\tilde{h}}\right) f(U_{\bar{x}},U_{\tilde{x}})dU_{\bar{x}}dU_{\tilde{x}}$$

$$= \int\int U(\tilde{h}\tilde{v} + \tilde{x}) f\left(\bar{h}\bar{v} + \bar{x} - \sum_{\bar{\alpha}=-m}^{m} U_{\bar{\alpha}}\right) f\left(\tilde{h}\tilde{v} + \tilde{x} - \sum_{\tilde{\alpha}=-m}^{m} U_{\tilde{\alpha}}\right)\mathcal{K}(\bar{v})\mathcal{K}(\tilde{v})d\bar{v}d\tilde{v}$$

$$= \left[f(\bar{x}) - \sum_{\bar{\alpha}=-m}^{m} U_{\bar{\alpha}} \sum_{s=1}^{q_1} \frac{\partial f(\bar{x})}{\partial x_s} h_s + \frac{1}{2}\left(\sum_{\bar{\alpha}=-m}^{m} U_{\bar{\alpha}}\right)^2 \sum_{s=1}^{q_1} \frac{\partial^2 f(\bar{x})}{\partial x_s \partial x_{s'}} h_s^2 + \mathcal{O}\left(\bar{h}\bar{v} - \sum_{\bar{\alpha}=-m}^{m} U_{\bar{\alpha}}\right)^2\right] [u(\tilde{x})f(\tilde{x})$$

$$- u(\tilde{x}) \sum_{\tilde{\alpha}=-m}^{m} U_{\tilde{\alpha}} \sum_{s=q_1+1}^{q} \frac{\partial f(\tilde{x})}{\partial x_s} h_s + \frac{1}{2} u(\tilde{x})\left(\sum_{\tilde{\alpha}=-m}^{m} U_{\tilde{\alpha}}\right)^2 \sum_{s=q_1+1}^{q} \frac{\partial^2 f(\tilde{x})}{\partial x_s \partial x_{s'}} h_s^2$$

$$+ u(\tilde{x})\mathcal{O}\left(\tilde{h}\tilde{v} - \sum_{\tilde{\alpha}=-m}^{m} U_{\tilde{\alpha}}\right)^2 + f(\tilde{x})\mathcal{O}(\tilde{h}^2) - \sum_{\tilde{\alpha}=-m}^{m} U_{\tilde{\alpha}} \sum_{s=q_1+1}^{q} \frac{\partial f(\tilde{x})}{\partial x_s} h_s\, \mathcal{O}(\tilde{h}^2)$$

$$+ \frac{1}{2}\left(\sum_{\tilde{\alpha}=-m}^{m} U_{\tilde{\alpha}}\right)^2 \sum_{s=q_1+1}^{q} \frac{\partial^2 f(\tilde{x})}{\partial x_s \partial x_{s'}} h_s^2\, \mathcal{O}(\tilde{h}^2) + \mathcal{O}\left(\bar{h}^3 + \left(\tilde{h}\tilde{v} - \sum_{\tilde{\alpha}=-m}^{m} U_{\tilde{\alpha}}\right)^2\right)\Bigg],$$

where we set $U_{\bar{\alpha}} = \bar{h}\bar{v} + \bar{x} - \sum_{\bar{\alpha}=-m}^{m} U_{\bar{\alpha}}$, $U_{\tilde{\alpha}} = \tilde{h}\tilde{v} + \tilde{x} - \sum_{\tilde{\alpha}=-m}^{m} U_{\tilde{\alpha}}$ respectively. Similarly, we can get

$$\mathbb{E}\left(\mathcal{K}\left(\frac{X-x}{h}\right)\right)$$

$$= \mathbb{E}\left(\mathcal{K}\left(\frac{\bar{X}-\bar{x}}{\bar{h}}\right)\mathcal{K}\left(\frac{\tilde{X}-\tilde{x}}{\tilde{h}}\right)\right)$$

$$= \int\int \mathcal{K}\left(\frac{\bar{X}-\bar{x}}{\bar{h}}\right)\mathcal{K}\left(\frac{\tilde{X}-\tilde{x}}{\tilde{h}}\right) f(\bar{X},\tilde{X})d\bar{X}d\tilde{X}$$

$$= \int\int f\left(\bar{h}\bar{v} + \bar{x} - \sum_{\bar{\alpha}=-m}^{m} U_{\bar{\alpha}}\right) f\left(\tilde{h}\tilde{v} + \tilde{x} - \sum_{\tilde{\alpha}=-m}^{m} U_{\tilde{\alpha}}\right)\mathcal{K}(\bar{v})\mathcal{K}(\tilde{v})d\bar{v}d\tilde{v}$$

$$=\left[\mathrm{f}(\bar{\mathrm{x}})-\sum_{\bar{\alpha}=-\mathrm{m}}^{\mathrm{m}}\mathrm{U}_{\bar{\alpha}}\sum_{\mathrm{s}=1}^{\mathrm{q}_1}\frac{\partial \mathrm{f}(\bar{\mathrm{x}})}{\partial \mathrm{x}_\mathrm{s}}\mathrm{h}_\mathrm{s}+\frac{1}{2}\left(\sum_{\bar{\alpha}=-\mathrm{m}}^{\mathrm{m}}\mathrm{U}_{\bar{\alpha}}\right)^2\sum_{\mathrm{s}=1}^{\mathrm{q}_1}\frac{\partial^2 \mathrm{f}(\bar{\mathrm{x}})}{\partial \mathrm{x}_\mathrm{s}\partial \mathrm{x}_{\mathrm{s}'}}\mathrm{h}_\mathrm{s}^2+\mathcal{O}\left(\bar{\mathrm{h}}\bar{\mathcal{v}}-\sum_{\bar{\alpha}=-\mathrm{m}}^{\mathrm{m}}\mathrm{U}_{\bar{\alpha}}\right)^2\right][\mathrm{f}(\tilde{\mathrm{x}})$$

$$-\sum_{\tilde{\alpha}=-\mathrm{m}}^{\mathrm{m}}\mathrm{U}_{\tilde{\alpha}}\sum_{\mathrm{s}=\mathrm{q}_1+1}^{\mathrm{q}}\frac{\partial \mathrm{f}(\tilde{\mathrm{x}})}{\partial \mathrm{x}_\mathrm{s}}\mathrm{h}_\mathrm{s}+\frac{1}{2}\left(\sum_{\tilde{\alpha}=-\mathrm{m}}^{\mathrm{m}}\mathrm{U}_{\tilde{\alpha}}\right)^2\sum_{\mathrm{s}=\mathrm{q}_1+1}^{\mathrm{q}}\frac{\partial^2 \mathrm{f}(\tilde{\mathrm{x}})}{\partial \mathrm{x}_\mathrm{s}\partial \mathrm{x}_{\mathrm{s}'}}\mathrm{h}_\mathrm{s}^2+\mathcal{O}\left(\tilde{\mathrm{h}}\tilde{\mathcal{v}}-\sum_{\tilde{\alpha}=-\mathrm{m}}^{\mathrm{m}}\mathrm{U}_{\tilde{\alpha}}\right)^2\right],$$

and

$$\mathbb{E}\left(\mathrm{g}(\mathrm{x})\mathrm{K}\left(\frac{\mathrm{X}-\mathrm{x}}{\bar{\mathrm{h}}}\right)\right)$$

$$=\mathbb{E}\left(\mathrm{g}\left(\sum_{\alpha=-\mathrm{m}}^{\mathrm{m}}\mathrm{U}_\alpha+\mathrm{U}_\mathrm{x}\right)\mathcal{K}\left(\frac{\sum_{\bar{\alpha}=-\mathrm{m}}^{\mathrm{m}}\mathrm{U}_{\bar{\alpha}}+\mathrm{U}_{\bar{\mathrm{x}}}-\bar{\mathrm{x}}}{\bar{\mathrm{h}}}\right)\mathcal{K}\left(\frac{\sum_{\tilde{\alpha}=-\mathrm{m}}^{\mathrm{m}}\mathrm{U}_{\tilde{\alpha}}+\mathrm{U}_{\tilde{\mathrm{x}}}-\tilde{\mathrm{x}}}{\tilde{\mathrm{h}}}\right)\right)$$

$$=\int\int \mathrm{g}(\bar{\mathrm{h}}\bar{\mathcal{v}}+\bar{\mathrm{x}},\tilde{\mathrm{h}}\tilde{\mathcal{v}}+\tilde{\mathrm{x}})\mathrm{f}\left(\bar{\mathrm{h}}\bar{\mathcal{v}}+\bar{\mathrm{x}}-\sum_{\bar{\alpha}=-\mathrm{m}}^{\mathrm{m}}\mathrm{U}_{\bar{\alpha}}\right)\mathrm{f}\left(\tilde{\mathrm{h}}\tilde{\mathcal{v}}+\tilde{\mathrm{x}}-\sum_{\tilde{\alpha}=-\mathrm{m}}^{\mathrm{m}}\mathrm{U}_{\tilde{\alpha}}\right)\mathcal{K}(\bar{\mathcal{v}})\mathcal{K}(\tilde{\mathcal{v}})\mathrm{d}\bar{\mathcal{v}}\mathrm{d}\tilde{\mathcal{v}}$$

$$=\left[\mathrm{g}(\mathrm{x})\mathrm{f}(\bar{\mathrm{x}})-\mathrm{g}(\mathrm{x})\sum_{\bar{\alpha}=-\mathrm{m}}^{\mathrm{m}}\mathrm{U}_{\bar{\alpha}}\sum_{\mathrm{s}=1}^{\mathrm{q}_1}\frac{\partial \mathrm{f}(\bar{\mathrm{x}})}{\partial \mathrm{x}_\mathrm{s}}\mathrm{h}_\mathrm{s}+\frac{1}{2}\mathrm{g}(\mathrm{x})\left(\sum_{\bar{\alpha}=-\mathrm{m}}^{\mathrm{m}}\mathrm{U}_{\bar{\alpha}}\right)^2\sum_{\mathrm{s}=1}^{\mathrm{q}_1}\frac{\partial^2 \mathrm{f}(\bar{\mathrm{x}})}{\partial \mathrm{x}_\mathrm{s}\partial \mathrm{x}_{\mathrm{s}'}}\mathrm{h}_\mathrm{s}^2+\mathrm{g}(\mathrm{x})\mathcal{O}\left(\bar{\mathrm{h}}\bar{\mathcal{v}}-\sum_{\bar{\alpha}=-\mathrm{m}}^{\mathrm{m}}\mathrm{U}_{\bar{\alpha}}\right)^2\right.$$

$$+\mathrm{f}(\bar{\mathrm{x}})\mathcal{O}\left(\bar{\mathrm{h}}^2\bar{\mathcal{v}}^2+\tilde{\mathrm{h}}^2\tilde{\mathcal{v}}^2\right)-\sum_{\bar{\alpha}=-\mathrm{m}}^{\mathrm{m}}\mathrm{U}_{\bar{\alpha}}\sum_{\mathrm{s}=1}^{\mathrm{q}_1}\frac{\partial \mathrm{f}(\bar{\mathrm{x}})}{\partial \mathrm{x}_\mathrm{s}}\mathrm{h}_\mathrm{s}\,\mathcal{O}\left(\bar{\mathrm{h}}^2\bar{\mathcal{v}}^2+\tilde{\mathrm{h}}^2\tilde{\mathcal{v}}^2\right)$$

$$+\frac{1}{2}\left(\sum_{\bar{\alpha}=-\mathrm{m}}^{\mathrm{m}}\mathrm{U}_{\bar{\alpha}}\right)^2\sum_{\mathrm{s}=1}^{\mathrm{q}_1}\frac{\partial^2 \mathrm{f}(\bar{\mathrm{x}})}{\partial \mathrm{x}_\mathrm{s}\partial \mathrm{x}_{\mathrm{s}'}}\mathrm{h}_\mathrm{s}^2\,\mathcal{O}\left(\bar{\mathrm{h}}^2\bar{\mathcal{v}}^2+\tilde{\mathrm{h}}^2\tilde{\mathcal{v}}^2\right)$$

$$\left.+\mathcal{O}\left(\bar{\mathrm{h}}^2\bar{\mathcal{v}}^2+\tilde{\mathrm{h}}^2\tilde{\mathcal{v}}^2\right)\mathcal{O}\left(\bar{\mathrm{h}}\bar{\mathcal{v}}-\sum_{\bar{\alpha}=-\mathrm{m}}^{\mathrm{m}}\mathrm{U}_{\bar{\alpha}}\right)^2\right]\left[\mathrm{f}(\tilde{\mathrm{x}})-\sum_{\tilde{\alpha}=-\mathrm{m}}^{\mathrm{m}}\mathrm{U}_{\tilde{\alpha}}\sum_{\mathrm{s}=\mathrm{q}_1+1}^{\mathrm{q}}\frac{\partial \mathrm{f}(\tilde{\mathrm{x}})}{\partial \mathrm{x}_\mathrm{s}}\mathrm{h}_\mathrm{s}\right.$$

$$\left.+\frac{1}{2}\left(\sum_{\tilde{\alpha}=-\mathrm{m}}^{\mathrm{m}}\mathrm{U}_{\tilde{\alpha}}\right)^2\sum_{\mathrm{s}=\mathrm{q}_1+1}^{\mathrm{q}}\frac{\partial^2 \mathrm{f}(\tilde{\mathrm{x}})}{\partial \mathrm{x}_\mathrm{s}\partial \mathrm{x}_{\mathrm{s}'}}\mathrm{h}_\mathrm{s}^2+\mathcal{O}\left(\tilde{\mathrm{h}}\tilde{\mathcal{v}}-\sum_{\tilde{\alpha}=-\mathrm{m}}^{\mathrm{m}}\mathrm{U}_{\tilde{\alpha}}\right)^2\right],$$

substitute the former equations into (A.4), and by Lemma A.2, we finally get

$$\hat{\mathrm{g}}(\mathrm{x})\longrightarrow_\mathrm{p}\left(\left[\mathrm{g}(\mathrm{x})+\mathcal{O}\left(\bar{\mathrm{h}}^2\bar{\mathcal{v}}^2+\tilde{\mathrm{h}}^2\tilde{\mathcal{v}}^2\right)\right]\left[\mathrm{f}(\bar{\mathrm{x}})-\sum_{\bar{\alpha}=-\mathrm{m}}^{\mathrm{m}}\mathrm{U}_{\bar{\alpha}}\sum_{\mathrm{s}=1}^{\mathrm{q}_1}\frac{\partial \mathrm{f}(\bar{\mathrm{x}})}{\partial \mathrm{x}_\mathrm{s}}\mathrm{h}_\mathrm{s}+\frac{1}{2}\left(\sum_{\bar{\alpha}=-\mathrm{m}}^{\mathrm{m}}\mathrm{U}_{\bar{\alpha}}\right)^2\sum_{\mathrm{s}=1}^{\mathrm{q}_1}\frac{\partial^2 \mathrm{f}(\bar{\mathrm{x}})}{\partial \mathrm{x}_\mathrm{s}\partial \mathrm{x}_{\mathrm{s}'}}\mathrm{h}_\mathrm{s}^2\right.\right.$$

$$\left.+\mathcal{O}\left(\bar{\mathrm{h}}\bar{\mathcal{v}}-\sum_{\bar{\alpha}=-\mathrm{m}}^{\mathrm{m}}\mathrm{U}_{\bar{\alpha}}\right)^2\right]\left[\mathrm{f}(\tilde{\mathrm{x}})-\sum_{\tilde{\alpha}=-\mathrm{m}}^{\mathrm{m}}\mathrm{U}_{\tilde{\alpha}}\sum_{\mathrm{s}=\mathrm{q}_1+1}^{\mathrm{q}}\frac{\partial \mathrm{f}(\tilde{\mathrm{x}})}{\partial \mathrm{x}_\mathrm{s}}\mathrm{h}_\mathrm{s}\right.$$

$$\left.+\frac{1}{2}\left(\sum_{\tilde{\alpha}=-\mathrm{m}}^{\mathrm{m}}\mathrm{U}_{\tilde{\alpha}}\right)^2\sum_{\mathrm{s}=\mathrm{q}_1+1}^{\mathrm{q}}\frac{\partial^2 \mathrm{f}(\tilde{\mathrm{x}})}{\partial \mathrm{x}_\mathrm{s}\partial \mathrm{x}_{\mathrm{s}'}}\mathrm{h}_\mathrm{s}^2+\mathcal{O}\left(\tilde{\mathrm{h}}\tilde{\mathcal{v}}-\sum_{\tilde{\alpha}=-\mathrm{m}}^{\mathrm{m}}\mathrm{U}_{\tilde{\alpha}}\right)^2\right]$$

$$+\left[\mathrm{u}(\tilde{\mathrm{x}})+\mathcal{O}\left(\tilde{\mathrm{h}}^2\right)\right]\left[\mathrm{f}(\bar{\mathrm{x}})-\sum_{\bar{\alpha}=-\mathrm{m}}^{\mathrm{m}}\mathrm{U}_{\bar{\alpha}}\sum_{\mathrm{s}=1}^{\mathrm{q}_1}\frac{\partial \mathrm{f}(\bar{\mathrm{x}})}{\partial \mathrm{x}_\mathrm{s}}\mathrm{h}_\mathrm{s}+\frac{1}{2}\left(\sum_{\bar{\alpha}=-\mathrm{m}}^{\mathrm{m}}\mathrm{U}_{\bar{\alpha}}\right)^2\sum_{\mathrm{s}=1}^{\mathrm{q}_1}\frac{\partial^2 \mathrm{f}(\bar{\mathrm{x}})}{\partial \mathrm{x}_\mathrm{s}\partial \mathrm{x}_{\mathrm{s}'}}\mathrm{h}_\mathrm{s}^2\right.$$

$$\left.+\mathcal{O}\left(\bar{\mathrm{h}}\bar{\mathcal{v}}-\sum_{\bar{\alpha}=-\mathrm{m}}^{\mathrm{m}}\mathrm{U}_{\bar{\alpha}}\right)^2\right]$$

$$\cdot\left[f(\tilde{x}) - \sum_{\tilde{\alpha}=-m}^{m} U_{\tilde{\alpha}} \sum_{s=q_1+1}^{q} \frac{\partial f(\tilde{x})}{\partial x_s} h_s + \frac{1}{2}\left(\sum_{\tilde{\alpha}=-m}^{m} U_{\tilde{\alpha}}\right)^2 \sum_{s=q_1+1}^{q} \frac{\partial^2 f(\tilde{x})}{\partial x_s \partial x_{s'}} h_s^2 + \mathcal{O}\left(\tilde{h}\tilde{\varpi} - \sum_{\tilde{\alpha}=-m}^{m} U_{\tilde{\alpha}}\right)^2\right]\Bigg)$$

$$\Bigg/\left[f(\bar{x}) - \sum_{\bar{\alpha}=-m}^{m} U_{\bar{\alpha}} \sum_{s=1}^{q_1} \frac{\partial f(\bar{x})}{\partial x_s} h_s + \frac{1}{2}\left(\sum_{\bar{\alpha}=-m}^{m} U_{\bar{\alpha}}\right)^2 \sum_{s=1}^{q_1} \frac{\partial^2 f(\bar{x})}{\partial x_s \partial x_{s'}} h_s^2 + \mathcal{O}\left(\bar{h}\bar{\varpi} - \sum_{\bar{\alpha}=-m}^{m} U_{\bar{\alpha}}\right)^2\right][f(\tilde{x})$$

$$- \sum_{\tilde{\alpha}=-m}^{m} U_{\tilde{\alpha}} \sum_{s=q_1+1}^{q} \frac{\partial f(\tilde{x})}{\partial x_s} h_s + \frac{1}{2}\left(\sum_{\tilde{\alpha}=-m}^{m} U_{\tilde{\alpha}}\right)^2 \sum_{s=q_1+1}^{q} \frac{\partial^2 f(\tilde{x})}{\partial x_s \partial x_{s'}} h_s^2 + \mathcal{O}\left(\tilde{h}\tilde{\varpi} - \sum_{\tilde{\alpha}=-m}^{m} U_{\tilde{\alpha}}\right)^2\Bigg]$$

$$= g(x) + u(\tilde{x}) + o_p(1)$$
$$= g(x) + \Phi(\tilde{x}) + o_p(1)$$
$$= g(x) + o_p(1).$$

(ii) From $\mathbb{E}(x_i\Phi(\tilde{x}))$ shown in the part (ii) of the proof of Theorem 2.3, by Lemma A.2, we can directly see that

$$\mathbb{E}(x_i\Phi(\tilde{x})) = o_p(1)\ , i = 1,2,\ldots,q_1,$$

substitute it into the expansion of $\hat{\beta}$ shown in the part (ii) of the proof of Theorem 2.3, we then get

$$\hat{\beta}_{LM} \longrightarrow_p \beta + CE(x\Phi(\tilde{x})) = \beta + o_p(1).$$

(iii) Similarly, from the asymptotic expansions of the partial linear model's estimation in the end of section 2, we can get

$$\hat{\beta}_{PLM} \longrightarrow_p \beta + \mathcal{C}_0\big(o_p(1) - o_p(1) - o_p(1) + o_p(1)\big) = \beta + o_p(1).$$

Hence the entire theorem is proved. □

ONLINE APPENDIX B: MONTE CARLO SIMULATIONS

In the online appendix B, we will illustrate the Monte Carlo simulation results of section 2 which is omitted from the main text. In order to study and compare the large sample and small sample estimation performances of the parametric and nonparametric regression models (2.3), (2.4) in section 2, a Monte Carlo simulation study is carried out. We first consider the simple DGPs as follows

$$\mathrm{DGP1}: y = x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + x_4\beta_4 + \varepsilon_1, \tag{B.1}$$

$$\mathrm{DGP2}: y = \sin(x_1)\beta_1 + x_2^2\beta_2 + x_3\beta_3 + x_4\beta_4 + \varepsilon_2, \tag{B.2}$$

where all the variables x_i are drawn from the uniform distributions, i=1,2,3,4. DGP1 simulates a linear process, and $\beta_1 = 1, \beta_2 = 2, \beta_3 = 3, \beta_4 = 4, \varepsilon_1 \sim \mathbb{N}(0,1)$; while DP2 simulates a nonlinear process, and $\beta_1 = 1, \beta_2 = 1, \beta_3 = 3, \beta_4 = 1, \varepsilon_2 \sim \mathbb{N}(0,1)$. $n = 50,500,1000$ respectively. We use the correlations among the variables $x = (x_1, x_2, x_3, x_4)'$ to describe the degree of model endogeneity, the correlation matrices ρ are

$$\begin{pmatrix} 1 & 0.2 & 0.21 & 0.23 \\ 0.2 & 1 & 0.22 & 0.21 \\ 0.21 & 0.22 & 1 & 0.22 \\ 0.23 & 0.21 & 0.22 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 & 0.57 & 0.56 \\ 0.5 & 1 & 0.22 & 0.2 \\ 0.57 & 0.22 & 1 & 0.22 \\ 0.56 & 0.2 & 0.22 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0.8 & 0.57 & 0.61 \\ 0.8 & 1 & 0.22 & 0.21 \\ 0.57 & 0.22 & 1 & 0.22 \\ 0.61 & 0.2 & 0.22 & 1 \end{pmatrix}$$

representing the weak endogeneity, median endogeneity and strong endogeneity respectively, the i's row j's column of the matrices ρ is the correlation coefficient between the variable x_i and the variable x_j, $i \neq j = 1,2,3,4$ (see Florens et al., 2012).

The regression biases are measured and computed by the distance metric between the sampling distribution of the estimators when there are no variables omitted and the sampling distribution of the estimators when there are variables omitted. As for the parametric estimation, what we are concerned is the estimation bias of $\widehat{\beta}_1$, the sampling distribution of $\widehat{\beta}_1$ is generated by the bootstrap method (resampling 50 times out of n); while as for the nonparametric estimation, what we are concerned is the bias of $\hat{y} = \hat{g}(x)$, and the sampling distribution of $\hat{g}(x)$ is replaced by $\hat{y}$ using bootstrapping methods. We use the entropy-based metric proposed by Granger et.al (2001), Maasoumi and Racine (2002), Li qi et.al (2009) and Hayfield and Racine (2008) to measure the estimation biases caused by omitting variables

$$\begin{aligned} \text{Bias}: \mathcal{S}_\rho &= \frac{1}{2}\int \left(f_1^{1/2} - f_2^{1/2}\right)^2 d\beta_1 \\ &= \frac{1}{2}\int \left(1 - \frac{f_1^{1/2}}{f_2^{1/2}}\right)^2 d\mathcal{F}_1(\beta), \end{aligned} \tag{B.3}$$

where f_1, f_2 are the marginal densities of the estimator $\widehat{\beta}_1$ when there are no variables omitted and there are variables omitted respectively. The second expression is in a moment form which is often replaced with a sample average. Similarly, we can measure and compute the bias of the nonparametric estimator $\hat{y}$. $0 \le S_\rho \le 1$, so a bigger value of $\mathcal{S}_\rho$ means a larger distance between the distributions f_1, f_2, and hence a larger biases between the estimators. Under the same model endogenous degree, if the nonparametric estimator's metric is smaller than the parametric estimator's metric, then the bias of the nonparametric estimation is smaller than the parametric OLS estimation. Apart from this, the models' fitness are also compared, we use the method proposed by Hayfield and Racine (2008)

$$\mathcal{R}^2 = \frac{\left(\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{y})\right)^2}{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (\hat{y}_i - \bar{y})^2} \tag{B.4}$$

to measure the goodness-of-fit of the nonparametric estimation, where $\hat{y}_i$ is the nonparametric prediction and $\bar{y}$ is the sample mean.[1] Hence, there exists comparability between the goodness-of-fit of

[1] By Schwarz-Cauchy inequality, we are able to show that this measure of $\mathcal{R}^2$ will always lie in the range [0,1], and it is identical to the standard measure computed as $\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 / \sum_{i=1}^n (y_i - \bar{y})^2$ when the model is linear,

the parametric and nonparametric estimations. The Root-Mean-Squared-error (Rmse) of the models is defined as $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}$, see e.g., Hall et al. (2007). The OLS and the local constant estimation results of the DGPs (5.1), (5.2) with sample size $N = 50{,}500{,}1000$ respectively are given then. From Tables C.1, C.2 and C.3 provided in the online Appendix C, we make the following observations.

1. Under the same degree of endogeneity, the asymptotic biases of the parametric and nonparametric estimations are all related with the numbers of the variables omitted whether the true DGP is linear or nonlinear. The more variables omitted, the larger asymptotic biases will be, and the asymptotic bias of the nonparametric estimation will be smaller than that of the parametric estimation when certain explanatory variables are omitted.
2. When the same explanatory variables are omitted, the asymptotic biases of the parametric and nonparametric estimations are all related with the degree of model endogeneity (the correlations among the included variables and the omitted variables) whether the real DGP is linear or nonlinear. The larger degree of model endogeneity, the larger asymptotic biases will be, and the asymptotic bias of the nonparametric estimation will be smaller than that of the parametric estimation under certain degree of endogeneity.
3. The goodness-of-fits and the precisions of predictions are all related with the degree of model endogeneity and the numbers of variables omitted whether the real DGP is linear or not. The more variables omitted and the larger degree of model endogeneity, the worse the goodness-of-fits and precisions will be. The goodness-of-fits and precisions of predictions of the nonparametric estimations are always better than the parametric estimations under certain degree of endogeneity.

These above results imply that when there are explanatory variables omitted from regression models, the asymptotic bias of the nonparametric estimation will be smaller than the asymptotic bias of the parametric OLS estimation when the sample size is large enough and the included variables are finite. Our simulations support the results of Theorem 2.3, 2.4.

We then investigate the asymptotic performances of the partial linear model. The DGP considered for the partial liner model is as follows

$$y = x_1\beta_1 + g(x_2,x_3,x_4) + \varepsilon_3, \tag{B.5}$$

where all the variables x_i are drawn from uniform distributions, $\beta_1 = 1$, $\varepsilon_3 \sim \mathbb{N}(0,1)$, and the DGP of the parametric linear model is (5.1) as usual. The degree of model endogeneity is defined in section 5.1, while the estimation biases of the parameter β_1 are defined as $|\hat{\beta}_{LM} - \beta|$ and $|\hat{\beta}_{PLM} - \beta|$ for the linear model and the partial linear model respectively.[2] What we are concerned now is the comparisons between these two estimators, the simulation results are then given in Table A.4. From Table C.4 provided in the online Appendix C, we find that

1. The comparison results between the linear model's estimation and the partial linear model's estimation are similar to the results 1-3 shown in section 5.1. Under the same degree of model endogeneity and when the same variables are omitted, the asymptotic bias of the partial linear model is smaller than the asymptotic bias of the linear model when the sample size is large enough.
2. When the degree of model endogeneity is increasing, the absolute distance between the

fitted with least squares and includes an intercept term.

[2] One can also use the entropy-based metric method used in section 5.1, which is equivalent to the method of this section, to measure the biases of the linear and partial linear models' estimations. However, the absolute distance metric used in this section is more suitable to this situation and much more convenient and intuitionistic.

asymptotic bias of the linear model and the asymptotic bias of the partial linear model is also getting larger whether the real DGP is linear or nonlinear. These hence imply that when the degree of model endogeneity is increasing, the asymptotic bias of the partial linear model will be much smaller than the asymptotic bias of the linear model if certain explanatory variables are omitted.

3. The asymptotic bias of the partial linear model is not uniformly smaller than the asymptotic bias of the linear model when the degree of model endogeneity is weak; while uniformly smaller when the degree of model endogeneity is stronger.

These conclusions show that the partial linear model's estimation owns a smaller asymptotic bias than the linear model's estimation when there are several variables omitted and the sample size is large enough. Our simulations support the results at the end of section 2 and Theorem 2.5.

Following Florens et.al (2012), we next use the method and DGP in their paper to compute and compare the OLS estimator and the IV estimator of the linear and partial linear models with $\mathrm{N} = 100,250,500$ respectively when there are endogeneities in the models. The results are shown in Table C.5 provided in the online appendix C. Let us now consider a univariate endogenous regression model $y = \beta x + \epsilon, \mathrm{Cov}(x,\epsilon) \neq 0$, by Cameron and Trivedi (2005), Godffrey (1999) etc., we are able to show that $\widehat{\beta}_{LM-IV} = (z'xx'y)^{-1}(z'yx'x)\widehat{\beta}_{LM}$ where z is the instrumental variable. Set $\mathbb{E}(\widehat{\beta}_{LM}) - \beta \triangleq \Delta$ to denote the small sample bias, then we can get $\mathbb{E}(\widehat{\beta}_{LM-IV}) - \beta = \mathrm{E}\left(\left(z'xx'y\right)^{-1}\left(z'yx'x\right) - 1\right)\widehat{\beta}_{LM} + \Delta$.

This implies that, in an endogenous linear model, if the small sample estimation bias of the OLS is getting larger, then the small sample estimation bias of the IV is also getting larger. In this paper we find this conclusion also holds true for the partial linear models. As shown in Table C.5, the absolute bias of the partial linear model is smaller than that of the linear model, hence the absolute IV estimation bias of the partial linear model is also smaller than the linear model when there exist endogeneities; and the convergent rate to the true parameter value of the partial linear model's IV estimation is faster than that of the linear model's.

REFERENCES

Florens, J. P., Johannes, J., and S. V. Bellegem (2012). Instrumental regression in partially linear models. Econometrics Journal 15, 304–324.

Granger, C. W., Maasoumi, E. and J. Racine (2001). A metric entropy and a new test of independence. UCSD Working paper, University of California, San Diego.

Hall, P., Li, Q. and J. S. Racine (2007). Nonparametric estimation of regression functions in the presence of irrelevant regressors. Review of Economics & Statistics 89, 784-789.

Li, Q., Maasoumi, E., and J. S. Racine (2009). A nonparametric test for equality of distributions with mixed categorical and continuous data. Journal of Econometrics 148, 186-200.

Hayfield, T and J.S. Racine (2008). Nonparametric econometrics: the np package. Journal of Statistical Software 27, 2008.

ONLINE APPENDIX C: TABLES

Table C.1	Stron	Nonli	Bias	0.349	0.862	0.989	Bias	0.746	0.832	0.921

Table

Stron																		
Nonli		Weak endogeneity						Moderate endogeneity										
Bias		Linear(DGP1)			Nonlinear(DGP2)			Linear(DGP1)			Nonlinear(DGP2)			Linear(DGP1)				
	LS	R^2	Rmse	Bias	R^2	Rmse	Bias	R^2	Rmse	Bias	R^2	Rmse	Bias	R^2	Rmse	Bias	R^2	Rmse
0.364	Omitted: 0	0.972	1.134		0.892	2.184		0.984	1.046		0.936	1.566		0.984	1.043		0.949	1.552
0.884	Omitted: 1	0.644	4.028	0.415	0.829	2,717	0.115	0.774	4.187	0.950	0.933	1.587	0.103	0.938	2.014	1	0.946	1.571
0.996	Omitted: 2	0.421	5.077	0.880	0.670	3.731	0.647	0.688	4.574	0.972	0.787	2.794	0.984	0.936	2.022	1	0.896	2.161
Bias	Omitted: 3	0.284	5.591	0.874	0.059	6.235	0.577	0.641	4.854	0.989	0.462	4.390	0.999	0.906	2.433	1	0.872	2.373
	NW	R^2	Rmse	Bias	R^2	Rmse	Bias	R^2	Rmse	Bias	R^2	Rmse	Bias	R^2	Rmse	Bias	R^2	Rmse
0.136	Omitted: 0	0.991	0.627		0.997	0.364		0.986	1.003		0.987	0.697		0.992	0.686		0.991	0.614
0.663	Omitted: 1	0.907	2.013	0.008	0.994	0.498	0.015	0.791	3.761	0.182	0.979	0.852	0.113	0.952	1.707	0.681	0.991	0.615
0.892	Omitted: 2	0.519	4.536	0.219	0.789	2.924	0.032	0.742	4.139	0.269	0.802	2.722	0.473	0.946	1.813	0.790	0.936	1.673
	Omitted: 3	0.327	5.562	0.277	0.054	6.271	0.806	0.651	4.792	0.397	0.549	3.962	0.822	0.929	2.082	0.899	0.894	2.136

Table

	Weak endogeneity						Moderate endogeneity											Stron
	Linear(DGP1)			Nonlinear(DGP2)			Linear(DGP1)			Nonlinear(DGP2)			Linear(DGP1)					Nonli
LS	R^2	Rmse	Bias	R^2	Rmse	Bias	R^2	Rmse	Bias	R^2	Rmse	Bias	R^2	Rmse	Bias	R^2	Rmse	Bias
Omitted: 0	0.975	1.035		0.941	1.445		0.979	1.067		0.924	1.651		0.981	0.962		0.933	1.583	
Omitted: 1	0.644	3.873	0.381	0.916	1.716	0.045	0.749	3.510	0.884	0.903	1.870	0.153	0.927	1.982	0.999	0.927	1.603	0.303
Omitted: 2	0.385	5.082	0.557	0.673	3.391	0.294	0.676	3.980	0.999	0.734	3.091	0.844	0.926	2.042	1	0.832	2.432	0.673
Omitted: 3	0.169	5.901	0.684	0.039	5.802	0.568	0.666	4.042	0.999	0.458	4.409	0.999	0.844	2.796	1	0.796	2.675	0.868
NW	R^2	Rmse	Bias	R^2	Rmse	Bias	R^2	Rmse	Bias	R^2	Rmse	Bias	R^2	Rmse	Bias	R^2	Rmse	Bias
Omitted: 0	0.987	0.774		0.979	0.865		0.983	0.957		0.981	0.827		0.985	0.890		0.978	0.901	
Omitted: 1	0.725	3.441	0.043	0.952	1.312	0.022	0.769	3.389	0.096	0.966	1.109	0.095	0.932	1.849	0.322	0.975	0.941	0.247
Omitted: 2	0.425	4.951	0.056	0.704	3.929	0.301	0.722	3.682	0.562	0.821	2.546	0.572	0.929	1.861	0.783	0.894	1.927	0.586
Omitted: 3	0.180	5.861	0.329	0.057	5.748	0.542	0.685	3.916	0.784	0.474	4.339	0.778	0.847	2.746	0.914	0.807	2.598	0.778

	Weak endogeneity						Moderate endogeneity										
	Linear(DGP1)			Nonlinear(DGP2)			Linear(DGP1)			Nonlinear(DGP2)			Linear(DGP1)				
LS	R^2	Rmse	Bias	R^2	Rmse	Bias	R^2	Rmse	Bias	R^2	Rmse	Bias	R^2	Rmse	Bias	R^2	Rmse
Omitted: 0	0.979	0.993		0.929	1.707		0.981	0.993		0.932	1.578		0.980	1.013		0.924	1.698
Omitted: 1	0.662	3.999	0.556	0.899	1.976	0.070	0.786	3.302	0.827	0.917	1.746	0.215	0.920	2.042	0.999	0.920	1.74
Omitted: 2	0.438	5.154	0.724	0.669	3.577	0.299	0.714	3.818	0.995	0.769	2.912	0.771	0.916	2.101	1	0.842	2.448
Omitted: 3	0.226	6.047	0.803	0.061	6.022	0.516	0.687	3.989	0.999	0.469	4.408	0.987	0.839	2.903	0.999	0.821	2.607
NW	R^2	Rmse	Bias	R^2	Rmse	Bias	R^2	Rmse	Bias	R^2	Rmse	Bias	R^2	Rmse	Bias	R^2	Rmse
Omitted: 0	0.986	0.843		0.979	0.932		0.986	0.875		0.981	0.832		0.985	0.922		0.976	0.967
Omitted: 1	0.699	3.815	0.067	0.949	1.414	0.053	0.807	3.159	0.366	0.966	1.131	0.133	0.941	1.763	0.664	0.972	1.043
Omitted: 2	0.461	5.068	0.154	0.723	3.279	0.125	0.730	3.711	0.465	0.828	2.524	0.389	0.925	1.989	0.784	0.892	2.033
Omitted: 3	0.229	6.039	0.351	0.068	5.999	0.348	0.693	3.948	0.686	0.475	4.386	0.557	0.843	2.869	0.912	0.839	2.477

Table C.4. Finite sample biases of the linear and partial linear estimators with omitted variables.

		Weak endogeneity		Moderate endogeneity		Strong endogeneity	
		Linear	Nonlinear	Linear	Nonlinear	Linear	Nonlinear
N=50	Linear model	Bias	Bias	Bias	Bias	Bias	Bias
	Omitted: 1	1.0659	0.9615	2.0254	1.7797	6.1255	4.6629
	Omitted: 2	1.3299	1.5392	3.1514	3.7747	4.6495	5.3690
	Partial linear model	Bias	Bias	Bias	Bias	Bias	Bias
	Omitted: 1	1.0030	0.8819	1.4035	1.7529	4.7517	3.9157
	Omitted: 2	1.2358	1.4254	2.2827	3.7802	3.2262	4.7217
N=500	Linear model	Bias	Bias	Bias	Bias	Bias	Bias
	Omitted: 1	0.3124	0.3122	2.2645	2.4012	6.3112	6.6131
	Omitted: 2	0.6914	1,0820	3,1145	4.6536	4.2922	7.2762
	Partial linear model	Bias	Bias	Bias	Bias	Bias	Bias
	Omitted: 1	0.3153	0.3173	2.2329	2.3969	6.1883	6.5867
	Omitted: 2	0.6934	0.8735	3.0904	4.6685	4.1572	7.2550
N=1000	Linear model	Bias	Bias	Bias	Bias	Bias	Bias
	Omitted: 1	0.5787	0.5816	2.0965	2.0865	6.4468	6.0783
	Omitted: 2	0.8297	1.0744	3.1655	4.4086	4.1018	6.4189
	Partial linear model	Bias	Bias	Bias	Bias	Bias	Bias
	Omitted: 1	0.5841	0.5882	2.1003	2.0964	6.3600	6.0340
	Omitted: 2	1.8321	1.0733	3.1859	4.4327	4.0227	6.3869

Table C.5. Estimates of the endogenous linear and partial linear models.

	$\mathrm{N} = 100$	$\mathrm{N} = 250$	$\mathrm{N} = 500$
Estimators	Estimates	Estimates	Estimates
$\widehat{\beta}_{LM}$	0.9347(0.0653)	0.8572(0.1428)	1.0372(0.0372)
$\widehat{\beta}_{PLM}$	0.9719(0.0281)	0.9797(0.0203)	0.9886(0.0114)
$\widehat{\beta}_{LM-IV}$	1.0203(0.0203)	0..8302(0.1698)	1.1267(0.1267)
$\widehat{\beta}_{PLM-IV}$	0.9808(0.0192)	0.9939(0.0061)	1.0005(0.0005)

Note: The true value of β is 1. LM, PLM represents the OLS estimates of the linear model and the Robinson estimator of the partial linear model respectively; $LM - IV$, $PLM - IV$ represents the instrumental estimation of the linear and partial linear models respectively. The absolute biases between the estimator and the true value are shown in brackets.