THREE ESSAYS ON NONPARAMETRIC AND SEMIPARAMETRIC METHODS AND THEIR APPLICATIONS

A Dissertation

by

CARL DAVID AUGUST GREEN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Qi Li
Committee Members,	Li Gan
	Mark Hoekstra
	Lori Taylor
Head of Department,	Timothy Gronberg

May 2015

Major Subject: Economics

ABSTRACT

This dissertation contains three essays on nonparametric and semiparametric regression methods.

In the first essay, we consider the problem of nonparametric regression with mixed discrete and continuous covariates using the k-nearest neighbor (k-nn) method. We derive the asymptotic normality of the proposed estimator and use Monte Carlo simulations to demonstrate its finite sample performance. We apply the method to estimate corn yields in Iowa as a function of agricultural district, temperature, and precipitation.

In the second essay, we consider the problem of testing error serial correlation in fixed effects panel data models in a nonparametric framework. We show that our test statistic has a standard normal distribution under the null hypothesis of zero serial correlation. The test statistic diverges to infinity at the rate of $\sqrt{N}$ under the alternative hypothesis that errors are serially correlated, where N is the cross-sectional sample size. We propose a bootstrap version of the test which we show to perform well in finite sample applications.

In the third essay, we consider estimation of varying-coefficient single-index models with an endogenous regressor. We propose a multi-step instrumental variables procedure to estimate the coefficient function and the corresponding index parameters. We prove the consistency of the estimators, and we present Monte Carlo simulations demonstrating their finite sample performance. We then apply the proposed method to examine the determinants of aggregate illiquidity in the U.S. stock market.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Qi Li, my committee members, Drs. Li Gan, Mark Hoekstra, and Lori Taylor, and finally Drs. Ke-Li Xu and Ximing Wu, for their support and guidance. I would like to thank Dr. Yu Yvette Zhang, Dr. Cheng Hsiao, Wei Long, and Dr. Zhongjian Lin for joint work included in this dissertation. Finally, I would like to thank the Economics Department for its support.

.

TABLE OF CONTENTS

Page

ABSTRACT . . . ii

ACKNOWLEDGEMENTS . . . iii

TABLE OF CONTENTS . . . iv

LIST OF FIGURES . . . vi

LIST OF TABLES . . . vii

1. INTRODUCTION AND SUMMARY . . . 1

2. ESTIMATION OF NONPARAMETRIC REGRESSION MODELS WITH MIXED DISCRETE AND CONTINUOUS COVARIATES BY THE K-NN METHOD, WITH QI LI AND YU YVETTE ZHANG . . . 5

   2.1 Introduction . . . 5
   2.2 K-nn Estimation and Cross-Validation . . . 7
   2.3 Monte Carlo Simulations . . . 11
   2.4 Empirical Application . . . 14
   2.5 Conclusion . . . 15

3. TESTING ERROR SERIAL CORRELATION IN FIXED EFFECTS NONPARAMETRIC PANEL DATA MODELS, WITH WEI LONG AND CHENG HSIAO . . . 17

   3.1 The nonparametric fixed effects panel data model . . . 19
   3.2 A bootstrap test procedure . . . 24
   3.3 Monte Carlo simulations . . . 25

4. VARYING-COEFFICIENT SINGLE-INDEX MODELS WITH ENDOGENEITY: THEORY AND APPLICATION, WITH ZHONGJIAN LIN . . . 28

   4.1 Introduction . . . 28
   4.2 The Varying Coefficient Single Index Model . . . 30
   4.3 The Estimation Method . . . 32
   4.4 Monte Carlo Simulations . . . 37

4.5 An Empirical Application 39
4.5.1 Model, Variable Description, and Data 43
4.5.2 Empirical Results 44
4.6 Conclusion 46

5. CONCLUSION 48

REFERENCES 49

APPENDIX A. APPENDIX TO SECTION 2 56

APPENDIX B. APPENDIX TO SECTION 3 78

B.1 Proof of Theorem 1 78
B.2 Appendix B: Proofs of (7) to (9) 81
B.3 Results of Monte Carlo Simulations 91

APPENDIX C. APPENDIX TO SECTION 4: PROOFS OF THEOREMS . 95

LIST OF FIGURES

FIGURE Page

2.1 Plot of $x/(x^4+1)$ 13

2.2 Estimated marginal effects 16

4.1 Estimated price impact 47

4.2 Price impact estimated using instrument (solid) and treating X_{it} as exogenous (dashed) 47

LIST OF TABLES

TABLE Page

2.1 Summary of simulation results 13

4.1 Mean squared errors for the four specifications for $\beta(\cdot)$. 38

B.1 Rejection frequency for DGP 1,2,3 under case (i): i.i.d error (estimated size). 91

B.2 Rejection frequency for DGP 1,2,3 under case (ii): AR(1) (estimated power). 92

B.3 Rejection frequency for DGP 1,2,3 under case (iii): MA(1) (estimated power). 93

B.4 Rejection frequency for DGP 2 under case (i)',(ii)' and (iii)'. 94

B.5 Rejection frequency for DGP 2 under case (i)',(ii)' and (iii)': an asymptotic test. 94

1. INTRODUCTION AND SUMMARY

Nonparametric regression methods have the advantage that they do not impose strong restrictions on the structure of the relationship between the dependent and explanatory variables, instead allowing this structure to be revealed by the data. Semiparametric regression methods offer some of the same flexibility, but often with the advantages of lower dimensionality and improved interpretability. Since researchers often have little prior information on the relationships being studied, this flexibility of nonparametric and semiparametric methods is often of great value. It is therefore not surprising that these methods have received much recent attention from econometricians and statisticians. This dissertation adds to the literature on nonparametric and semiparametric regression methods in the following ways.

In the first study, we adapt the k-nearest neighbor (k-nn) method for nonparametric regression to the case of mixed continuous and discrete explanatory variables.

Under the k-nn method, regression estimates are driven by the same number of observations at every point in the range of the explanatory variables, no matter how dense or sparse the data around the point. Thus, unlike fixed-bandwidth kernel methods, the k-nn method does not suffer from the problem of estimates assigning undue weight to single observations in areas where the data are particularly sparse. Researchers may therefore prefer the k-nn method when the explanatory variables are distributed unevenly over their range.

Ouyang, Li, and Li (2006) study the k-nn method for the case in which all explanatory variables are continuous. We generalize their results to our proposed method, which admits not only continuous, but also discrete explanatory variables. Our proposed method smooths both types of explanatory variable – the continuous variables by the k-nn method, and the discrete variables as in Aitchison and Aitken (1976) and Racine and Li (2004). Just like smoothing of continuous covariates, smoothing of discrete covariates introduces bias in finite samples; however, it may substantially decrease estimation variance, and thus lead to more accurate estimates.

We consider the selection of smoothing parameters by least-squares cross-validation, and we derive the probability order of the parameters thus selected. We then derive the asymptotic normality of our regression estimator that uses these cross-validated smoothing parameters. We provide Monte Carlo simulations demonstrating the finite sample performance of our estimators, and we apply our method to investigate how corn yields in Iowa depend upon agricultural district, temperature, and precipitation.

In the second study, we develop a test for error serial correlation in fixed effects nonparametric panel data models.

It is important to test for error serial correlation in these (and other) models, for at least three reasons. First, in the presence of error serial correlation researchers need to use methods of estimating standard errors that account for this autocorelation. While in many settings such methods have become the default choice for applied researchers, in other settings researchers may prefer to avoid them due to their high computational cost or their inferior performance in the absence of error autocorrelation. Therefore, researchers may wish to first test whether robust standard errors are indeed necessary before using them. Second, if errors are serially correlated it may be possible to obtain more efficient estimators by taking this autocorrelation into account. Third, strong serial correlation in errors often indicates that some important explanatory variables are missing from the model.

Li and Hsiao (1998) propose a test for zero error serial correlation in a nonparametric model; we generalize this test to the case of a fixed effects nonparametric panel data model. We propose a test statistic, and we derive its asymptotic normal distribution under the null hypothesis of zero error serial correlation. However, our simulations show that this asymptotic distribution is a poor approximation in finite samples; we therefore introduce a bootstrap test procedure, which simulations show to perform well in finite samples.

In the third study, we consider estimation of varying-coefficient single-index models with an endogenous explanatory variable.

Varying-coefficient models are increasingly popular, and they have some particularly attractive properties in the case where some explanatory variables are endogenous. In the presence of endogeneity, estimation of fully nonparametric models may become difficult or even impossible, but estimation of varying-coefficient models often remains straightforward. Provided that all endogenous variables enter the model linearly, instrumental variable estimation methods are available; see for example the method of Cai, Xiong, and Wu (2006).

Our model is similar to that considered by Cai, Xiong, and Wu, but we restrict the coefficient functions to follow a single-index specification in order to reduce the dimensionality of the model. Xia and Li (1999) and Fan et al. (2003) consider the estimation of such varying-coefficient single-index models in the case where all explanatory variables are exogenous; to the best of our knowledge, ours is the first paper to consider varying-coefficient single-index models with endogenous explanatory variables. For expositional simplicity, we restrict our analysis to the case of a single endogenous explanatory variable.

We propose a multi-step instrumental variable procedure for estimating our model. We derive the $\sqrt{n}$-consistency of our index parameter estimators, and we show that our coefficient function estimator converges to its true value at the standard rate for single-index models. We provide Monte Carlo simulations demonstrating the finite sample performance of our estimators.

Finally, we apply our estimation procedure to analyze the determinants of aggregate liquidity in the U.S. stock market. We use a measure of illiquidity based on the "price impact" of trades – the coefficient in a pooled regression of stock returns on stock-specific order flow (i.e., on signed trading volume). We argue that the price impact of interest is that of an uninformed trade; thus, as order flow consists of informed as well as uninformed trades, we argue that it should be viewed as endogenous. This endogeneity can be addressed by an instrumental variables approach, with order flow (or rather our proxy of order flow) instrumented by a component of order flow that consists solely of uninformed trades. Inspired by Coval and Stafford (2007), we identify as such a component those trades made by mutual funds with extreme inflows or outflows of funds. Using this instrumental variable and the methods developed in our paper, we estimate price impact and examine how it varies according to two interest rate variables.

2. ESTIMATION OF NONPARAMETRIC REGRESSION MODELS WITH MIXED DISCRETE AND CONTINUOUS COVARIATES BY THE K-NN METHOD, WITH QI LI AND YU YVETTE ZHANG

2.1 Introduction

Nonparametric methods of conditional mean estimation offer flexibility, imposing only relatively weak assumptions on the form of the conditional mean function. These methods can provide consistent estimators in situations where parametric estimators are biased even asymptotically. Thus, researchers may prefer nonparametric methods to parametric alternatives when they are reluctant to make assumptions about the form of the conditional mean function. However, in finite samples nonparametric methods require smoothing in order to balance bias against variance, and their performance depends critically on the degree of smoothing that they entail. Therefore, the choice of the parameters controlling this degree of smoothing is an important subject of study. Data-driven methods for selecting smoothing parameters have been proposed; see, for example, Härdle, Hall, and Marron (1988,1992), and Gao and Tong (2004) and the references therein. Of particular relevance to our paper, Racine and Li (2004) study data-driven smoothing parameter selection, and the asymptotic behavior of the corresponding estimator, for the kernel method introduced by Aitchison and Aitken (1976). This method allows for both discrete and continuous explanatory variable, and it smooths both types of variable. In this paper, we consider a method that allows for continuous and discrete explanatory variables, and smooths the discrete variables as in Aitchison and Aitken (1976) and Racine and Li (2004), but smooths the continuous variables using the k-nearest neighbor method rather than fixed-bandwidth methods. We study the selection of smoothing parameters by least-squares cross-validation, and the asymptotic behavior of the corresponding conditional mean function estimator.

The k-nearest neighbor method for nonparametric regression has the advantage that, in cases where the explanatory variables are unevenly distributed over their range, it automatically adjusts the size of the bandwidth to account for the relative density or sparseness of the data. It is generally best to decrease the size of the bandwidth where the data are dense and increase the size of the bandwidth where the data are sparse; by always using the k nearest observations, the k-nearest neighbor method makes such an adjustment automatically. This may be of particular help in areas where the data are sparse, as it ensures that even here estimates are driven by a sufficient number of observations, with no single observation receiving undue weight.

Ouyang, Li, and Li (2006) study the k-nearest neighbor method for nonparametric regression for the case in which all explanatory variables are continuous. They derive the probability order of smoothing parameters selected by least-squares cross-validation, and the asymptotic normal distribution of the corresponding conditional mean function estimators. We generalize their results to the case of mixed continuous and discrete explanatory variables.

The remainder of this paper is organized as follows: In section 2.2 we describe our proposed method, which smooths the continuous explanatory variables using the k-nearest neighbor method and smooths the discrete explanatory variables as in Aitchison and Aitkin (1976) and Racine and Li (2004). We examine the asymptotic behavior of smoothing parameters selected by least-squares cross-validation. We show that these smoothing parameters are asymptotically equivalent to non-stochastic smoothing parameters that minimize a weighted mean square error. We then derive the asymptotic normal distribution of the corresponding conditional mean function estimator. In section 2.3 we present Monte Carlo simulations that demonstrate the good finite sample performance of the proposed estimator. In section 2.4 we use our method to estimate the conditional mean of corn yield in Iowa as a function of agricultural district, average annual temperature and precipitation. Section 2.5 concludes the paper. All proofs are relegated to the appendix.

2.2 K-nn Estimation and Cross-Validation

We consider a nonparametric regression model with both discrete and continuous regressors. We write the vector of regressors X_i as $X_i = (X_i^c, X_i^d)$, where X_i^c is a $q \times 1$ vector of continuous regressors and X_i^d is an $m \times 1$ vector of discrete regressors. We let $\mathcal{D} \subset \mathcal{R}^m$ denote the range of X_i^d. We consider the nonparametric regression model

$$Y_i = g(X_i) + u_i, \qquad i = 1, 2, ..., n \tag{2.1}$$

where the functional form of $g(\cdot)$ is not specified. We assume that the data $(Y_i, X_i)_{i=1}^n$ are independent and identically distributed.

We use the k-nearest neighbor method to smooth the continuous regressors, and we smooth the discrete regressors in the spirit of Racine and Li (2004) and Hall, Li, and Racine (2007). For the discrete regressors, we first define a univariate kernel function given by

$$l(X_{s,i}^d, x_s^d, \lambda_s) = \begin{cases} 1 & \text{if } X_{s,i}^d = x_s^d, \\ \lambda_s & \text{if } X_{s,i}^d \neq x_s^d \end{cases} \tag{2.2}$$

where the range of λ_s is $[0, 1]$. The product kernel is then defined by $L(X_i^d, x^d, \lambda) = \prod_{s=1}^m l(X_{s,i}^d, x_s^d, \lambda_s)$.

Next we consider the continuous regressors. Given a vector $x^c \in \mathcal{R}^q$, we let $D_n(x^c, k)$ denote the distance between x^c and its k^{th}-nearest neighbor. We then define

$$R_x = D_n(x^c, k) \equiv \text{Euclidean distance between } x^c \text{ and its } k^{th}\text{-nearest neighbor among } \{X_j^c\}_{j=1}^n. \tag{2.3}$$

We let R_i denote the distance between X_i^c and its k^{th}-nearest neighbor, i.e.

$$R_i = D_n(X_i^c, k) \equiv \text{Euclidean distance between } X_i^c \text{ and its } k^{th}\text{-nearest neighbor among } \{X_j^c\}_{j\neq i}. \tag{2.4}$$

Next, let $w(\cdot) : \mathcal{R}^q \to \mathcal{R}$ be a bounded non-negative weight function satisfying $w(v) = w(-v)$, $\int w(v)dv = 1$, and $w(v) = 0$ for $\|v\| \geq 1$, where $\|\cdot\|$ denotes the Euclidean norm. The k-nearest neighbor estimator of $f(x)$, the density function of X_i, is given by

$$\hat{f}(x) = \frac{1}{nR_x^q} \sum_{i=1}^n w\left(\frac{X_i^c - x^c}{R_x}\right) L(X_i^d, x^d, \lambda). \tag{2.5}$$

Then the local constant k-nearest neighbor estimator of $g(x)$ is given by

$$\hat{g}(x) = \frac{1}{nR_x^q} \sum_{i=1}^n Y_i \, w\left(\frac{X_i^c - x^c}{R_x}\right) L(X_i^d, x^d, \lambda)/\hat{f}(x) \tag{2.6}$$

We consider the selection of k and λ by leave-one-out least squares cross-validation. Specifically, we choose $\hat{k}$ and $\hat{\lambda}$ to minimize

$$CV(k, \lambda) = \sum_{i=1}^n \left(Y_i - \hat{g}_{-i}(X_i)\right)^2 M(X_i) \tag{2.7}$$

where $\hat{g}_{-i}(X_i)$ is the leave-one-out k-nearest neighbor estimator of $g(X_i)$, given by

$$\hat{g}_{-i}(X_i) = \sum_{j \neq i} Y_j \, w\left(\frac{X_j^c - X_i^c}{R_i}\right) L(X_j^d, X_i^d, \lambda)/\hat{f}_{-i}(X_i),$$

and

$$\hat{f}_{-i} = \sum_{j \neq i} w\left(\frac{X_j^c - X_i^c}{R_i}\right) L(X_j^d, X_i^d, \lambda)$$

is the leave-one-out estimator of $f(X_i)$; and $M(X_i)$ is a non-negative weight function that trims away data near the boundary of the data support.

Before analyzing the asymptotic properties of the cross validation selected smoothing parameters, we first list some regularity conditions.

Assumption 1 (i) $(X_i, Y_i)_{i=1}^n$ are i.i.d as (X, Y); (ii) $u_i \equiv Y_i - g(X_i)$ has zero mean and finite fourth moment; (iii) $g(\cdot, x^d)$ and $f(\cdot, x^d)$ are both continuously differentiable up to the fourth order for all $x^d \in D$; (iv) defining $\sigma^2(x) = E(u_i^2 | X_i = x)$, $\sigma^2(\cdot)$ is continuous in x; (v) $f(x)$ is bounded from below on the support of $M(\cdot)$.

Assumption 2 (i) The kernel function $w(\cdot)$ is bounded, symmetric, and non-negative; (ii) $w(v) = 0$ for all v outside the unit sphere; (iii) $\int w(v)dv = 1$; (iv) $\int w(v)vv'dv = c_w I_q$, where c_w is a positive constant and I_q is the $q \times q$ identity matrix; (v) $\int w^2(v)dv = d_w$, where d_w is a positive constant; (vi) $\int w^2(v)vv'dv = \nu_w I_q$, where ν_w is a positive constant.

Assumption 3 $(\hat{\lambda}, \hat{k}) \in \Lambda \times K$, where, for some $C_0 > 0$, $\Lambda = \{\lambda \in \mathcal{R}^m | \; \|\lambda\| < C_0 (log\ n)^{-1})\}$; and $K = [n^{\delta}, n^{1-\delta}]$ for some arbitrarily small $\delta \in (0, 1/2)$.

In analyzing the asymptotic behavior of the smoothing parameters, we first give the leading term of the CV function in the following theorem:

Theorem 1 *Under assumptions 1-3,*

$$CV(k,\lambda) = CV_0(k,\lambda) + (s.o.),$$

where

$$CV_0(k,\lambda) = B_1\left(\frac{k}{n}\right)^{\frac{4}{q}} + B_2\left(\frac{1}{k}\right) + B_3(\lambda) + B_4(\lambda)\left(\frac{k}{n}\right)^{\frac{2}{q}},$$

where B_1 *and* B_2 *are positive constants,* $B_3(\lambda)$ *and* $B_4(\lambda)$ *can be written as* $B_3(\lambda) = \sum_{s=1}^{m}\lambda_s^2 d_s + 2\sum_{s=1}^{m-1}\sum_{t>s}^{m}\lambda_s\lambda_t d_{ts}$ *and* $B_4(\lambda) = \sum_{s=1}^{m}\lambda_s c_s$*, where* d_t*,* d_{ts} *and* c_s *are some constants,* $(s.o.)$ *denotes terms have probability orders smaller than* $CV_0(k,\lambda)$*.*

Let $\hat{k}$ and $\hat{\lambda}$ denote, respectively, the values of k and λ selected by least squares cross-validation, and let k_0 and λ_0 denote, respectively, the values of k and λ that minimize $CV_0(k,\lambda)$. From Theorem 1 we immediately have the following result:

Theorem 2 *Under the same assumptions as in Theorem 1, we have that*

$$\hat{k} = k_0 + o_p(k_0)$$
$$\hat{\lambda} = \lambda_0 + o_p(\|\lambda_0\|)$$

We further derive the rate of convergence of $\hat{k}$ and $\hat{\lambda}$ to k_0 and λ_0, respectively:

Theorem 3 *Under assumptions 1-3, we have*

(i) If $q \leq 3$*,* $(\hat{k}-k_0)/k_0 = O_p\left(n^{-q/[2(4+q)]}\right)$ *and* $\hat{\lambda}-\lambda_0 = O_p\left(n^{-1/2}\right)$.

(ii) If $q \geq 4$*,* $(\hat{k}-k_0)/k_0 = O_p\left(n^{-2/(4+q)}\right)$ *and* $\hat{\lambda}-\lambda_0 = O_p\left(n^{-4/(4+q)}\right)$.

Let $\gamma = (k,\lambda')$ and $\hat{\gamma} = (\hat{k},\hat{\lambda}')$. We use $\hat{g}_{\hat{\gamma}}(x)$ to denote the local constant k-nearest neighbor estimator of $g(x)$ using $k = \hat{k}$ and $\lambda = \hat{\lambda}$. Theorem 4 gives the asymptotic distribution of $\hat{g}_{\hat{\gamma}}(x)$:

Theorem 4 *Under assumptions 1-3, for each $x \in \mathcal{R}^{q+m}$ such that $f(x) > 0$, we have that*

$$\hat{k}^{1/2}\left[\hat{g}_{\hat{\gamma}}(x) - g(x) - \left(\mu_k(x)\left(k_0/n\right)^{2/q} + \hat{\lambda}'\mu_l(x)\right)\right] \xrightarrow{d} N\left(0, c_0 d_w \sigma^2(x)\right)$$

where

$$\mu_k(x) = \quad c_w(c_0 f(x))^{-2/q}\left[tr[\nabla^2 g(x)]/2 + [\nabla f(x)'\nabla g(x)]/f(x)\right],$$

$\mu_l(x)$ is an $m \times 1$ vector whose s-th element is given by

$$\mu_{ls}(x) = \quad \sum_{z_d \in D} \mathbf{1}_s\left(z^d, x^d\right)\left[g\left(x^c, z^d\right) - g(x)\right] f\left(x^c, z^d\right)/f(x),$$

c_0 is the volume of a unit ball in $\mathcal{R}^q$, $c_w = \int w(v)v_s^2 dv$, $d_w = \int w^2(v)dv$, and $\mathbf{1}_s\left(z^d, x^d\right) = \mathbf{1}(z_s^d \neq x_s^d)\prod_{t\neq s}\mathbf{1}(z_t^d = x_t^d)$.

2.3 Monte Carlo Simulations

In this section, we use Monte Carlo simulations to examine the finite sample performance of the proposed estimator. We consider the following two data generating processes, for $i = 1, \ldots, n$,

- DGP1:

$$Y_i = X_{1,i}^d + X_{2,i}^d + 3\frac{X_{1,i}^c}{\{X_{1,i}^c\}^4 + 1} + u_i, ;$$

- DGP2:

$$Y_i = X_{1,i}^d + X_{2,i}^d + 3\frac{X_{1,i}^c}{\{X_{1,i}^c\}^4 + 1} + \log(X_{2,i}^c) + u_i,$$

where $X_{j,i}^d \in \{0, 1\}$ with $\mathrm{P}(X_{j,i}^d = 1) = 0.5$ for $j = 1, 2$, $X_{1,i}^c \sim N(0, 2)$, $X_{2,i}^c \sim \chi_3^2$, $u_i \sim N(0, 1)$, and $X_{1,i}^d, X_{2,i}^d, X_{1,i}^c, X_{2,i}^c$, and u_i are independent. We experiment with three sample sizes: $n = 50, 100$ and 200. Each experiment is repeated 1,000 times.

We plot in Figure 2.1 the function $f(x) = x/(x^4+1)$ for $x \in [-5, 5]$. This curve is rather smooth except for a sharp peak and a sharp trough about zero. This kind of abrupt change in curvature calls for local or variable bandwidths in kernel regressions, and the nearest neighbor method is likely to outperform the kernel method. We consider in our experiments:

- K_{NW}: the (Nadaraya-Watson) kernel estimator for mixed discrete and continuous regressors of Racine and Li (2004).

- $K_{nn,1}$: the k-nearest neighbor estimator for mixed discrete and continuous regressors with a single k for all continuous variables.

For the sake of comparison, we also consider the following alternative nearest neighbor estimators:

- $K_{nn,f}$: 'frequency' nearest neighbor estimator, which estimates a separate model for each possible outcome of the discrete variables.

- $K_{nn,2}$: the k-nearest neighbor estimator for mixed discrete and continuous regressors with k being allowed to differ across continuous variables.

We employ the method of least-squares cross validation, as is discussed in the previous section, to select the bandwidths in all experiments. We use the package `np` for nonparametric methods in R to implement the simulations. We use the mean integrated square error to measure the overall performance of these estimators, where the mean integrated square error is approximated by the sum of mean square errors at all sample points. The simulation results are reported in Table 2.1.

Across the three sample sizes and two DGP's, we observe that the nearest neighbor estimators outperform the kernel estimators in our experiments. Similar to the results of Racine and Li (2004) on the kernel estimators, the kernel smoothing of

Figure 2.1: Plot of $x/(x^4+1)$

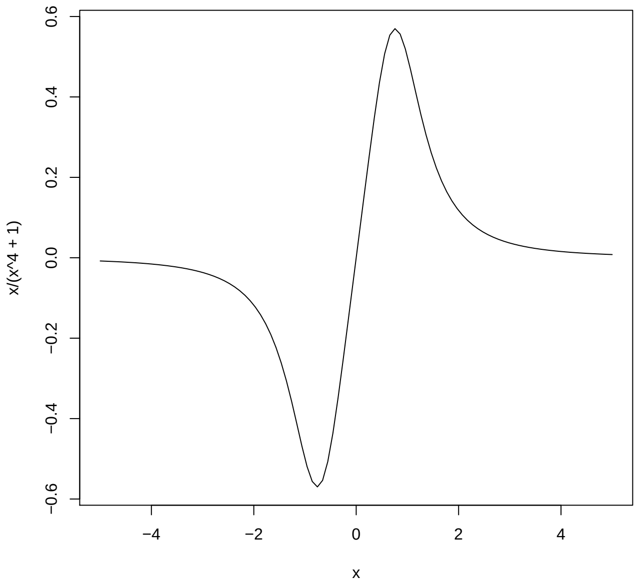


Table 2.1: Summary of simulation results

		K_{NW}	$K_{nn,1}$	$K_{nn,f}$	$K_{nn,2}$
DGP1					
$n = 50$	Mean MSE	0.418	0.352	0.455	
	Median MSE	0.401	0.337	0.445	
$n = 100$	Mean MSE	0.277	0.225	0.302	
	Median MSE	0.269	0.215	0.294	
$n = 200$	Mean MSE	0.179	0.138	0.181	
	Median MSE	0.177	0.133	0.178	
DGP2					
$n = 50$	Mean MSE	0.680	0.614	0.732	0.616
	Median MSE	0.655	0.594	0.715	0.589
$n = 100$	Mean MSE	0.538	0.489	0.612	0.478
	Median MSE	0.525	0.479	0.597	0.471
$n = 200$	Mean MSE	0.412	0.376	0.467	0.371
	Median MSE	0.408	0.369	0.460	0.364

discrete variables in the nearest neighbor estimators significantly improves upon the frequency estimator, as is evident by the considerable performance gap between $K_{nn,1}$ and $K_{nn,f}$. In DGP2 with two continuous variables, we observe little difference between $K_{nn,1}$ and $K_{nn,2}$, where the latter allows different k's for different variables. This result seems to support using a single k in multivariate nearest neighbor regression. Practitioners may benefit from this recommendation since using a single k would undoubtedly reduce the computation cost of cross-validation-bandwidth-selection, especially when the dimension of continuous variables is large.

2.4 Empirical Application

In this section, we apply the proposed estimator to estimate the relationship between crop yield and climate conditions. We are concerned with the average corn yield as a function of average temperature and precipitation in Iowa, the largest corn-producing state in the United States. Our data consist of annual average corn yield (in bushes per acre), growing season temperature (centigrades), precipitation (in inches) for 9 agriculture reporting districts in Iowa from 1990 through 2011. The data were downloaded from the United States Department of Agriculture (USDA) website.

We apply the nearest neighbor estimator $K_{nn,1}$ and the kernel estimator K_{NW} for mixed discrete and continuous variables to our dataset. We use data from the years 1990 through 2005 for estimation. Our model takes the form, for $i = 1, \ldots, 9$ and $t = 1990, \ldots, 2005$,

$$\text{Yield}_{i,t} = g(\text{District}_i, \text{Temperature}_{i,t}, \text{Precipitation}_{i,t}) + u_{i,t},$$

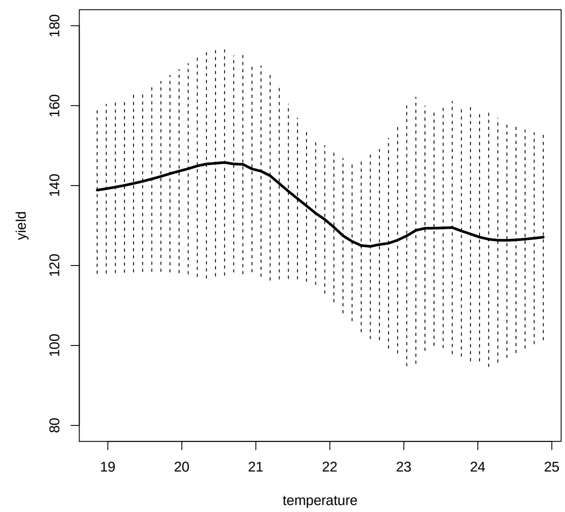
where the District$i$'s are dummy variables for agriculture districts in Iowa, g is the conditional mean of average corn yield given agriculture district, temperature and precipitation, and $u_{i,t}$ is an error term with mean zero and finite variance.

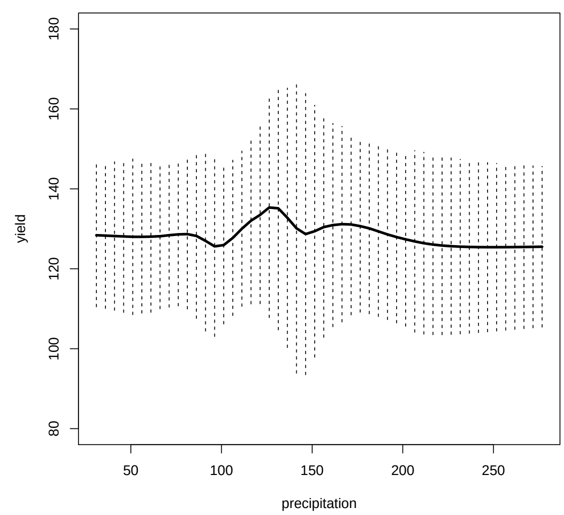
We use the out-of-sample mean square error (MSE) on data from the years 2006 through 2011 to assess the performance of our estimators. The out-of-sample MSE from the $K_{nn,1}$ estimator is 1,713, considerably lower than 2,099 from the K_{NW}. Below in Figure 2.2 we examine the 'marginal effect' of the two climate variables on corn yield by plotting the conditional mean of corn yield as a function of temperature or precipitation, evaluated at the median of the other regressors. We indicate the asymptotic 95% confidence interval by dotted lines above each evaluation point. Our estimation results clearly suggest nonlinear relationships between corn yield and climate variables. It is seen that average corn yield increases with average growing season temperature between 19 and 20.5 centigrade, declines between 20.5 and 22.5 centigrade, and largely levels off above 22.5 centigrade. Similarly, moderate levels of average precipitation (between 100 and 150 inches per year) are associated with higher corn yields relative to a lower or higher amount of precipitation.

2.5 Conclusion

In this paper, we consider a method of nonparametric conditional mean function estimation for data with both continuous and discrete regressors whereby the continuous regressors are smoothed using the k-nearest neighbor method and the discrete regressors are smoothed as in Racine and Li (2004). We analyze the asymptotic behavior of smoothing parameters selected by least-squares cross-validation, and we derive the asymptotic normal distribution of the corresponding regression function estimator. Our Monte Carlo simulations demonstrate the good finite sample performance of the proposed estimator. We then apply our method to the estimation of conditional corn yield in Iowa as a function of agricultural district, annual average temperature and precipitation.

Figure 2.2: Estimated marginal effects





3. TESTING ERROR SERIAL CORRELATION IN FIXED EFFECTS NONPARAMETRIC PANEL DATA MODELS, WITH WEI LONG AND CHENG HSIAO

Nonparametric and semiparametric methods allow for the estimation of panel data models that impose relatively few assumptions. This flexibility has made these methods increasingly popular among applied researchers. An early paper by Li and Stengos (1996) proposes a method for estimating a fixed effects panel data model that uses standard methods for estimating nonparametric additive models such as the marginal integration method of Linton and Nielson (1995) or a backfitting method such as in Opsomer and Rupert (1997) or Mammen, Linton, and Nielsen(1999). However, this method does not take full advantage of the structure of the model, and several more recent papers introduce methods that use more of this structure. Baltagi and D. Li (2002) propose a method that uses a series approximation to estimate the regression function. Henderson, Carroll, and Li (2008) introduce an iterative nonparametric kernel estimator and conjecture its asymptotic distribution.

At the same time, parametric dynamic panel models, which allow for the inclusion of lagged dependent variables as regressors, are also becoming more popular. Dynamic panel models are useful not only in applications in which the relationship between the dependent variable and its lagged values are of direct interest, but also in applications in which the lagged dependent variable is an important control variable. For an overview of dynamic panel models, see Baltagi (2008). While parametric dynamic panel models are increasingly popular, until very recently few, if any, estimators for dynamic panel models allowed the lagged dependent variable to enter the regression function nonparametrically. A recent paper by Su and Lu

(2013) addresses this gap in the literature. The authors introduce a recursive local polynomial estimation method for fixed effects dynamic panel models. They use methods developed in Mammen, Støve, and Tjøstjeim (2009) to derive the uniform consistency and asymptotic normality of the estimators under the assumption of zero serial correlation in the idiosyncratic errors.

We propose a test for the null hypothesis zero serial correlation. As argued in Li and Hsiao (1998), testing for serial correlation has long been a standard practice in applied econometric analysis because if the errors are serially correlated, not only an estimator ignoring serial correlation is generally inefficient, it can be inconsistent if the regressors contain lagged dependent variables. Moreover, strong serial correlation is often an indication of omitting important explanatory variables. Hence, testing autocorrelation is important because the choice of an appropriate estimation procedure for a given panel data model crucially depends on the error structure assumed by the model. Often the estimation methods could be considerably simplified if the errors are not autocorrelated. In this paper, we will generalize Li and Hsiao's test for zero error serial correlation in a nonparametric model to a fixed effects nonparametric model.

The remainder of the paper is organized as follows: Section 2 introduces the test statistic for a nonparametric model fixed effects model and derives its asymptotic distribution. Section 3 proposes using a bootstrap method to better approximate the null distribution of the test statistics. Section 4 reports Monte Carlo simulation results to examine the finite sample performance of the proposed test. Section 5 concludes the paper. The proofs of the main results are given in the two appendices.

3.1 The nonparametric fixed effects panel data model

We consider the following fixed effects nonparametric panel data model:

$$y_{it} = g(x_{it}) + \mu_i + \nu_{it}, \quad i = 1, ..., N;\ t = 1, ..., T, \tag{2.1}$$

where $x_{it} = (y_{i,t-1}, \tilde{x}'_{it})'$, $\tilde{x}_{it}$ is of dimension $(d-1) \times 1$ ($d \geq 2$) vector of explanatory variable that does not contain any lagged value of the dependent variable, μ_i is the fixed effect term.

We are interested in testing the null hypothesis that there is zero first order serial correlation in ν_{it}. That is, we test

$$H_0: \qquad E(\nu_{it}\nu_{i,t-1}) = 0.$$

We would like to test H_0 against an alternative that $E(\nu_{it}\nu_{i,t-1}) \neq 0$. However, since we have to first remove the fixed effects μ_i by first difference, the first difference error $\epsilon_{it} \equiv \nu_{it} - \nu_{i,t-1}$ at an MA(1) error structure when ν_{it} is serially correlated, our test statistic will be based on the sample analogue of $E(\epsilon_{it}\epsilon_{i,t-1})$ which equals to zero under H_0. If H_0 is false, ν_{it} is serially correlated, then $E(\epsilon_{it}\epsilon_{i,t-1}) = E[(\nu_{it} - \nu_{i,t-1})(\nu_{i,t-1} - \nu_{i,t-3})] = 2\gamma_2 - \gamma_1 - \gamma_3$, where $\gamma_j = E(\nu_{i,t-j}\nu_{it})$. Thus, our test will have power against the alternative hypothesis that $2\gamma_2 - \gamma_1 - \gamma_3 \neq 0$.

Because ν_{it} is not observable, we need to first estimate the $g(\cdot)$ function in order to estimate ν_{it}. Also, since the fixed effects can be arbitrarily correlated with the regressor x_{it} and there are no instrumental variables available that can take care of the correlation between x_{it} and μ_t, following Henderson et al (2008) and Su and Lu

(2013) we take a first difference to remove the fixed effects:

$$y_{it} - y_{i,t-1} = g(x_{it}) - g(x_{i,t-1}) + \nu_{it} - \nu_{i,t-1}. \tag{2.2}$$

Model (6) is an additive model with the restriction that, except for the negative sign in front of the second function, the two additive functions have identical functional forms. Henderson et al (2008) proposed using a profile likelihood back-fitting method to estimate model (6) under the assumptions that x_{it} and ν_{js} are independent with each other for all it and js. Su and Lu (2013) consider a similar dynamic panel data model in which x_{it} contains one lagged dependent variable, $y_{i,t-1}$, and propose to use a local polynomial method to estimate the $g(\cdot)$ function using a back-fitting method. In this paper we will adopt the estimation method proposed by Su and Lu (2013).

Note that x_{it} contains $y_{i,t-1}$ which is correlated with $\nu_{i,t-1}$. However, given that ν_{it} is a serially uncorrelated process, $x_{i,t-1} = (y_{i,t-2}, \tilde{x}'_{i,t-1})'$ is uncorrelated with $\nu_{it} - \nu_{i,t-1}$. Hence, taking the conditional expectation of (6) conditional on $x_{i,t-1} = x$, we obtain

$$E(\Delta y_{it} | x_{i,t-1} = x) = E[g(x_{it}) | x_{i,t-1} = x] - g(x). \tag{2.3}$$

Let $f_{t,t-1}(z|x)$ denote the conditional density function of x_{it} at $x_{it} = z$ conditional on $x_{i,t-1} = x$ and define $r(x) = -E(\Delta y_{it} | x_{i,t-1} = x)$. Then we can re-write (6) as

$$r(x) = g(x) - \int f_{t,t-1}(z|x) g(z) dz \equiv g(x) - (Ag)(x), \tag{2.4}$$

where $(Ag)(x) = \int f_{t,t-1}(z|x) g(z) dz$.

Note that A is a linear operator. Equations (7) or (8) suggest a recursive (back-fitting) method to estimate $g(x)$. For expositional simplicity we will discuss a local constant recursive estimator below; see Su and Lu (2013) for a general local polynomial estimator. Let $\hat{g}_{[l-1]}(x)$ denote the $l-1$ step estimate of $g(x)$. Then the next step estimator is given by

$$\hat{g}_{[l]}(x) = \hat{r}(x) + \hat{E}[g_{[l-1]}(x_{it})|x_{i,t-1} = x], \tag{2.5}$$

where

$$\begin{aligned} \hat{r}(x) &= -\frac{\frac{1}{NT_3}\sum_{j=1}^{N}\sum_{s=4}^{T}\Delta y_{js}K_{j,s-1,x}}{\hat{f}(x)}, & (2.6)\\ \hat{E}[g_{[l-1]}(x_{it})|x_{i,t-1} = x] &= \frac{\frac{1}{NT_3}\sum_{j=1}^{N}\sum_{s=4}^{T}\hat{g}_{[l-1]}(x_{js})K_{j,s-1,x}}{\hat{f}(x)}, & (2.7)\\ \hat{f}(x) &= \frac{1}{NT_3}\sum_{j=1}^{N}\sum_{s=4}^{T}K_{js,x}, & (2.8) \end{aligned}$$

where $T_j = T - j$ and $K_{js,x} = K((x_{js} - x)/h) = \prod_{m=1}^{d} k((x_{js,m} - x_m)/h_m)$ is the product kernel function.

The above estimation procedure requires one to use an initial estimator to start the iterative procedure. Following Henderson et al (2008) and Su and Lu (2013) we use a nonparametric series estimator as an initial estimator. Letting $p(x)$ be a $L \times 1$ vector of series base functions, we use the linear combination of them: $p(x)'\beta$ to approximate $g(x)$, so that the initial estimator of $g(x)$ is given by

$$\hat{g}_{[0]}(x) = p(x)'\hat{\beta} = p(x)'(\tilde{P}'\tilde{P})^{-1}\tilde{P}\Delta Y,$$

where $\tilde{P}$ is a $(nT_3) \times L$ matrix with a typical row given by $p(x_{it})' - p(x_{i,t-1})'$ and ΔY is $(nT_3) \times 1$ with a typical element given by $y_{it} - y_{i,t-1}$.

We define ϵ_{it} and $\hat{\epsilon}_{it}$ as follows:

$$\begin{aligned}\epsilon_{it} &= \nu_{it} - \nu_{i,t-1}, \\ \hat{\epsilon}_{it} &= y_{it} - y_{i,t-1} - (\hat{g}_{it} - \hat{g}_{i,t-1}),\end{aligned}$$

where g_{it} denotes $g(x_{it})$.

Then our test statistic I_N is based on the sample analogue of $E(\epsilon_{it}\epsilon_{i,t-2})$ defined as follows:

$$I_N \equiv \frac{1}{NT_3}\sum_{i=1}^{N}\sum_{t=4}^{T}\hat{\epsilon}_{it}\hat{\epsilon}_{i,t-2}. \tag{2.9}$$

We derive the asymptotic distribution of I_N under zero serial correlation in ν_{it} under the following assumptions which are similar to the ones imposed in Su and Lu (2013):

Assumption A1

(i) The random variables $(y_i, x_i, \mu_i, \nu_i), i = 1, ..., N$ are independent and identically distributed across the i index, where $y_i = (y_{i1}, ..., y_{iT}), x_i = (x_{i1}, ..., x_{iT}), \nu_i = (\nu_{i1}, ..., \nu_{iT})$.

(ii) $(y_{it}, x_{it}, \nu_{it})$ is strictly stationary in t.

(iii) $E[\epsilon_{it}^2|x_{it}] = \sigma_\epsilon^2$.

(iv) Let $f_t(\cdot)$ denote the PDF of x_{it}, and let $\mathcal{D}$ denote its support. We assume that $\mathcal{D}$ is a compact set.

(v) The PDF $f_t(\cdot)$ is uniformly bounded and is bounded below from 0 on its support.

(vi) $E(\nu_{it}|x_{it}, x_{i,t-1}, ..., x_{i1}) = 0$ a.s. under H_0.

(vii) $\|g\|_2 < C$ for some $C < \infty$, where $\|g\|_2 \equiv \left(\int g(x)^2 f(x)dx\right)^{1/2}$.

(viii) $\int\int[g(z) - g(x)]^2 f_t(x) f_{t|t-1}(z|x)dxdz > 0$ for $t = 2, .., T$.

(ix) $\int\int\left[\frac{f_{t|t-1}(z|x)}{f_t(z)}\right]^2 f_t(z)f_{t-1}(x)dzdx < \infty$.

(x) $sup_{z\in\mathcal{D}}\int|g(z)|f_{t|t-1}(z|x)dz < \infty$.

(xi) The functions $f_t(\cdot)$ and $g(\cdot)$ have up to second-order partial derivatives that exist and are uniformly continuous.

(xii) The kernel function $k : \mathbb{R} \to \mathbb{R}$ is a symmetric and continuous PDF that has compact support.

(xiii) T is fixed. As $N \to \infty$, $\|h\| \to 0$, $(Nh_1...h_d)/logN \to \infty$, $N\|h\|^8 \to 0$, where $\|h\| = \sqrt{\sum_{j=1}^d h_j^2}$ is the Euclidean norm.

Assumption A1 (i)-(ii) assume that the data is iid across the i index, and stationary across the t index, the stationarity assumption can be dropped (i.e., Su and Lu 2013), but it will make proof arguments longer. The conditional homoskedasticity assumption A1 (iii) can also be relaxed to allow for conditional heteroskedastic errors as in Li and Hsiao (1998), and Su and Lu (2013), again this will make the proofs longer. Assumption A1 (iv)-(v) assume that x_{it} has a compact support and that its density function is bounded below away from zero in its support which. is also assumed in Su and Lu (2013). This assumption can be relaxed by using a density weight to modify our test statistic or using some trimming function to trim small values of the estimated density as in Robinson (1988). Assumption A1 (vi) basically requires that ν_{it} is a martingale difference process and is also uncorrelated with $\tilde{x}_{it}$ under H_0. Assumptions A1 (vii)-(xi) impose some restrictions on $g(\cdot)$, $f(\cdot)$ and

$f_{t|t-1}(\cdot)$. They are quite standard and similar to those in Su and Lu (2013). Finally, A1 (xii)-(xiii) impose restrictions on the kernel function and the smoothing parameters. Assumption A1 (xiii) restrict that $d < 8$. Given the 'curse of dimensionality' of nonparametric estimation method, it is unlikely that one apply nonparametric estimation method to a model with $d \geq 8$. Even so the condition that $N\|h\|^8 \to 0$ as $N \to \infty$ in assumption A1 (xiii) can be relaxed to $N\|h\|^{4\nu} \to 0$ as $N \to \infty$, where $\nu \geq \min\{2, [d/4] + 1\}$ is an even integer, and also replace the second order kernel by a higher ν^{th} order kernel.

Theorem 1 gives the asymptotic distribution of the test statistic.

Theorem 1 *Under the null hypothesis of no serial correlation in ν_{it}, and under Assumption A1, we have that*

$$J_N = \frac{\sqrt{NT_3} I_N}{\hat{\sigma}_\epsilon^2} \xrightarrow{d} N(0, 1),$$

where $\hat{\sigma}_\epsilon^2 = \frac{1}{NT_2} \sum_{i=1}^N \sum_{t=3}^T \hat{\epsilon}_t^2$ is a consistent estimator of σ_ϵ^2.

It is easy to show that under H_1 that ν_{it} is serially correlated, our test statistic $I_N \xrightarrow{p} E(\epsilon_{it}\epsilon_{i,t-1}) = 2\gamma_2 - \gamma_1 - \gamma_3 \neq 0$, where $\gamma_j = E(\nu_{it}\nu_{i,t-j})$. This together with $\hat{\sigma}_\epsilon^2 = O_p(1)$ imply that $J_N = \sqrt{NT_3} I_N / \hat{\sigma}_\epsilon^2 \to \infty$ at the rate of $\sqrt{N}$. Hence, our test will reject H_0 with probability approaching one as $N \to \infty$.

3.2 A bootstrap test procedure

In order to improve the finite-sample performance of the test, we propose a bootstrap procedure. The procedure consists of the following steps:

1. Estimate $g(\cdot)$ as discussed in Section 2 and obtain $\hat{u}_{it} = y_{it} - \hat{g}_{it}$, $i = 1, 2, ..., N; t = 1, ..., T$.

2. First-difference $\hat{u}_{it}$ to obtain $\hat{\epsilon}_{it} = \hat{u}_{it} - \hat{u}_{i,t-1}$ for $i = 1, .., N$ and $t = 2, ..., T$.

3. Obtain bootstrap errors ϵ^*_{it} for $i = 1, .., N$ and $t = 1, ...T$ by random sampling without replacement from $\{\hat{\epsilon}_{it}\}_{i=1,t=1}^{N,T}$.

4. Compute the test statistic J^*_N as in section 2, replacing $\hat{\epsilon}_{it}$ with ϵ^*_{it}.

5. Repeat steps 2 through 4 a large number (call it B) times and obtain the empirical distribution of $\{J^*_{N,j}\}_{j=1}^{B}$, from which we can obtain bootstrap critical values.

Note that when generating the bootstrap sample we have imposed the null hypothesis that ϵ^*_{it} is serially uncorrelated so that J^*_N mimics the null distribution of J_N whether the null hypothesis is true or false. This procedure is relatively simple and computationally efficient, as it does not require repeated estimation of $g(\cdot)$. In the next section we show that this procedure performs well in finite sample applications. We leave the theoretical justification to the above proposed bootstrap as a future research topic.

3.3 Monte Carlo simulations

In this section we present the results of some Monte Carlo simulations to examine the finite sample performance of the bootstrap test procedure described in the previous section. We consider the following three data generating processes (DGP) which are similar to the cases considered in Su and Lu (2013):

DGP 1: $y_{it} = 1 + x_{it} + \mu_i + \nu_{it}$;

DGP 2: $y_{it} = 1 + x_{it} + x^2_{it} + \mu_i + \nu_{it}$;

DGP 3: $y_{it} = 1 + \sin(2x_{it}) + \mu_i + \nu_{it}$, where $x_{it} = 0.3x_{i,t-1} + \xi_{it}$, with ξ_{it} i.i.d U[-1,1]; $\mu_i = \bar{x}_i + \omega_i$, where $\bar{x}_i \equiv (1/T)\sum_{t=1}^{T} x_{it}$ and ω_i is i.i.d N(0,1). DGP 1 specifies a linear model while DGP 2 and 3 specify nonlinear panel data models. We use the Gaussian kernel function with bandwidth $h = \hat{\sigma}_x(nt)^{-1/5}$, where $\hat{\sigma}_x$ is the sample standard deviation of $\{x_{it}\}_{i=1,t=1}^{N,T}$. We follow Li and Hsiao (1998) and use three different processes for the error ν_{it}:

(i) i.i.d: ν_{it} i.i.d N(0,1);

(ii) AR(1): $\nu_{it} = 0.3\nu_{i,t-1} + u_{it}$, where u_{it} is i.i.d N(0,0.91);

(iii) MA(1): $\nu_{it} = 0.3u_{i,t-1} + u_{it}$, where u_{it} is i.i.d N(0,0.91).

Case (i) above corresponds to the null hypothesis that ν_{it} is serially uncorrelated, while cases (ii) and (iii) correspond to the alternative hypothesis that ν_{it} is a serially corrected process. Table B.1-B.3 gives the rejection frequencies for the three DGPs. Table B.1 corresponds to that the null hypothesis is true, while Tables B.2 and B.3 correspond to the case that the alternative hypothesis holds. The number of replications is 2000 and the number of bootstraps, B, is 1000. Table B.1 reports the estimated size of the bootstrap test when the error is i.i.d $N(0,1)$ and T equals to 5 and 10. It shows that, for all the three DGPs, the estimated sizes are slightly lower than the nominal size when $T = 5$. However, the estimated sizes increase toward their nominal size as N becomes larger.

Table B.2 and B.3 show that our test is quite powerful in detecting the AR(1) and MA(1) serially correlated errors: for MA(1) error process, even when $N = 50$, the estimated power is close to 1 and when $N = 100$ or 200, the estimated power equals to 1 for almost all the cases. Thus, the Monte-Carlo simulation results show that our proposed bootstrap test performs reasonably well in finite samples for the DGPs we considered.

For a robustness check, we further consider a conditional heteroskedastic error case: $u_{it} = \sigma_{it}\epsilon_{it}$, where ϵ_{it} is i.i.d N(0,0.91) and $\sigma_{it} = \sqrt{1 + x_{it}^2}$. Specifically, we consider the following three cases:

(i)' ν_{it} is $\sigma_{it}N(0, 1)$;

(ii)' AR(1): $\nu_{it} = 0.3\nu_{i,t-1} + u_{it}$, where u_{it} is $\sigma_{it}N(0, 0.91)$;

(iii)' MA(1): $\nu_{it} = 0.3u_{i,t-1} + u_{it}$, where u_{it} is $\sigma_{it}N(0, 0.91)$.

To save space, we only consider DGP 2 and fix $T = 10$. The results are presented in Table B.4. From Table B.4 we observe that the testing results for DGP 2 are very similar to its conditional homoskedastic error cases (compared with the results for DGP 2 in Table B.1, B.2 and B.3). The estimated sizes are quite close to the nominal sizes and the test remains powerful under conditional heteroskedatic errors.

Finally we report simulation result using critical values from the asymptotic standard normal distribution. The results for DGP 2 with $T = 10$ and with conditional heteroskedastic error cases (i)', (ii)' and (iii)' are displayed in Table B.5. We also report the mean and standard errors of the test statistic J_N. From Table B.5 we observe that the estimated sizes are significantly larger than the nominal sizes for all cases. This shows the necessity of using bootstrap method to overcome the size distortion of the asymptotic test.

4. VARYING-COEFFICIENT SINGLE-INDEX MODELS WITH ENDOGENEITY: THEORY AND APPLICATION, WITH ZHONGJIAN LIN

4.1 Introduction

Varying-coefficient models retain much of the flexibility of fully nonparametric models while offering improved interpretability and mitigating the "curse of dimensionality" through dimension reduction. Because of these and other advantages, varying-coefficient models have been the subject of much research. In an early paper, Hastie and Tibshirani (1993) introduce spline and kernel estimation methods for varying-coefficient models. Fan, Yao, and Cai (2003) and Cai, Fan, and Yao (2000) consider estimation of varying-coefficient models for panel data; Chen and Tsay (1993) and Cai, Fan, and Li (2000) consider the case of stationary times series; Cai, Li, and Park (2009) , Xiao (2009), and Sun, Cai, and Li (2013) consider the case of nonstationary time series.

Cai, Das, Xiong, and Wu (2006) argue that varying-coefficient models represent a particularly attractive compromise between parametric and nonparametric methods when some regressors are endogenous: by restricting the endogenous variables to enter the model linearly but allowing their coefficients to vary according to unknown functions of the exogenous variables, one can avoid some of the difficulties of estimating nonparametric models under endogeneity without sacrificing much in the way of flexibility. The authors propose a two-step instrumental variables procedure for estimating this type of model. Cai and Xiong (2012) consider instrumental variables estimation of a more general partially varying coefficient model.

While the model considered in Cai, Das, Xiong, and Wu (2006) has lower dimensionality than a fully nonparametric model, further dimension reduction might be desirable if the number of exogenous explanatory variables is large. In our model, such dimension reduction is achieved by restricting the coefficients to depend only on a single index of the exogenous variables. Single-index models have proven popular, due in part to their effective treatment of high dimensionality and their relative ease of interpretation (see Härdle and Hall (1993), Ichimura (1993), Li and Racine (2007) and the references therein), and single-index structure can be introduced quite naturally in varying-coefficient models. Xia and Li (1999) and Fan, Yao, and Cai (2003) consider estimation of varying-coefficient single-index models with exogenous regressors; to the best of our knowledge, ours is the first paper to consider varying-coefficient single-index models with endogenous regressors. For expositional simplicity, we restrict our analysis to the case of a single endogenous regressor.

We propose a multi-step estimation procedure. In the first step, the coefficient function is estimated using existing multivariate kernel instrumental variable methods which do not take advantage of the function's single-index structure. The first-step coefficient estimates are then fit to a single-index model using an iterative procedure adapted from Xia and Härdle (2006). Xia and Härdle show that their procedure has the advantage of not requiring $\sqrt{n}$-consistent pilot estimators in order to achieve $\sqrt{n}$ -consistent estimation of the index parameters; we show that the same is true of our procedure. In addition, we show that our estimator of the coefficient function converges to its true value at the standard rate for single-index models.

Finally, we apply the proposed method to study the determinants of aggregate liquidity in the U.S. stock market. We use a measure of illiquidity based on "price impact"– the responsiveness of stock returns to stock-specific order flow (i.e., to signed trading volume). In particular, to isolate illiquidity from information effects, we measure illiquidity by the price impact of order flow coming from uninformed traders. Inspired by Coval and Stafford (2007), we identify as uninformed traders those mutual funds that are experiencing extreme inflows or outflows of funds: as Coval and Stafford argue, trades made by these funds are likely motivated more by the need to quickly adjust the size of their portfolio than by any new information about the stocks involved. With a measure of this "forced trading" by mutual funds serving as an instrument for our proxy of order flow, we are able to estimate the price impact with a regression of stock returns on this proxy. Using quarterly data on a sample of S&P 500 stocks, we apply this instrumental variable approach and our varying-coefficient methods to investigate how the price impact varies according to market conditions. Due to sample size concerns, we restrict our analysis to the relationship between liquidity and interest rates – specifically, the Federal Funds Rate and a term spread variable.

The remainder of the paper is organized as follows: In section 4.2 we describe our model. In section 4.3 we describe our estimation method and present our asymptotic results. In section 4.4 we discuss a Monte Carlo simulation demonstrating the finite-sample performance of the estimators. In section 4.5 we present our empirical application. Section 4.6 concludes the paper.

4.2 The Varying Coefficient Single Index Model

We consider the following varying-coefficient single-index model

$$Y_t = X_t \beta(Z_t^T \gamma_0) + u_t, \qquad (t = 1, ..., n), \tag{4.1}$$

where Y_t, X_t and u_t are scalars,[1] Z_t is $q \times 1$ and γ_0 is a $q \times 1$ vector of unknown parameters; the functional form of $\beta(\cdot)$ is not specified. Z_t is exogenous. We allow for X_t to be endogenous, so we can have $E(X_t u_t) \neq 0$. We assume that there exists

[1] It straightforward to generalize the model for X_t to be a vector of random variables, for expositional simplicity, we will only consider the scalar X_t case in the paper.

an instrumental variable W_t such that $E(W_tX_t) \neq 0$ and $E(W_tu_t) = 0$. In fact, we will further assume that $E(u_t|W_t, Z_t) = 0$. Then multiplying (4.1) by W_t and taking conditional expectations we obtain

$$E(W_tY_t|Z_t) = E(W_tX_t|Z_t)\beta(Z_t^T\gamma_0) \tag{4.2}$$

because $E(W_tu_t|Z_t) = E[W_tE(u_t|Z_t, W_t)|Z_t] = 0$. Provided that $E[W_tX_t|Z_t]$ is invertible, equation (4.2) leads to

$$\beta(Z_t^T\gamma_0) = E[W_tX_t|Z_t]^{-1}E(W_tY_t|Z_t) \equiv g(Z_t). \tag{4.3}$$

The conditional mean functions in (4.3) are unknown, but they can be consistently estimated by nonparametric methods. In this paper we will estimate $g(Z_t)$ by the local-constant kernel method:

$$\tilde{g}(Z_t) = \hat{E}[W_tX_t|Z_t]^{-1}\hat{E}(W_tY_t|Z_t), \tag{4.4}$$

where $\hat{E}(A_t|Z_t) = \sum_{s=1}^n A_sH_{b,st}M_{n,s}/\sum_{s=1}^n H_{b,st}M_{n,s}$ is the local constant kernel estimator of $E(A_t|Z_t)$, where $H_{b,st} = b^{-q}\prod_{j=1}^q L((Z_{sj} - Z_{tj}/b)$ is the product kernel function,[2] and Z_{tj} is the j^{th} component of Z_t; $M_{n,s}$ is a trimming function that trims out data near the boundary of the support so that we can obtain a uniform convergent rate for $\max_{1\leq t\leq n}|\tilde{g}(z_t) - g(z_t)|$ over $z_t \in M_{n,t}$. Consider the simple case that $Z_t \in [0,1]^q$: the trimming function $M_{n,t}$ can be chosen as $M_{n,t} = \prod_{l=1}^q \mathbf{1}[\delta_n \leq Z_{tj} \leq 1-\delta_n]$, where $\delta_n \to 0$ and $b/\delta_n \to 0$ as $n \to \infty$. The use of the trimming function guarantees that the estimation bias is the same whether at the interior point or at the boundary

[2]For expositional simplicity, we assume that $b_1 = b_2 = ... = b_q = b$. In practice, one should always use a different b_l for different l, for $l = 1, ..., q$.

point of the trimmed data support. Also note that $\lim_{n\to\infty} M_{n,t} = \prod_{l=1}^{q}[0 \leq Z_{tj} \leq 1]$ so that asymptotically we include all observations in the data support when computing the nonparametric conditional mean functions.

We discuss how to estimate γ_0 and the unknown function $\beta(\cdot)$ in the next section.

4.3 The Estimation Method

Our estimation strategy follows similar steps as in Xia and Härdle (2006); see also Chen, Gao, and Li (2013). The main difference is that we allow X_t to be endogenous, and so our estimation method uses instrumental variable and nonparametric kernel estimation methods.

For the index coefficient vector γ_0 to be identified, we need an identification condition. We assume that there is a unique vector $\gamma_0 \in \mathcal{R}^q$ with $\gamma_0^T\gamma_0 = 1$ that makes

$$E[(g(Z_t) - \beta(Z_t^T\gamma_0)]^2 = 0.$$

Or equivalently, we assume that for all $\gamma \neq \gamma_0$ with $\gamma^T\gamma = 1$, we have

$$E[(g(Z_t) - \beta(Z_t^T\gamma)]^2 > 0.$$

Note that

$$\gamma_0 = argmin_\gamma E\left\{\left[g(Z_t) - \beta(Z_t^T\gamma)\right]^2\right\} \tag{5}$$

subject to $\gamma^T\gamma = 1$. By conditioning on $Z^T\gamma$, we observe that (5) equals to $E(\sigma^2(Z_t^T\gamma))$, where

$$\sigma^2(Z_t^T\gamma) = E\left[\left(g(Z_t) - \beta(Z_t^T\gamma)\right)^2 |Z_t^T\gamma\right]. \tag{6}$$

By the law of iterative expectations, it follows that

$$E\left\{\left[g(Z_t)-\beta(Z_t^T\gamma)\right]^2\right\}=E[\sigma^2(Z_t^T\gamma)].$$

Hence,

$$\gamma_0=argmin_\gamma E[\sigma^2(Z_t^T\gamma)] \tag{7}$$

subject to $\gamma^T\gamma=1$.

When Z_s is close to Z_t (the closeness will be guaranteed by a kernel weight function), we have

$$\beta(Z_s^T\gamma_0)\approx\beta(Z_t^T\gamma_0)+\beta'(Z_t^T\gamma_0)(Z_s-Z_t)^T\gamma_0.$$

For a given vector γ, an infeasible estimate of $\sigma^2_\gamma(Z_t^T\gamma)$ is given by (infeasible because $\beta'(\cdot)$ and γ are unknown)

$$\hat{\sigma}^2_\gamma(Z_t^T\gamma)=\frac{1}{n}\sum_{s=1}^n\left[\tilde{g}(Z_s)-\tilde{g}(Z_t)-\beta'(Z_t^T\gamma)(Z_s-Z_t)^T\gamma\right]^2H_b(Z_s,Z_t), \tag{8}$$

where $H_b(Z_s,Z_t)=b^{-q}\prod_{l=1}^q L\left(\frac{Z_{sl}-Z_{tl}}{b}\right)$ is the multivariate product kernel function.

We estimate $\beta'(Z_t^T\gamma_0)$ and γ_0 by d_t and γ, where d_t and γ minimize the following objective function (sample analogue of $E[\sigma^2(Z_t^T\gamma)]$):

$$\frac{1}{n}\sum_{t=1}^n\hat{\sigma}^2_\gamma(Z_t^T\gamma) \;=\; \frac{1}{n^2}\sum_{s=1}^n\sum_{t=1}^n\left[\tilde{g}_s-\tilde{g}_t-d_t(Z_s-Z_t)^T\gamma\right]^2H_b(Z_s,Z_t). \tag{9}$$

We will use the shorthand notation: $Z_{st}=Z_s-Z_t$ and $H_{b,st}=H_b(Z_s,Z_t)$. For a given vector γ with $\gamma^T\gamma = 1$, minimizing (9) with respect to d_t gives

$$d_t = \left[\sum_{s=1}^{n}(Z_{st}^T\gamma)^2 H_{b,st}\right]^{-1} \sum_{s=1}^{n} Z_{st}^T\gamma(\tilde{g}_s - \tilde{g}_t)H_{b,st}. \tag{10}$$

Also, for a given d_t, minimizing (9) with respect to γ leads to

$$\gamma = \left[\sum_{s=1}^{n}\sum_{t=1}^{n} d_t^2 Z_{st} Z_{st}^T H_{b,st}\right]^{-1} \sum_{s=1}^{n}\sum_{t=1}^{n} d_t Z_{st}(\tilde{g}_t - \tilde{g}_s)H_{b,st}, \tag{11}$$

and we standardize $\gamma = sgn_1\, \gamma/|\gamma|$, where sgn_1 is the sign of the first component of γ, $|\gamma|= \sqrt{\gamma^T\gamma}$ is the Euclidean norm of γ.

In practice we iterate between (10) and (11) until convergence. We use $\bar{\gamma}$ to denote the convergent value of γ.

The above estimator uses a multivariate kernel and is not efficient since it does not utilize the univariate single-index structure. Next, we replace the multivariate (product) kernel function $H_{b,st}$ by a univariate kernel function $K_{h,st}^{\gamma} = h^{-1}K((Z_s^T\gamma - Z_t^T\gamma)/h)$, where $K(\cdot)$ is a univariate symmetric density function. Then the estimates for $\beta'(Z_t^T\gamma_0)$ and γ_0 are given by $\tilde{d}_t$ and $\tilde{\gamma}$ which are based on the following iterative procedures:

$$\tilde{d}_t = \left[\sum_{s=1}^{n}(Z_{st}^T\tilde{\gamma})^2 K_{h,st}^{\tilde{\gamma}}\right]^{-1} \sum_{s=1}^{n} Z_{st}^T\tilde{\gamma}(\tilde{g}_s - \tilde{g}_t)K_{h,st}^{\tilde{\gamma}}\ , \tag{12}$$

and

$$\tilde{\gamma} = \left[\sum_{s=1}^{n}\sum_{t=1}^{n} \tilde{d}_t^2 Z_{st} Z_{st}^T K_{h,st}^{\tilde{\gamma}}\right]^{-1} \sum_{s=1}^{n}\sum_{t=1}^{n} \tilde{d}_t Z_{st}(\tilde{g}_s - \tilde{g}_t)K_{h,st}^{\tilde{\gamma}}\ , \tag{13}$$

and we standardize $\tilde{\gamma} = sgn_1\, \tilde{\gamma}/|\tilde{\gamma}|$, where sgn_1 is the sign of the first component of $\tilde{\gamma}$, $|\tilde{\gamma}|$ is the Euclidean norm of $\tilde{\gamma}$.

In (12) and (13), the initial values for d_t and γ are obtained from the convergent values of d_t and γ using the multivariate kernel function, i.e., from (10) and (11).

Before discussing our asymptotic results, we list the assumptions underlying them.

Assumptions

1. The observations $\{(X_t, Y_t, Z_t)\}_{t=1}^{T}$ are a stationary β-mixing, where the mixing rate $\beta(\tau) = O(\rho^{-\tau})$ for some $0 < \rho < 1$.

2. With probability 1, Z_t lies in a compact set $\mathcal{D}$. .

3. The kernel functions $H(\cdot)$ and $K(\cdot)$ are symmetric, second-order kernel functions with bounded derivatives and compact support. The kernel $H(\cdot)$ is Lipschitz continuous. $K(\cdot)$ has a finitely integrable Fourier transform. .

4. Let $f_Z(\cdot)$ and $f_\theta(\cdot)$ denote the marginal density functions of Z_t and $Z_t^T\theta$, respectively. Then: $f_Z(\cdot)$ has bounded derivatives; $f_\theta(\cdot)$ has bounded derivate for any θ such that $|\theta|= 0$; there exists a compact set B such that $inf_{z\in B} f_z(z) > 0$.

5. $\beta(\cdot)$ has bounded, continuous ν^{th} derivative function, where $\nu \geq 2$ is a positive integer. The functions $E(Z_t|\, Z_t^T\gamma = v)$ and $E(Z_t Z_t^T|\, Z_t^T\gamma = v)$ have bounded derivatives; for some r¿3, $E(y^r|\, Z_t = z)$ is bounded.

6. $E(W_t X_t|\, Z_t = z) \neq 0$ for all $z \in \mathcal{D}$.

7. $E(u_t|\, W_t, Z_t) = 0$.

8. $H(v) = \prod_{l=1}^{q} L(v_l)$ is a product kernel with $L(\cdot)$ being a second order univariate bounded symmetric density function.

9. Let b denote the bandwidth used in the multivariate kernel stage. Then $b \to 0$ and $nb^{q+2}/logn \to \infty$.

10. Let h denote the bandwidth used in the univariate kernel stage. Then $h \to 0$ and $nh^3/logn \to \infty$.

11. The univariate kernel function $K(\cdot)$ is a ν^{th} order kernel function, where $\nu \geq 2$ is a positive integer.

The following theorem shows that the convergent value $\bar{\gamma}$ obtained using the multivariate kernel weights is a consistent estimator of γ_0.

THEOREM 4.3.1 *Under assumptions 1 - 9 we have* $\bar{\gamma} - \gamma_0 = o_p(1)$.

The proof of Theorem 4.3.1 is given in the appendix.

Theorem 4.3.1 states that for *any* initial value γ with $\gamma^T\gamma = 1$, iterations between (10) and (11) lead to a consistent convergent value $\bar{\gamma}$.

These consistent initial estimators are sufficient to allow for the $\sqrt{n}$-consistency of the resulting estimators using the single-index kernel weights: letting $\hat{\gamma}$ denote the convergent value of $\tilde{\gamma}$, we have

THEOREM 4.3.2 *Under assumptions 1 - 11, we have*

$$\hat{\gamma} - \gamma = O_p(b^\nu + h^2 + n^{-1/2}). \tag{14}$$

The proof of Theorem 4.3.2 is given in the appendix.

If one further imposes the conditions that $b^\nu = O_p(n^{-1/2})$ and $h = O_p(n^{-1/4})$ (an under-smoothing condition), then (14) implies that $\hat{\gamma} - \gamma = O_p(n^{-1/2})$.

With the result of Theorem 4.3.2, our estimator for $\beta(z^T\gamma_0)$ is given by

$$\hat{\beta}(z^T\gamma_0) = \frac{n^{-1}\sum_{s=1}^{n} \tilde{g}_s K_h((Z_s - z)^T\hat{\gamma}_0/h)}{\hat{f}_\gamma(z^T\gamma_0)}, \tag{15}$$

where $K_h(v) = h^{-1}K(v)$, $\hat{f}_\gamma(z^T\gamma_0) = n^{-1}\sum_{s=1}^{n} K_h((Z_s - z)^T\hat{\gamma}_0/h)$ is the kernel estimator of the univariate density function $f_\gamma(z^T\gamma_0)$. Using $\hat{\gamma} - \gamma_0 = O_p(n^{-1/2})$, it is easy to show that

$$\hat{\beta}(z^T\gamma_0) - \beta(z^T\gamma_0) = O_p(b^\nu + h^2 + (nh)^{-1/2}).$$

In the next section, we conduct Monte Carlo simulations to examine the finite-sample performances of our proposed semiparametric estimators $\hat{\gamma}$ and $\hat{\beta}(z'\hat{\gamma})$.

4.4 Monte Carlo Simulations

In our simulations we consider models of the form

$$Y_t = X_t\beta(Z_t^T\gamma) + u_t, \tag{16}$$

where

$$x_t = 0.8w_t + 0.6\nu_t\ , \tag{17}$$

$$u_t = 0.5\nu_t + 0.5e_t\ , \tag{18}$$

and $Z_t = (Z_{1t}, Z_{2t})^T$, with Z_{1t} and Z_{2t} i.i.d uniformly on the interval $[-\sqrt{3}, \sqrt{3}]$; W_t and ν_t are i.i.d uniformly on the interval $[-\sqrt{3}, \sqrt{3}]$; e_t is i.i.d as $N(0,1)$; $Z_{1t}, Z_{2t}, W_t, \nu_t$, and e_t are independent of one another.

We consider four specifications for $\beta(\cdot)$:

1. $\beta(v) = 1.5\, sin\left(2^{-3/2}\pi\, v\right)$

2. $\beta(v) = 1.5\, sin\left(2^{-1/2}\pi\, v\right)$

3. $\beta(v) = cos\left(2^{-3/2}\pi\, v\right) - 0.8$

4. $\beta(v) = 0.8v^2 - 0.4\sqrt{2}v - 1$

In each replication, we use a random starting value of γ. Before normalization, the starting values of γ_1 and γ_2 are distributed independently, uniformly on the interval $[-1, 1]$. All kernel functions are Gaussian. In the steps involving multivariate kernel weighting we use a product kernel, with corresponding smoothing parameters $h = n^{-1/6}\hat{\sigma}_Z$, where $\hat{\sigma}_Z = (\hat{\sigma}_{Z_1}, \hat{\sigma}_{Z_2})^T$ is a vector of sample standard deviations of the Z_t's . In the steps involving univariate kernel weighting based on the single index, we use smoothing parameter $h = n^{-1/5}\hat{\sigma}_v$, where $\hat{\sigma}_v$ is the sample standard deviation of the single index $v_t \equiv Z_t^T\gamma$. Table 4.1 provides the results from 1,000 Monte Carlo replications for each of the sample sizes $n = 50, 100, 200$, and 400. It can be seen that the MSEs are often quite high when $n = 50$, but fall rapidly as the sample size increases.

Table 4.1: Mean squared errors for the four specifications for $\beta(\cdot)$.

specification	T	γ_1	γ_2	$\beta(\cdot)$
1	50	0.0157	0.04816	1.89728
	100	0.0062	0.01115	0.10984
	200	0.00177	0.00179	0.06646
	400	0.00081	0.0008	0.04291
2	50	0.02378	0.13206	0.78059
	100	0.00575	0.00785	0.47587
	200	0.00185	0.00188	0.34763
	400	0.00061	0.00061	0.24847
3	50	0.03821	0.20072	7.16526
	100	0.01654	0.03837	1.47082
	200	0.00623	0.00622	0.05939
	400	0.00284	0.00286	0.03909
4	50	0.01891	0.09582	0.38415
	100	0.00568	0.01057	0.31509
	200	0.00182	0.0018	0.13405
	400	0.00071	0.00071	0.08987

4.5 An Empirical Application

In this section we consider an empirical application to the determinants of liquidity in the stock market. We introduce a measure of aggregate liquidity, and we investigate how this measure varies according to market conditions.

Liquidity is defined by Pastor and Stambaugh (2003) as "a broad and elusive concept that generally denotes the ability to trade large quantities quickly, at low cost, and without moving the price". Liquidity has been a major subject of research in the finance literature. Chordia, Roll, and Subrahmanyan (2000), Hasbrouck and Seppi (2001), Huberman and Halka (2001) , and Jones (2002) show that a common component explains much of the variation in liquidity; Chordia, Roll, and Subrahmanyan (2001) study the determinants of this common component. A number of papers – including Pastor and Stambaugh (2003) , Acharya and Pedersen (2005), Watanabe and Watanabe (2008), Sadka (2006), Bekaert, Harvey, and Lundblad (2007), Korajczyk and Sadka (2008) , and Hasbrouck (2009) – show that investors require higher returns to hold assets that are less liquid and assets whose returns are more sensitive to fluctuations in market liquidity.

A natural measure of the illiquidity of an asset is the "price impact"– the change in the asset's price resulting from a trade of a given size. (See Goyenko, Holden, andTrzcinka (2009) for a discussion of short-term price impact measures, as well as other measures of liquidity.) In order to estimate an average price impact for a sample of stocks, we might consider running a pooled regression of the stocks' returns on their order flow. However, this regression may not yield consistent estimates of the average price impact of interest. This may be the case for at least two reasons.

First, if measured over a sufficiently long time horizon, the price impact estimated by the above procedure will likely be driven not just by illiquidity, but also by information effects. The immediate price impact of a trade is likely to be independent of its information content, as only the market participant who initiates it can determine the extent to which the trade was driven by information. However, the same is unlikely to be true of the longer-term price impact: while we would expect that the immediate price impact of an uninformed trade will eventually be reversed, that of an informed trade may persist as more market participants acquire the same information. Thus, since our hypothetical regression does not distinguish between order flow coming from uninformed and informed traders, the price impacts that it estimates will capture information effects as wells as illiquidity if measured over a sufficiently long time horizon. (For a discussion of the relationship between information content and price impact, and how it evolves, see Hasbrouck (1988)).

Second, market participants may have information about the current level of aggregate liquidity, as well as the relative liquidity of different stocks – information that is likely to influence their trading behavior. Thus, even after controlling for market conditions observable to us as researchers, market participants may tend to make more trades at times when aggregate liquidity is higher, and to trade liquid stocks more than illiquid ones. Thus, absolute order flow may be positively correlated with liquidity in the cross section and in the time series. If this is the case, our hypothetical regression will assign more weight to individual stocks and periods with higher liquidity, resulting in downward-biased estimates of average price impact.

Our analysis is further complicated by the fact that we do not directly observe signed order flow; following Pastor and Stambaugh (2003), we proxy signed order flow by volume signed by contemporaneous returns in excess of the market return. The use of this proxy adds to our concerns about endogeneity.

We address this potential endogeneity with an instrumental variables approach. Inspired by Coval and Stafford (2007), among others, we instrument our order flow proxy by a measure of "forced trading" by mutual funds - trades made by funds with extreme inflows or outflows of capital. Coval and Stafford document that funds experiencing extreme inflows tend to expand existing positions, while those experience extreme outflows contract existing positions. Based on this observation, they argue that trading by firms with extreme inflows or outflows can be seen as uninformed trading. Consistent with this theory, they find that flow-induced selling exerts downward pressure on security prices, flow-induced buying exerts upward pressure, and these effects are eventually reversed.

A number of recent papers consider the implications of the flow-induced pressure identified by Coval and Stafford (2007). Edmans, Goldstein, and Jiang (2012) and Khan, Kogan, and Serafeim (2012) identify some of its real effects. They find, respectively, that flow-induced price increases are associated with increased probability of takeover; and with increased probability of a seasoned equity offering, insider sales, and undertaking a merger or acquisition. Lou (2012) demonstrates that flow-induced trading can explain in full or in part three well-documented patterns: the persistence of mutual fund performance, the "smart money" effect, and stock price momentum. Jotikasthira, Lundblad, and Ramadorai (2012) find that flow-induced trading of emerging market stocks by funds in developed countries has significant impact on emerging market equity prices.

Using our instrumental variables approach to identify the average price impact, we examine how this price impact varies according to market conditions. The question of how liquidity and trading activity vary according to market conditions is considered extensively by Chordia, Roll, and Subrahmanyan (2001). Using data on a large sample of NYSE stocks spanning the period 1988 to 1998, they measure liquidity by daily market averages of quoted spreads, effective spreads, and market depth, and they measure trading activity by dollar volume and the total number of transactions. They perform first-differenced regressions to examine the effects of a number of explanatory variables on their measures of liquidity and trading activity; here we briefly summarize their results. In their regressions, both the Federal Funds rate and the term spread have positive effects on the quoted and realized spread (the effects are statistically significant for the quoted spread) and negative and statistically significant effects on depth and the measures of trading activity. They find little evidence that the default spread, measured by the difference in yield between Moody's Baa or better corporate bond index and that on a 10-year constant maturity Treasury bond, significantly affects liquidity or trading activity. They also find substantial day-of-the-week effects, substantial effects of contemporaneous market returns, and that depth and trading activity are higher in the two days leading up to an announcement of GDP, unemployment, or inflation figures.

Chordia, Sarkar, and Subrahmanyam (2003) show that increases in volatility predict decreases in liquidity, unexpected decreases in the Federal Funds rate increase liquidity, and increases in aggregate fund flows predict decreases in liquidity. Hameed, Kang, Viswanathan (2010) find that market returns affect liquidity, and the effect is asymmetric: negative returns lower liquidity more than positive returns increase it.

Due to sample size concerns, we concentrate on studying the relationship between interest rates and illiquidity. We consider two interest rate variables – the Federal Funds rate and a term spread variable. Chordia, Roll, and Subrahmanyan (2001) argue that there should be a negative relationship between interest rates and liquidity, as higher interest rates increase inventory costs for market makers and increase the costs of margin trading. Thus, we expect a positive relationship between the Federal Funds rate and our measure of illiquidity. Chordia, Roll, and Subrahmanyan (2001) argue that an increase in longer-term bond yields may cause investors to reallocate their portfolios, raising volume and in turn affecting liquidity. Thus, we may expect liquidity to be related to the term spread, although the sign of the relationship is unclear.

4.5.1 Model, Variable Description, and Data

We estimate the model

$$R_{it} = X_{it}\beta(Z_t) + u_{it} \ , \tag{19}$$

where R_{it} is the return in excess of the market return (here defined as the Fama-French Big Neutral benchmark) for stock i in quarter t; X_{it} is volume, signed by R_{it}, as a fraction of the number of shares outstanding; Z_t is a vector containing the Federal Funds rate and the term spread – the return on a 10-year treasury bond minus the Federal Funds rate.

As described above, we use a flow-induced mutual fund trading variable as an instrument for X_{it} in 19. We define flow-induced trading as the sum of a) purchases made by funds with flow in the top decile, and b) sales from funds with flow in the bottom decile. We construct our flow-induced trading variable W_{it} using the formula of Khan, Kogan, and Serafeim (2012), which is similar to the preferred formula of Coval and Stafford (2007). This formula can be expressed as

$$W_{it} = \frac{\sum_{j=1}^{J} \left(max(0, \Delta shr_{ijt}) \mathcal{I}(flow_{jt} > Percentile(90^{th})) - max(0, -\Delta shr_{ijt}) \mathcal{I} \left(flow_{jt} < Percentile(10^{th}) \right) \right)}{shrout_{it}} \ , \tag{20}$$

where Δshr_{ijt} is the first difference of the number of shares of stock i held by fund j in quarter t; $flow_{jt}$ is the inflow for fund j in quarter t as a fraction of its total net assets in quarter $t-1$; $shrout_{it}$ is the number of shares of stock i outstanding in quarter t; $\mathcal{I}(\cdot)$ is the indicator function.

Following Coval and Stafford (2007), we define the flow for mutual fund j in quarter t as

$$flow_{j,t} = \frac{TNA_{j,t} - (1 + R_{j,t})TNA_{j,t-1}}{TNA_{j,t-1}} \tag{21}$$

where $TNA_{j,t}$ is fund j's total net assets in quarter t, and $R_{j,t}$ is the return for fund j in quarter t. We calculate the quarterly flow by summing the monthly flows.

Our data on stock returns come from CRSP, and our data on mutual fund flows and mutual fund holdings come from the CRSP Survivor-Bias-Free U.S. Mutual Fund Database. Our sample consists of 418 stocks that were in the S&P 500 in 2003. Our sample period extends from the third quarter of 2003 through the fourth quarter of 2012.

4.5.2 Empirical Results

For purposes of comparison, we first consider estimation of (19) treating X_{it} as exogenous. In this estimation, we use a procedure which is identical to that detailed above, except that first-step coefficient estimates $\tilde{g}(Z_{it})$ are obtained using standard kernel methods for varying-coefficient models with exogenous regressors instead of by instrumental variables methods. Specifically, rather than estimating $g(\cdot)$ from (4.4) by

$$\tilde{g}(Z_t) = \hat{E}\left[W_t X_t | Z_t\right]^{-1} E(W_t Y_t | Z_t) \tag{22}$$

we instead estimate it by

$$\tilde{g}(Z_t) = \hat{E}\left[X_t^2 | Z_t\right]^{-1} E(X_t Y_t | Z_t) \tag{23}$$

where, as before, the conditional expectations are estimated using standard kernel methods.

This estimation yielded index coefficient $\hat{\gamma}_1 = 0.94$ for the Federal Funds rate and coefficient $\hat{\gamma}_2 = 0.33$ for the term spread. Panel (a) of Figure 4.1 gives the estimated price impact as a function of the single index $V_t = Z_t^T \hat{\gamma}$. This price impact falls into a narrow range, from 0.121 to 0.138, with a median of 0.130.

We next present results of instrumental variable estimation of (19). This estimation yielded index coefficient $\hat{\gamma}_1 = .16$ for the Federal Funds Rate and $\hat{\gamma}_2 = 0.99$ for the term spread. Noting that the standard deviation of the Federal Funds rate for our sample is 1.93 percentage points, while the standard deviation of the term spread is 1.34 percentage points, it can be seen that, for any fixed interest rates, the estimated standardized marginal effect of the term spread on $\beta(\cdot)$ is greater in absolute value than that of the Federal Funds rate. This is contrary to the results of our estimation using equation (22) (treating X_{it} as exogenous). Panel (b) of figure 4.1 shows the estimated price impact as a function of the index V_t ; it can be seen that this function is nonlinear. The price impact is increasing in V_t (and thus in the interest rate variables, as $\hat{\gamma}$ is positive) for most values of V_t, although for high values it is decreasing in V_t. Thus, our results are consistent with Chordia, Roll, and Subrahmanyan (2001) in that the Federal Funds rate and the term spread are generally positively related with illiquidity. The estimated price impact ranges widely, from 0.024 to 0.249, with a median of 0.201.

In order to highlight the differences between the results obtained using the instrumental variable and those obtained by treating X_{it} as exogenous, we combine the price impact curves from panels (a) and (b) of Figure 4.1 in Figure 4.2. It should be noted that, since the index coefficients – and thus the indexes for the two estimations – differ, a given index value does not correspond to the same interest rates for each curve. Our purpose here is simply to compare the distributions of the price impacts and the strength and linearity of their relationship with the index and thus the interest rate variables. It can be seen from Figure 4.2 that ignoring the endogeneity of X_{it} results in an estimated price impact curve that is relatively flat (the dashed curve), while our instrumental variable estimation shows a much more pronounced nonlinear relationship between price impact and the interest rate variables (the solid curve). It can also be seen that the instrumental variable estimates of price impact are generally higher.

While we do not present statistical tests of the endogeneity of X_{it}, these differences in the results obtained using the instrument and those obtained from treating X_{it} as exogenous may suggest that the the latter method underestimates the price impact, the degree to which the price impact depends on the interest rates, and the importance of the term spread relative the Federal Funds rate for liquidity.

4.6 Conclusion

We introduce instrumental variable methods for estimating varying-coefficient single-index models with an endogenous regressor. These methods provide an effective treatment of endogeneity, and of high dimensionality of the exogenous regressors, while retaining a good deal of flexibility. We show the consistency of our estima-

Figure 4.1: Estimated price impact

(a) Treating X_{it} as exogenous (b) Instrumental variables estimates

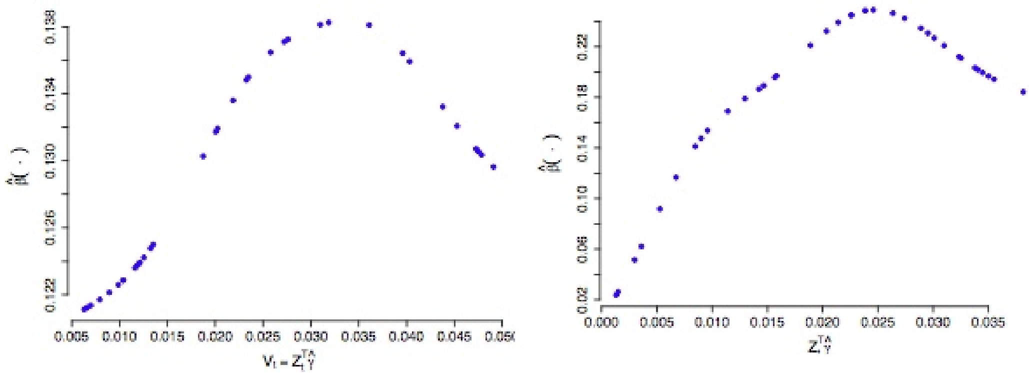
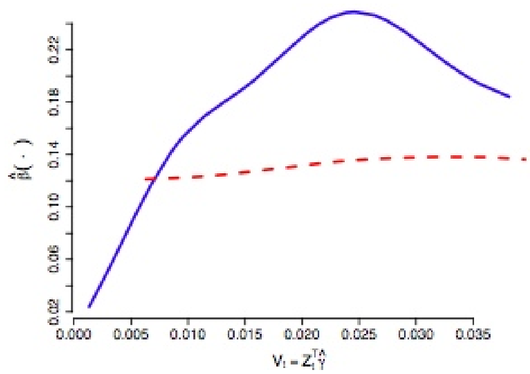


Figure 4.2: Price impact estimated using instrument (solid) and treating X_{it} as exogenous (dashed)



tors of the coefficient function, and the $\sqrt{n}$-consistency of our estimator of the index parameters. We demonstrate the practical power of our estimators in Monte Carlo simulations. An empirical application to liquidity in U.S. stock markets suggests that liquidity has a nonlinear and generally negative relationship with two interest rate variables – the Federal Funds rate and a term spread variable.

5. CONCLUSION

This dissertation adds to the literature on nonparametric and semiparametric regression methods. We consider nonparametric regression with mixed continuous and discrete explanatory variables. We introduce a test for error serial correlation in fixed effects nonparametric panel data models. Finally, we consider the estimation of varying-coefficient single-index models with an endogenous explanatory variable.

REFERENCES

Acharya, V., Pedersen, L. (2005): Asset pricing with liquidity risk. Journal of Financial Economics 77, 375-410.

Aitchison, J., Aitken, C. (1976): Multivariate binary discrimination by kernel method. Biometrika 63, 413-420.

Baltagi, B. (2008): Econometric Analysis of Panel Data. 4th ed. West Sussex: John Wiley and Sons.

Baltagi, B., Li, D. (2002): Series estimation of partially linear panel data models with fixed effects. Annals of Economics and Finance 3: 103-116.

Bachetta, P., van Wincoop, E. (2006): Can information heterogeneity explain the exchange rate determination puzzle? The American Economic Review 96, 552-576.

Bekaert, G., Harvey, C., Lundblad, C. (2007): Liquidity and expected returns: lessons from emerging markets. Review of Financial Studies 20, 1783-1831.

Breedon, F., Vitale, P. (2010): An empirical study of portfolio-balance and information effects of order flow on exchange rate. Journal of International Money and Finance 29, 504-524.

Cai, Z., Fan, J., Li, R. (2000): Efficient estimation and inferences for varying-coefficient models. Journal of the American Statistical Association 95, 888-902.

Cai, Z., Fan, J., Yao, Q. (2000): Functional-coefficient regression models for nonlinear time series. The Journal of the American Statistical Association 95, 941-956.

Cai, Z., Das, M., Xiong, H., Wu, X. (2006): Functional coefficient instrumental variables models. Journal of Econometrics 133, 207-241.

Cai, Z., Li, Q., Park, J. (2009): Functional-coefficient models for nonstationary time series data. Journal of Econometrics 148, 101-113.

Cai, Z., Xiong, H. (2012): Partially varying coefficient instrumental variables models. Statistica Neerlandica 66, 85-110.

Chen, R., Tsay, R.S. (1993): Functional-coefficient autoregressive models. Journal of the American Statistical Association 88, 298-308.

Chen, J., Gao, J., Li, D. (2013): Estimation in partially linear single-index panel data models with fixed effects. Journal of Business and Economic Statistics 31, 315-330.

Chordia, T., Roll, R., Subrahmanyan, A. (2000): Commonality in liquidity. Journal of Financial Economics 56, 3-28.

Chordia, T., Roll, R., Subrahmanyan, A. (2001): Market liquidity and trading activity. The Journal of Finance LVI, 501-530.

Chordia, T., Sarkar, A., Subrahmanyam, A. (2003): An empirical analysis of stock and bond market liquidity. Staff Report, Federal Reserve Bank of New York.

Coval, J., Stafford, E. (2007): Asset fire sales (and purchases) in equity markets. Journal of Financial Economics 86, 479-512.

Edmans, A., Goldstein, I., Jiang, W. (2012): The real effects of financial markets: the impact of prices on takeovers. The Journal of Finance LVII, 933-971.

Engle, R.F., Grange, C., Rice, J., Weiss, A. (1986): Semiparametric estimates of the relationship between weather and electricity sales. Journal of the American Statistical Association 81, 310-320.

Fan, Y., Li, Q. (1999): Root-N-consistent estimation of partially linear time series models. Journal of Nonparametric Statistics 11, 251-269.

Fan, J., Yao, Q., Cai, Z. (2003): Adaptive varying-coefficient linear models. Journal of the Royal Statistical Society 65, 57-80.

Frazzini, A., Lamont, A. O. (2008): Dumb money: mutual fund flows and the cross-section of stock returns. Journal of Financial Economics 88, 299-322.

Gao, J., Tong, H. (2004): Semi-parametric nonlinear time series model. Journal of Royal Statistical Society 66, 321-336.

Goyenko, R., Holden, C., Trzcinka, C. (2009): Do liquidity measures measure liquidity? The Journal of Financial Economics 92, 153-181.

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

Hameed, A., Kang, W., Viswanathan, S. (2010): Stock market declines and liquidity. The Journal of Finance 65, 257-293.

Härdle, W., Hall, P., Marron, J. (1988): How far are automatically chosen regression smoothing parameters from their optimum? Journal of the American Statistical Association 83, 86-89.

Härdle, W., Hall, P., Marron, J. (1992): Regression smoothing parameters that are not far from their optimum. Journal of the American Statistical Association 87, 223-233.

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

Hasbrouck, J. (1988): Trades, quotes, inventories, and information. Journal of Financial Economics 22, 229-252.

Hasbrouck, J., Seppi, D. (2001): Common factors in prices, order flow, and liquidity. Journal of Financial Economics 59, 383-411.

Hasbrouck, J. (2009): Trading costs and returns for US equities: estimating effective costs from daily data. Journal of Finance LXIV, 1445-1477.

Hastie, T. J., Tibshirani, R. J. (1993): Varying-coefficient models. Journal of the Royal Statistical Society.

Henderson, D., Carroll, R., Li, Q. (2008): Nonparametric estimation and testing of fixed effect panel data models. Journal of Econometrics 144, 257-275.

Huang, J. Z., Wu, C. O., Zhou, L. (2004): Polynomial spline estimation and inference for varying coefficient models with longitudinal data. Statistica Sinica 14, 763-788.

Huberman, G., Halka, D. (2001): Systematic liquidity. Journal of Financial Research XXIV, 161-178.

Ichimura, H. (1993): Semiparametric least square (SLS) and weighted SLS estimation of single-index models. Journal of Econometrics 58, 71-120.

Jones, C. (2002): A century of stock market liquidity and trading costs. Working Paper, Columbia University.

Jotikasthira, C., Lundblad, C., Ramadorai, T. (2012): Asset fire sales and purchases and the international transmission of funding shocks. The Journal of Finance 67, 2015-2050.

Khan, M., Kogan, L., Serafeim, G. (2012): Mutual fund trading pressure: firm-level stock price impact and timing of SEOs. The Journal of Finance LXVII, 1371-1394.

Korajczyk, R., Sadka, R. (2008): Pricing the commonality across alternative measures of liquidity. Journal of Financial Economics 87, 45-72.

Li, Q., Stengos, T. (1996): Semiparametric estimation of partially linear panel data models. Journal of Econometrics 71: 289-397.

Li, Q., Hsiao, C. (1998): Testing serial correlation in semiparametric panel data models. Journal of Econometrics, 87, 207-237.

Li, Q., Racine, J. (2007): Nonparametric Econometrics. Princeton: Princeton University Press.

Linton, O., Nielsen, J. (1995): A kernel method of estimating structured nonparametric regression based on marginal integration. Biometrika 82: 93-101.

Liu, Z., Lu, X. (1997): Root-N-consistent estimation of partially linear model by K-nn method. Econometric Reviews 16, 41120.

Lou, D. (2012): A flow-based explanation for return predictability. The Review of Financial Studies 25, 3457 - 3489.

Mack, Y., Rosenblatt, M. (1979): Multivariate k-nearest neighbor density estimates. Journal of Multivariate Analysis 9, 115.

Mammen, E., Linton, O., Nielsen, J. (1999): The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. Annals of Statistics 27: 1443:1490.

Mammen, E., Støve, B., Tjøstjeim, D. (2009): Nonparametric additive models for panels of time series. Econometric Theory 25, 442-481.

Opsomer, J.D., Ruppert, D. (1997): Fitting a bivariate additive model by local polynomial regression. Annals of Statistics 25, 186-211.

Ouyang, D., Li, D., Li, Q. (2006): Cross-validation and non-parametric *k* nearest-neighbour estimation. Econometrics Journal, 9, 448-471.

Pastor, L., Stambaugh, R. (2003): Liquidity risk and expected stock returns. The Journal of Political Economy 111, 642-692.

Racine, J., Li, Q. (2004): Nonparametric estimation of regression functions with mixed discrete and continuous regressors. Journal of Econometrics 119, 99130.

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

Sadka, R. (2006): Liquidity Risk and asset pricing. Journal of Financial Economics 80, 309-349.

Su, L., Lu, X. (2013): Nonparametric dynamic panel data models: kernel estimation and specification testing. Journal of Econometrics 176, 112-133.

Sun, Y., Cai, Z., Li, Q. (2013): Semiparametric functional coefficient models with integrated covariates. Econometric Theory 29, 659-672.

Watanabe, A., Watanabe, M. (2008): Time-varying liquidity risk and the cross section of stock returns. Review of Financial Studies 21, 2449-2486.

Xia, Y., Li, W.K. (1999): On single-index coefficient regression models. Journal of the American Statistical Association 94, 1275-1285.

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

Xiao, Z. (2009): Functional-coefficient cointegration models. Journal of Econometrics 152, 81-92.

APPENDIX A

APPENDIX TO SECTION 2

We re-state Lemma 1 and Lemma 2 from Ouyang, Li, and Li (2006) below for ease of reference.

Let $S_r = \{v : ||v-x|| < r\}$ (a ball centered at x with radius r), $G(r) = Prob[X_i \in S_r]$, $S_n = \{v : ||v-x|| < R_x\}$ and $P(S_n) = Prob[X_i \in S_n]$. Obviously $G(R_x) = P(S_n)$.

Lemma 1 *Let $h(r) = 1/[r^\mu G^\gamma(r)]$, μ and γ are integers such that $E[h(R_x)]$ exists, then*

$$E[h(R_i)|X_i] = (c_0 f(X_i))^{\mu/q} \left(\frac{k}{n}\right)^{-(\mu/q)-\gamma} [1 + o_p(1)] \tag{1}$$

where $c_0 = \pi^{q/2}/\Gamma((q+2)/2)$ is the volume of unit ball in $\mathcal{R}^q$.

Proof: See page 459 of Ouyang, Li, and Li (2006).

Remark 1: *Using equation (12) of Mack and Rosenblatt (1979), Liu and Lu (1997) have shown that (see lemma 1 of Liu and Lu) for $\xi = (\mu+\eta)/q$, where μ is an integer and η is a nonnegative integer less than or equal to $q-1$ $(0 \leq \eta \leq q-1)$,*

$$E[h(R_i)|X_i] = (c_0 f(X_i))^{\mu/q} \frac{n!}{k!} \frac{(k-\xi-\gamma)!}{(n-\xi-\gamma)!} \left(\frac{(k-\xi-\gamma)}{(n-\xi-\gamma)}\right)^{\eta/q} [1 + o_p(1)]. \tag{2}$$

Note that $\frac{n!}{k!}\frac{(k-\xi-\gamma)!}{(n-\xi-\gamma)!}\left(\frac{(k-\xi-\gamma)}{(n-\xi-\gamma)}\right)^{\eta/q} = \left(\frac{n}{k}\right)^{\xi+\gamma}\left(\frac{k}{n}\right)^{\eta/q}[1+o(1)] = \left(\frac{k}{n}\right)^{\eta/q-\xi-\gamma}[1+o(1)] = \left(\frac{k}{n}\right)^{-\lambda/q-\gamma}[1+o(1)]$. Substituting this into (2) proves lemma 1.

Lemma 2 *Let $A(x)$ be a measurable function of x. Then*

$$E[A(X_j)w(\frac{X_j - X_i}{R_i})|X_i, R_i] = \frac{(k-1)}{nG(R_i)} \int_{||x_j - X_i|| < R_i} f(x_j)A(x_j)w(\frac{x_j - X_i}{R_i})dx_j. \qquad (3)$$

Proof: It follows directly from equation (22) of Mark and Rosenblatt (1979) and the fact that $w(\frac{x_j - X_i}{R_i}) = 0$ for $||x_j - X_i|| \geq R_i$. See page 459 of Ouyang, Li, and Li (2006).

Proof of Theorem 1

Using $Y_i = g(X_i) + u_i$, we have $(M_i = M(X_i))$

$$\begin{aligned} CV(k,\lambda) =& \frac{1}{n}\sum_{i=1}^{n}\left[g\left(X_i\right) - \hat{g}_{-i}\left(X_i\right)\right]^2 M_i + \frac{2}{n}\sum_{i=1}^{n}\left[g\left(X_i\right) - \hat{g}_{-i}\left(X_i\right)\right] u_i M_i + \frac{1}{n}\sum_{i=1}^{n} u_i^2 M_i \\ &\equiv CV_1(k,\lambda) + CV_2(k,\lambda) + \frac{1}{n}\sum_{i=1}^{n} u_i^2 M_i, \end{aligned} \tag{4}$$

where $CV_1(k,\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left[g\left(X_i\right) - \hat{g}_{-i}\left(X_i\right)\right]^2 M_i$ and $CV_2(k,\lambda) = \frac{2}{n}\sum_{i=1}^{n}\left[g\left(X_i\right) - \hat{g}_{-i}\left(X_i\right)\right] u_i M_i$. The last term on the right of (4) does not depend on k or λ. We can show that $CV_2(k,\lambda)$ has smaller probability order than does $CV_1(k,\lambda)$. Hence, $CV_1(k,\lambda)$ is the leading term of $CV(k,\lambda)$. We first consider $CV_1(k,\lambda)$.

Using arguments similar to those used to prove lemma A.4 in Ouyang, Li, and Li (2006), we can show that the leading term of $CV_1(k,\lambda)$ is given by $CV_{1,1}(k,\lambda)$, which is defined below

$$\begin{aligned} CV_1(k,\lambda) &= \frac{1}{n}\sum_{i=1}^{n}\left[g\left(X_i\right) - \hat{g}_{-i}\left(X_i\right)\right]^2 M(X_i) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left[g\left(X_i\right) - \hat{g}_{-i}\left(X_i\right)\right]^2 \hat{f}_{-i}(X_i)^2 M(X_i)/\hat{f}_{-i}(X_i)^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}\left[g\left(X_i\right) - \hat{g}_{-i}\left(X_i\right)\right]^2 \hat{f}_{-i}(X_i)^2 M(X_i)/f_{-i}(X_i)^2 + (s.o.) \\ &= CV_{1,1}(k,\lambda) + (s.o.), \end{aligned} \tag{5}$$

where $CV_{1,1}(k,\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left[g\left(X_i\right) - \hat{g}_{-i}\left(X_i\right)\right]^2 \hat{f}_{-i}(X_i)^2 M(X_i)/f_{-i}(X_i)^2$ and $(s.o.)$ denotes terms that have probability order smaller than that of $CV_{1,1}(k,\lambda)$.

We can show, using U-statistic H-decomposition arguments similar to those used to prove lemma A.6 in Ouyang, Li, and Li (2006), that the leading term of $CV_{1,1}(k,\lambda)$

is $E[CV_{1,1}(k,\lambda)]$. Following Ouyang, Li, and Li (2006), we define

$$\hat{m}_{1i} = \frac{1}{(n-1)R_i^q} \sum_{j \neq i}^{n} \left(g\left(X_j\right) - g\left(X_i\right)\right) w(\frac{X_j^c - X_i^c}{R_i}) L(X_i^d, X_j^d, \lambda) \tag{6}$$

and

$$\hat{m}_{2i} = \frac{1}{(n-1)R_i^q} \sum_{j \neq i}^{n} u_j w(\frac{X_j^c - X_i^c}{R_i}) L(X_i^d, X_j^d, \lambda) \tag{7}$$

Then $(\hat{g}_{-i}(X_i) - g(X_i))\hat{f}_{-i}(X_i) = \hat{m}_{1i} + \hat{m}_{2i}$. We have that

$$\begin{aligned} E[CV_{1,1}(k,\lambda)] =& E[(\hat{m}_{1i} + \hat{m}_{2i})^2 M(X_i)/f(X_i)^2] \\ =& E[\hat{m}_{1i}^2 M(X_i)/f(X_i)^2] + E[\hat{m}_{2i}^2 M(X_i)/f(X_i)^2] \end{aligned} \tag{8}$$

because $E[\hat{m}_{1i}\hat{m}_{2i} M(X_i)/f(X_i)^2] = 0$.

We consider the first term in (8). Using Lemma 2 we have ($\int dx = \sum_{x^d} \int dx^c$):

$$\begin{aligned} & E\left(\hat{m}_{i1}^2 | X_i = x, R_i = r\right) \\ = & \frac{1}{(n-1)^2 r^{2q}} \left[E\left(\sum_{j \neq i}^{n} \left(g\left(X_j\right) - g\left(x\right)\right) w\left(\frac{X_j^c - x^c}{r}\right) L(X_j^d, x^d, \lambda) \right)^2 | X_i = x, R_i = r \right] \\ = & \frac{(k-1)(k-2)}{(n-1)^2 r^{2q} G(r)^2} \left(\int f(x_j) \left[(g(x_j) - g(x))\, w\left(\frac{x_j^c - x^c}{r}\right) L(x_j^d, x^d, \lambda) \right] dx_j \right)^2 \\ + & \frac{k-1}{(n-1)^2 r^{2q} G(r)} \left(\int f(x_j) \left[(g(x_j) - g(x))^2 w^2 \left(\frac{x_j^c - x^c}{r}\right) L^2(x_j^d, x^d, \lambda) \right] dx_j \right) \\ \equiv & A_{1n}(x,r) + A_{2n}(x,r). \end{aligned}$$

We first consider $A_{1n}(x,r)$. We have that

$$\begin{aligned}
&\sum_{x_j^d}\int f(x_j^c,x_j^d)\left[(g(x_j)-g(x))\,w\left(\frac{x_j^c-x^c}{r}\right)L(x_j^d,x^d,\lambda)\right]dx_j^c\\
=&\quad\int f(x_j^c,x^d)\left[\Big(g(x_j^c,x^d)-g(x)\Big)\,w\left(\frac{x_j^c-x^c}{r}\right)\right]dx_j^c\\
+&\quad\sum_{x_j^d\neq x^d}\int f(x_j^c,x_j^d)\left[(g(x_j)-g(x))\,w\left(\frac{x_j^c-x^c}{r}\right)L(x_j^d,x^d,\lambda)\right]dx_j^c. \qquad (9)
\end{aligned}$$

We consider the first term in (9):

$$\begin{aligned}
&\int f(x_j^c,x^d)\Big(g(x_j,x^d)-g(x)\Big)\,w\left(\frac{x_j^c-x^c}{r}\right)dx_j^c\\
=&\quad r^q\int f\Big(x^c+rv,x^d\Big)\left(g\left(x^c+rv\right)-g\left(x\right)\right)w\left(v\right)dv\\
=&\quad r^q\int\left(f\left(x\right)+rv'\nabla f\left(x\right)\right)\left(\nabla g\left(x\right)rv+\frac{1}{2}r^2v'\nabla^2 g\left(x\right)v\right)w\left(v\right)dv[1+o(1)]+(s.o.)\\
=&\quad r^{q+2}c_w\left[f\left(x\right)tr[\nabla^2 g(x)]/2+\nabla f(x)'\nabla g(x)\right]+O(r^{q+4})\\
\equiv&\quad r^{q+2}B_{1,1}(x)+O(r^{q+4}),
\end{aligned}$$

where $c_w=\int w(v)v_s^2dv$ and the definition of $B_{1,1}(x)$ should be apparent.

We now consider the second term in (9):

$$\begin{aligned}
&\sum_{x_j^d\neq x^d}\int f(x_j^c,x_j^d)\left(g(x_j)-g(x)\right)w\left(\frac{x_j^c-x^c}{r}\right)L\left(x_j^d,x^d,\lambda\right)dx_j^c\\
&=\sum_{x_j^d\in D}\sum_{s=1}^{m}\lambda_s\mathbf{1}_s(x_j^d,x^d)\int f(x_j^c,x_j^d)\left(g(x_j)-g(x)\right)w\left(\frac{x_j^c-x^c}{r}\right)dx_j^c+O(\|\lambda\|^2r^q).
\end{aligned}$$

(10)

where $\mathbf{1}_s(z^d,x^d)\equiv\mathbf{1}(z_s^d\neq x_s^d)\prod_{t\neq s}\mathbf{1}(z_t^d=x_t^d)$.

We have that, for $x_j^d \neq x^d$,

$$\begin{aligned}
&\int f\left(x_j^c, x_j^c\right)\left(g\left(x_j^c, x_j^d\right) - g(x)\right) w\left(\frac{x_j^c - x^c}{r}\right) dx_j^c \\
&= r^q \int f\left(x^c + rv, x_j^d\right)\left(g\left(x_j^c + rv, x_j^d\right) - g(x)\right) w(v) dv \\
&= r^q \int (f(x^c, x_j^d) + rv'\nabla f(x^c, x_j^d)) \Big[(g(x^c, x_j^d) - g(x)) + \nabla g(x^c, x_j^d) rv \\
&\qquad\qquad + \frac{1}{2} r^2 v' \nabla^2 g(x^c, x_j^d) v\Big] w(v)\, dv + (s.o.) \qquad (11) \\
&= r^q f(x^c, x_j^d)\left[\left(g(x^c, x_j^d) - g(x)\right) + c_w r^2 \left(tr[\nabla^2 g(x)]/2 + \nabla f(x)' \nabla g(x)\right)\right] + (s.o.) \\
&= r^q f(x^c, x_j^d)(g(x^c, x_j^d) - g(x)) + O(r^{q+2}).
\end{aligned}$$

Substituting (11) into (10), we have

$$\begin{aligned}
&\int f(x_j)\left(g(x_j) - g(x)\right) w\left(\frac{x_j^c - x^c}{r}\right) L\left(x_j^d, x^d, \lambda\right) dx_j \\
= &\sum_{x_j^d \in D}\left(\sum_{s=1}^m \lambda_s \mathbf{1}_s(x_j^d, x^d)\left(r^q f(x^c, x_j^d)\left(g(x^c, x_j^d) - g(x)\right) + O\left(r^{q+2}\right)\right)\right) + O\left(\|\lambda\|^2 r^q\right) \\
= &\ r^q \sum_{x_j^d \in D}\left(\sum_{s=1}^m \lambda_s \mathbf{1}_s(x_j^d, x^d) f(x^c, x_j^d)\left(g(x^c, x_j^d) - g(x)\right)\right) + O\left(\|\lambda\| r^{q+2} + \|\lambda\|^2 r^q\right) \\
\equiv &\ r^q B_{1,2}(x, \lambda) + O\left(\|\lambda\| r^{q+2} + \|\lambda\|^2 r^q\right), \qquad (12)
\end{aligned}$$

where the definition of $B_{1,2}(x)$ should be apparent.

We now consider $A_{2n}(x, r)$:

$$\begin{aligned}
&\sum_{x_j} \int f(x_j)\left(g(x_j) - g(x)\right)^2 w^2\left(\frac{x_j^c - x^c}{r}\right) L^2(x_j^d, x^d, \lambda) dx_j^c \\
&= \int f(x_j^c, x^d)(g(x_j^c, x^d) - g(x))^2 w^2\left(\frac{x_j^c - x^c}{r}\right) dx_j^c \\
&+ \sum_{x_j^d \neq x^d} \int f(x_j^c, x_j^d)(g(X_j) - g(x))^2 w^2\left(\frac{X_j^c - x^c}{r}\right) dx_j^c. \qquad (13)
\end{aligned}$$

We consider the first term in (13).

$$\begin{aligned}
&\int f(x_j^c, x^d)(g(x_j^c, x^d) - g(x))^2 w^2 \left(\frac{x_j^c - x^c}{r}\right) dx_j^c \\
= \quad & r^q \int f(x^c + rv, x^d)(g(x^c + rv, x^d) - g(x))^2 w^2(v) dv \\
&= r^{q+2} \nu_w f(x) [\sum_{s=1}^{q} g_s(x)]^2 + O(r^{q+4}) \\
&\equiv r^{q+2} B_3 + O(r^{q+4}), \qquad (14)
\end{aligned}$$

where $B_3 = \nu_w f(x)[\sum_{s=1}^{q} g_s(x)]^2$ and $\nu_w = \int w^2(v) v_s^2 dv$.

Now we consider the second term in (13)

$$\begin{aligned}
&\sum_{x_j^d \neq x^d} \int f(x_j^c, x_j^d)(g(x_j) - g(x))^2 w^2 \left(\frac{x_j^c - x^c}{r}\right) L^2(x_j^d, x^d, \lambda) dx_j^c \\
&= \sum_{x_j^d \in D} \left(\sum_{s=1}^{m} \lambda_s^2 \mathbf{1}_s(x_j^d, x^d) \int f(x_j^c, x_j^d) \left(g(x_j) - g(x)\right)^2 w^2 \left(\frac{x_j^c - x^c}{r}\right) dx_j^c \right) + O(r^q \|\lambda\|^3) \\
&= \quad \sum_{x_j^d \in D} \left(\sum_{s=1}^{m} \lambda_s^2 \mathbf{1}_s(x_j^d, x^d) \left[r^q f(x^c, x_j^d) \left(g(x^c, x_j^d) - g(x) \right)^2 + O(r^{q+2}) \right] \right) + O(r^q \|\lambda\|^3) \\
&= \quad r^q \sum_{x_j^d \in D} \left(\sum_{s=1}^{m} \lambda_s^2 \mathbf{1}_s(x_j^d, x^d) f(x^c, x_j^d) \left(g(x^c, x_j^d) - g(x) \right)^2 \right) + O(\|\lambda\|^2 r^{q+2} + r^q \|\lambda\|^3) \\
&\equiv r^q B_4(\lambda) + O(\|\lambda\|^2 r^{q+2} + r^q \|\lambda\|^3), \qquad (15)
\end{aligned}$$

where the definition of $B_4(\lambda)$ should be obvious.

Substituting from (11) and (12) and using Lemma 1, we have that

$$
\begin{aligned}
&E\left[A_{1n}(X_i,R_i)\mid X_i\right]\\
=\quad & E\left[\frac{k^2}{n^2R_i^{2q}G(R_i)^2}\left(R_i^{q+2}B_{1,1}(X_i)+R_i^qB_{1,2}(X_i,\lambda)\right)^2\mid X_i\right]+(s.o.)\\
=\quad & \frac{k^2}{n^2}B_{1,1}^2(X_i)E\left[\frac{R_i^4}{G(R_i)^2}\mid X_i\right]+\frac{k^2}{n^2}B_{1,2}^2(X_i,\lambda)E\left[\frac{1}{G(R_i)^2}\mid X_i\right]\\
+\quad & 2\left(\frac{k^2}{n^2}B_{1,1}(X_i)B_{1,2}(X_i,\lambda)E\left[\frac{R_i^2}{G(R_i)^2}\mid X_i\right]\right]+(s.o.)\\
=\quad & \left(\frac{1}{c_0f(X_i)}\right)^{\frac{4}{q}}B_{1,1}^2(X_i)\left(\frac{k}{n}\right)^{\frac{4}{q}}+B_{1,2}^2(X_i,\lambda)\\
+\quad & 2\left(\frac{1}{c_0f(X_i)}\right)^{\frac{2}{q}}B_{1,1}(X_i)B_{1,2}(X_i,\lambda)\left(\frac{k}{n}\right)^{\frac{2}{q}}+(s.o.)\\
=&\left(\frac{1}{c_0f(X_i)}\right)^{\frac{4}{q}}B_{1,1}^2(X_i)\left(\frac{k}{n}\right)^{\frac{4}{q}}+B_{1,2}^2(X_i,\lambda)+2B_{1,1}(X_i)B_{1,2}(X_i,\lambda)\left(\frac{1}{c_0f(X_i)}\right)^{\frac{2}{q}}\left(\frac{k}{n}\right)^{\frac{2}{q}}\\
&+O\left(\left(\frac{k}{n}\right)^{\frac{6}{q}}+\|\lambda\|\left(\frac{k}{n}\right)^{\frac{4}{q}}+\|\lambda\|^2\left(\frac{k}{n}\right)^{\frac{2}{q}}+\|\lambda\|^3\right).
\end{aligned}
$$

Substituting from (14) and (15) and using Lemma 1, we have that

$$
\begin{aligned}
&E\left[A_{2n}(X_i,R_i)\mid X_i\right]\\
=\quad & E\left[\frac{k}{n^2R_i^{2q}G(R_i)}\left(R_i^{q+2}B_3+R_i^2B_4(\lambda)\right)\mid X_i\right]+(s.o.)\\
=\quad & \frac{k}{n^2}B_3E\left[\frac{1}{R_i^{q-2}G(R_i)}\mid X_i\right]+\frac{k}{n^2}B_4(\lambda)E\left[\frac{1}{R_i^qG(R_i)}\mid X_i\right]+(s.o.)\\
=\quad & \left(c_0f(X_i)\right)^{\frac{q-2}{q}}B_3\frac{1}{k}\left(\frac{k}{n}\right)^{\frac{2}{q}}+\left(c_0f(X_i)\right)B_4(\lambda)\frac{1}{k}+(s.o.) \qquad (16)
\end{aligned}
$$

We now consider $E[\hat{m}_{2i}^2 M(X_i)/f(X_i)^2]$. We have that

$$E\left(\hat{m}_{i2}^2|X_i=x,R_i=r\right)= \tag{17}$$
$$\frac{1}{\left(n-1\right)^2 r^{2q}}E\left(\sum_{j\neq i}^{n}u_j^2 w^2\left(\frac{X_j^c-X_i^c}{R_i}\right)L^2(X_j^d,x^d,\lambda)|X_i=x,R_i=r\right).$$

We note that

$$\begin{aligned}
& E\left(\sum_{j\neq i}^{n}u_j^2 w^2\left(\frac{X_j^c-X_i^c}{R_i}\right)L^2(X_j^d,x^d,\lambda)|X_i=x,R_i=r\right)\\
= \quad & \frac{k-1}{G(r)}\sum_{x_j^d\in D}\left(L(x_j^d,x^d,\lambda)^2\int f(x_j^c,x_j^d)\sigma^2(x_j^c,x_j^d)w^2\left(\frac{x_j^c-x^c}{r}\right)dx_j^c\right)\\
= \quad & \frac{k-1}{G(r)}\int f(x_j^c,x^d)\sigma^2(x_j^c,x^d)w^2\left(\frac{x_j^c-x^c}{r}\right)dx_j^c\\
+ \quad & \frac{k-1}{G(r)}\sum_{x_j^d\in D}\sum_{s=1}^{m}\left(\lambda_s^2\mathbf{1}_s(x_j^d,x^d)\int f(x_j^c,x_j^d)\sigma^2(x_j^c,x_j^d)w^2\left(\frac{x_j^c-x^c}{r}\right)dx_j^c\right)+(s.o.).
\end{aligned} \tag{18}$$

We note that

$$\begin{aligned}
& \int f(x_j^c,x^d)\sigma^2(x_j^c,x^d)w^2\left(\frac{x_j^c-x^c}{r}\right)dx_j^c\\
= \quad & r^q\int f\left(x^c+rv,x^d\right)\sigma^2\left(x^c+rv,x^d\right)w^2\left(v\right)dv\\
= \quad & r^q f\left(x\right)\sigma^2\left(x\right)\int w^2\left(v\right)dv[1+O(r^2)]\\
& = r^q d_w\sigma^2\left(x\right)f\left(x\right)\left[1+O(r^2)\right]
\end{aligned} \tag{19}$$

where $d_w=\int w^2\left(v\right)dv$.

Substituting (19) into the second term in (18), we obtain

$$\sum_{x_j^d \in D} \sum_{s=1}^{m} \left(\lambda_s^2 \mathbf{1}_s(x_j^d, x^d) \int f(x_j^c, x_j^d) \sigma^2(x_j^c, x_j^d) w^2 \left(\frac{x_j^c - x^c}{r} \right) dx_j^c \right)$$
$$= \sum_{x_j^d \in D} \sum_{s=1}^{m} r^q \left(\lambda_s^2 \mathbf{1}_s(x_j^d, x^d) \sigma^2 \left(x^c, x_j^d \right) f \left(x^c, x_j^d \right) \int w^2 (v)\, dv [1 + o(1)] \right) = O(\|\lambda\|^2 r^q). \tag{20}$$

Combining the results from above, we have that

$$E \left[\hat{m}_{i2}^2 | X_i = x, R_i = r \right] = \frac{k-1}{(n-1)^2 r^q G(r)} \left(\sigma^2 (x) f (x) \int w^2 (v)\, dv + o(1) \right)$$

Thus, using Lemma 1, we have that

$$\begin{aligned} E \left[\hat{m}_{i2}^2 | X_i \right] &= \frac{k}{n^2} \left(\sigma^2 (X_i) f (X_i) \int w^2 (v)\, dv \right) E \left[\frac{1}{R_i^q G(R_i)} \mid X_i \right] \\ &= d_w \sigma^2(X_i) f(X_i) \left(c_0 f(X_i) \right) \frac{1}{k} + (s.o.) \\ &= c_0 d_w \sigma^2(X_i) f(X_i)^2 \left(\frac{1}{k} \right) + O \left(\left(\frac{1}{k} \right) \left(\frac{k}{n} \right)^{\frac{2}{q}} + \|\lambda\|^2 \left(\frac{1}{k} \right) \right). \end{aligned}$$

Combining the results from above, we have shown that

$$
\begin{aligned}
&E[CV_{1,1}(k,\lambda)] = E\left[\left(\hat{m}_{i1}^2+\hat{m}_{i2}^2\right)f(X_i)^{-2}M(X_i)\right] \\
&= \left(\frac{1}{c_0}\right)^{\frac{4}{q}} E\left[B_{1,1}^2(X_i)f(X_i)^{\frac{-4}{q}}M(X_i)\right]\left(\frac{k}{n}\right)^{\frac{4}{q}} + c_0 d_w E\left[\sigma^2(X_i)M(X_i)\right]\left(\frac{1}{k}\right) \\
&+ E\left[B_{1,2}^2(X_i,\lambda)f(X_i)^{-2}M(X_i)\right] \\
&+ 2\left(\frac{1}{c_0}\right)^{\frac{2}{q}} E\left[B_{1,1}(X_i)B_{1,2}(X_i,\lambda)f(X_i)^{-\frac{2+2q}{q}}M(X_i)\right]\left(\frac{k}{n}\right)^{\frac{2}{q}} \\
&+ O\left(\left(\frac{k}{n}\right)^{\frac{6}{q}} + \left(\frac{1}{k}\right)\left(\frac{k}{n}\right)^{\frac{2}{q}} + \|\lambda\|\left(\frac{k}{n}\right)^{\frac{4}{q}} + \|\lambda\|^2\left(\frac{k}{n}\right)^{\frac{2}{q}} + \|\lambda\|^3\right) \\
&\equiv B_1\left(\frac{k}{n}\right)^{\frac{4}{q}} + B_2\left(\frac{1}{k}\right) + B_3(\lambda) + B_4(\lambda)\left(\frac{k}{n}\right)^{\frac{2}{q}} \\
&+ O\left(\left(\frac{k}{n}\right)^{\frac{6}{q}} + \frac{1}{k}\left(\frac{k}{n}\right)^{\frac{2}{q}} + \|\lambda\|\left(\frac{k}{n}\right)^{\frac{4}{q}} + \|\lambda\|^2\left(\frac{k}{n}\right)^{\frac{2}{q}} + \|\lambda\|^3\right).
\end{aligned}
$$

Lemma 3 $CV_2(k,\lambda) = (nk)^{-1/2}\mathcal{C}_1 + n^{-1/2}(k/n)^{2/q}\mathcal{C}_2 + n^{-1/2}\mathcal{C}_3(\lambda)$, *where* $\mathcal{C}_1$ *and* $\mathcal{C}_2$ *are* $O_p(1)$ *and* $\mathcal{C}_3(\lambda)$ *is* $O_p(\|\lambda\|)$*. Therefore,* $CV_2(k,\lambda)$ *has a probability order smaller than that of* $CV_1(k,\lambda)$*.*

Let $w_{R_i,ij}$ denote $R_i^{-q}w\left(\frac{X_i^c-X_j^c}{R_i}\right)$ and let $l_{\lambda,ij}$ denote $L(X_i^d, X_j^d, \lambda)$. We can show that $CV_2(k,\lambda) = CV_{2,1}(k,\lambda)[1+o_p(1)]$, where $CV_{2,1}(k,\lambda) = n^{-1}\sum_i u_i(g_i-\hat{g}_{-i})\hat{f}_{-i}/f_i$. We can write $CV_{2,1}(k,\lambda)$ as $CV_{2,1}(k,\lambda) = B_{1n}+B_{2n}$, where $B_{1n} = [n(n-1)]^{-1}\sum_i\sum_{j\neq i} u_i(g_i-$

$g_j)w_{R_i,ij}l_{\lambda,ij}/f_{-i}$ and $B_{2n} = [n(n-1)]^{-1}\sum_i\sum_{j\neq i} u_iu_jw_{R_i,ij}\ l_{\lambda,ij}/f_{-i}$.

$$
\begin{aligned}
E[B_{1n}^2] = \frac{1}{n^2(n-1)^2}\sum_i\sum_{j\neq i}\sum_{l\neq i} E\left[u_i^2(g_i-g_j)w_{R_i,ij}l_{\lambda,ij}w_{R_i,il}l_{\lambda,il}/f_i^2\right] \\
&= \frac{1}{n^2(n-1)^2}\sum_i\sum_{j\neq i} E\left[\sigma^2(X_i)(g_i-g_j)^2w_{R_i,ij}^2l_{\lambda,ij}^2/f_i^2\right] \\
&+ \frac{1}{n^2(n-1)^2}\sum_i\sum_{j\neq i}\sum_{l\neq i,l\neq j} E\left[\sigma^2(X_i)(g_i-g_j)w_{R_i,ij}l_{\lambda,ij}(g_i-g_l)w_{R_i,il}l_{\lambda,il}/f_i^2\right] \\
&\equiv B_{1n,1}+B_{1n,2}.
\end{aligned}
$$

We first consider $B_{1n,1}$:

$$
\begin{aligned}
& E\left[(g(X_i)-g(X_j))^2w_{R_i,ij}^2l_{\lambda,ij}^2|X_i,R_i\right] \\
&= \int f(x_j^c,x^d)(g(x_i)-g(x_j^c,x^d))^2w_{R_i,ij}^2l_{\lambda,ij}^2dx_j^c \\
&+ \sum_{x_j\neq x^d} f(x_j^c,x_j^d)(g(x_i)-g(x_j^c,x^d))^2w_{R_i,ij}^2l_{\lambda,ij}^2dx_j^c \\
&= O\left(\frac{R_i^2}{R_i^q}\right)+O\left(\frac{\|\lambda\|^2}{R_i^q}\right).
\end{aligned}
$$

Thus $B_{1n,1} = n^{-2}O\left(E[R_i^{2-q}]+\|\lambda\|^2E(R_i^{-q})\right) = O(n^{-2}(k/n)^{\frac{2-q}{q}}+\|\lambda\|^2n^{-2}(k/n)^{-1}) = O((nk)^{-1}(k/n)^{2/q}+\|\lambda\|^2n^{-2}(k/n)^{-1})$ using Lemma 1.

Next we consider $B_{1n,2}$:

$$
\begin{aligned}
& E\left[(g_i-g_j)w_{R_i,ji}l_{\lambda,ij}|X_i,R_i\right] = \\
& \int f(x_j^c,x^d)(g(X_i)-g(x_j^c,x^d))w\left(\frac{X_i^c-x_j^c}{R_i}\right)dx_j^c \\
&+ \sum_{x_j^d\neq x^d}\int f(x_j^c,x_j^d)(g(X_i)-g(x_j^c,x_j^d))w\left(\frac{X_i^c-x_j^c}{R_i}\right)L(X_i^d,x_j^d,\lambda)dx_j^c \\
&= O\left(R_i^2\right)+O\left(\|\lambda\|\right).
\end{aligned}
$$

Thus $B_{1n,2} = n^{-1}O\left(E(R_i^4) + \|\lambda\|^2\right) = O\left(n^{-1}(k/n)^{4/q} + n^{-1}\|\lambda\|^2\right) =$

$O\left(n^{-1}(k/n)^{4/q} + n^{-1}\|\lambda\|^2\right)$ by using Lemma 1.

Thus, combining these results, we have that

$E\left[B_{1n}^2\right] = B_{1n,1} + B_{1n,2} =$

$O\left((nk)^{-1}(k/n)^{2/q} + n^{-1}(k/n)^{4/q} + \|\lambda\|^2 n^{-2}(k/n)^{-1} + n^{-1}\|\lambda\|^2\right)$

$= O\left((nk)^{-1}(k/n)^{2/q} + n^{-1}(k/n)^{4/q} + +n^{-1}\|\lambda\|^2\right).$

It follows that $B_{1n} = O_p\left(n^{-1/2}(k/n)^{2/q} + n^{-1/2}\|\lambda\|\right)$.

We now consider B_{2n}:

$$
\begin{aligned}
B_{2n} = & \quad \frac{1}{n(n-1)} \sum_i \sum_{j>i} u_i u_j \left[w_{R_i,ij} l_{\lambda,ij} + w_{R_j,ji} l_{\lambda,ij}\right]. \\
E[B_{2n}^2] = & \quad \frac{1}{n^2(n-1)^2} \sum_i \sum_{j>i} E\left\{u_i^2 u_j^2 \left[w_{R_i,ij} l_{\lambda,ij} + w_{R_j,ji} l_{\lambda,ij}\right]^2\right\} \\
= & \quad \frac{1}{n^2(n-1)^2} \sum_i \sum_{j>i} E\left\{\sigma^2(X_i)\sigma^2(X_j) \left[w_{R_i,ij} l_{\lambda,ij} + w_{R_j,ji} l_{\lambda,ij}\right]^2\right\} \\
\leq & \quad \frac{4}{n^2(n-1)^2} \sum_i \sum_{j>i} E\left[\sigma^2(X_i)\sigma^2(X_j) w_{R_i,ij}^2 l_{\lambda,ij}^2\right] \\
= & \quad n^{-2} O(E[R_i^{-q}]) = O(n^{-2}(k/n)^{-1}) = O((nk)^{-1})
\end{aligned}
$$

using Lemma 1.

Thus, we have that $B_{2n} = O_p\left((nk)^{-1/2}\right)$ and so

$B_{1n} + B_{2n} = O_p\left((nk)^{-1/2} + n^{-1/2}(k/n)^{2/q} + n^{-1/2}\|\lambda\|\right)$. It follows that we can write $CV_2(k,\lambda)$ as $CV_2(k,\lambda) = (nk)^{-1/2}\mathcal{C}_1 + n^{-1/2}(k/n)^{2/q}\mathcal{C}_2 + n^{-1/2}\mathcal{C}_3(\lambda)$, where $\mathcal{C}_1 = O_p(1)$, $\mathcal{C}_2 = O_p(1)$, and $\mathcal{C}_3(\lambda) = O_p(\|\lambda\|)$.

Proof of Theorem 3

We have that $CV(k,\lambda) = CV_1(k,\lambda) + CV_2(k,\lambda) +$ terms unrelated to (k,λ), and

$$CV_1(k,\lambda) = B_1\left(\frac{k}{n}\right)^{\frac{4}{q}} + B_2\frac{1}{k} + B_3(\lambda) + B_4(\lambda)\left(\frac{k}{n}\right)^{\frac{2}{q}}$$
$$+ \quad O\left(\left(\frac{k}{n}\right)^{\frac{6}{q}} + \frac{1}{k}\left(\frac{k}{n}\right)^{\frac{2}{q}} + \|\lambda\|\left(\frac{k}{n}\right)^{\frac{4}{q}} + \|\lambda\|^2\left(\frac{k}{n}\right)^{\frac{2}{q}} + \|\lambda\|^3\right).$$

Let $CV_0(k,\lambda)$ denote the leading term of $CV_1(k,\lambda)$. Then $CV_0(k,\lambda) = B_1\left(\frac{k}{n}\right)^{\frac{4}{q}} + B_2\left(\frac{1}{k}\right) + B_3(\lambda) + B_4(\lambda)\left(\frac{k}{n}\right)^{\frac{2}{q}}$. Let k_0 and λ_0 denote, respectively, the values of k and λ that minimize $CV_0(k,\lambda)$. In order to derive first-order conditions for k_0 and λ_0, we can first derive simple expressions for $B_3(\lambda)$ and $B_4(\lambda)$: We can write $B_{1,2}(x,\lambda)$ as:

$$\begin{aligned} B_{1,2}(x,\lambda) = & \sum_{z^d\in D}\left(\sum_{s=1}^{m}\lambda_s\mathbf{1}_s(z^d,x^d)f(x^c,z^d)\left(g(x^c,z^d)-g(x)\right)\right) \\ = & \sum_{s=1}^{m}\lambda_s\left(\sum_{z^d\in D}\mathbf{1}_s(z^d,x^d)f(x^c,z^d)\left(g(x^c,z^d)-g(x)\right)\right) \\ \equiv & \sum_{s=1}^{m}\lambda_s b_s(x), \end{aligned}$$

where the definition of $b_s(x)$ should be apparent.

Then

$$\begin{aligned} B_4(\lambda) \equiv & \; 2\left(\frac{1}{c_0}\right)^{2/q} E\left[B_{1,1}(X_i)B_{1,2}(X_i,\lambda)\right) \\ = & \; 2\left(\frac{1}{c_0}\right)^{2/q} E\left[B_{1,1}(X_i)\left(\sum_{s=1}^{m}\lambda_s b_s(X_i)\right)f(X_i)^{-\frac{2}{q}}M(X_i)\right] \\ = & \; 2\left(\frac{1}{c_0}\right)^{2/q}\sum_{s=1}^{m}\lambda_s\left(E\left[B_{1,1}(X_i)b_s(X_i)f(X_i)^{-\frac{2}{q}}M(X_i)\right]\right) \\ \equiv & \sum_{s=1}^{m}\lambda_s c_s, \end{aligned} \tag{21}$$

where the definition of c_s should be apparent.

Similarly, we can write $B_3(\lambda)$ as follows:

$$
\begin{aligned}
B_3(\lambda) &= E\left[B_{1,2}^2(X_i,\lambda)f(X_i)^{-2}M(X_i)\right] \\
&= E\left[\left(\sum_{s=1}^{m}\lambda_s c_s(X_i)\right)^2 f(X_i)^{-2}M(X_i)\right] \\
&= \sum_{s=1}^{m}\lambda_s^2 E\left[c_s^2(X_i)f(X_i)^{-2}M(X_i)\right] \\
&+ 2\sum_{s=1}^{m-1}\sum_{t>s}\lambda_s\lambda_t E\left[c_s(X_i)c_t(X_i)f(X_i)^{-2}M(X_i)\right] \\
&\equiv \sum_{s=1}^{m}\lambda_s^2 d_s + 2\sum_{s=1}^{m-1}\sum_{t>s}\lambda_s\lambda_t d_{ts},
\end{aligned} \tag{22}
$$

where the definitions of d_s and d_{ts} should be obvious.

Using (21) and (22), we can derive the following first-order conditions for k_0 and λ_0:

$$
\begin{aligned}
\frac{\partial CV_0}{\partial k} &= \left(\frac{4}{q}\right)B_1\frac{1}{k}\left(\frac{k}{n}\right)^{\frac{4}{q}} - B_2\frac{1}{k^2} + \left(\frac{2}{q}\right)B_4(\lambda)\frac{1}{k}\left(\frac{k}{n}\right)^{\frac{2}{q}} \\
&= 0, \frac{\partial CV_0}{\partial \vec{\lambda}} \\
&= 2\begin{pmatrix} d_1 & d_{12} & \cdots & d_{1m} \\ d_{21} & d_2 & & d_{2m} \\ \vdots & & \ddots & \\ d_{m1} & d_{m2} & & d_m \end{pmatrix}\begin{bmatrix}\lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_m\end{bmatrix} + \left(\frac{k}{n}\right)^{\frac{2}{q}}\begin{bmatrix}c_1 \\ c_2 \\ \vdots \\ c_m\end{bmatrix} \\
&= 0.
\end{aligned} \tag{23}
$$

After rewriting (23) as $A\lambda = -(k/n)^{2/q}c$, where the definition of the $m \times m$ matrix A should be clear from (23) , and c is an $m \times 1$ vector whose i^{th} element is c_i , we can solve to get $\lambda = -(k/n)^{2/q}A^{-1}c$. Thus, recalling that $B_4(\lambda) = \sum_{s=1}^{m}\lambda_s c_s$, we have that

$B_4(\lambda_0) = a_0(k_0/n)^{2/q}$ for some constant a_0. Substituting back into (23), we obtain

$$\left(\frac{4}{q}\right) B_1 \frac{1}{k_0} \left(\frac{k_0}{n}\right)^{\frac{4}{q}} - B_2 \frac{1}{k_0^2} + a_0 \left(\frac{2}{q}\right) \frac{1}{k_0} \left(\frac{k_0}{n}\right)^{\frac{4}{q}} = 0 \tag{24}$$

Solving for k_0, we obtain

$$\begin{aligned} k_0 &= \left(\frac{B_2 n^{\frac{4}{q}}}{\left(\frac{4}{q}\right) B_1 + a_0 \left(\frac{2}{q}\right)} \right)^{\frac{q}{4+q}} \\ &\equiv \quad b_1 n^{4/(4+q)}, \end{aligned} \tag{25}$$

where $b_1 = \left(\frac{(B_2)^{\frac{4}{q}}}{\left(\frac{4}{q}\right)B_1 + a_0\left(\frac{2}{q}\right)} \right)^{\frac{q}{4+q}}$. From $\lambda_0 = (k_0/n)^{2/q} A^{-1} c$ and (25), we have that $\lambda_0 = n^{-\frac{2}{4+q}} A^{-1} c$.

Let $\hat{k}$ and $\hat{\lambda}$ denote the values of k and λ that minimize $CV(k, \lambda)$. When $q \leq 3$,

$$CV(k, \lambda) =$$

$$B_1 \left(\frac{k}{n}\right)^{\frac{4}{q}} + B_2 \frac{1}{k} + B_3(\lambda) + B_4(\lambda) \left(\frac{k}{n}\right)^{\frac{2}{q}} + (nk)^{-1/2} \mathcal{C}_1 + n^{-\frac{1}{2}} \left(\frac{k}{n}\right)^{\frac{2}{q}} \mathcal{C}_2 + n^{-\frac{1}{2}} \mathcal{C}_3(\lambda) + (s.o.)$$

Noting that $\mathcal{C}_3(\lambda)$ can be written as $\mathcal{C}_3(\lambda) = \sum_{s=1}^{m} b_s \lambda_s$, we can derive the following first-order conditions:

$$\begin{aligned} &\frac{4}{q} B_1 \left(\frac{\hat{k}}{n}\right)^{\frac{4+q}{q}} - B_2 n^{-1} + \frac{2}{q} B_4(\hat{\lambda}) \left(\frac{\hat{k}}{n}\right)^{\frac{2+q}{q}} \\ &- \frac{1}{2} \mathcal{C}_1 n^{-1} \left(\frac{\hat{k}}{n}\right)^{\frac{1}{2}} \frac{2}{q} \mathcal{C}_2 n^{-1/2} \left(\frac{\hat{k}}{n}\right)^{\frac{2+q}{q}} + (s.o.) = 0, \\ &A\hat{\lambda} + \left(\frac{\hat{k}}{n}\right)^{\frac{2}{q}} c + n^{-1/2} b + (s.o.) = 0, \end{aligned} \tag{26}$$

where A was defined above, and b is an $m \times 1$ vector whose i^{th} element is b_i. From (26) we have that $\hat{\lambda} = (\hat{k}/n)^{2/q} A^{-1}c + n^{-1/2}A^{-1}b + (s.o.)$. Then, recalling that $B_4(\lambda) = \sum_{s=1}^{m} \lambda_s c_s$, we have that $B_4(\hat{\lambda}) = a_0(\hat{k}/n)^{2/q} + a_1 n^{-1/2} + (s.o.)$, where a_0 was defined above and a_1 is another constant. Substituting for $B_4(\hat{\lambda})$ into (26), we obtain

$$\left(\frac{4}{q}\right) B_1 \left(\frac{\hat{k}}{n}\right)^{\frac{4+q}{q}} - B_2 n^{-1} + \left(\frac{2}{q}\right) a_0 \left(\frac{\hat{k}}{n}\right)^{\frac{4+q}{q}} + \left(\frac{2}{q}\right) a_1 n^{-\frac{1}{2}} \left(\frac{\hat{k}}{n}\right)^{\frac{4+q}{q}}$$
$$- \frac{1}{2}\mathcal{C}_1 n^{-1} \left(\frac{\hat{k}}{n}\right)^{\frac{1}{2}} + \left(\frac{2}{q}\right) \mathcal{C}_2 n^{-1/2} \left(\frac{\hat{k}}{n}\right)^{\frac{2+q}{q}} + (s.o.) = 0 \qquad (27)$$

Let $\hat{k} = k_0 + \hat{k}_1$. Then, since $CV_0(\lambda, k)$ is the leading term of $CV(\lambda, k)$, we have that $\frac{\hat{k}_1}{k_0} = o_p(1)$ and $\frac{\hat{k}_1}{n} = o_p(1)$. By Taylor's Theorem

$$\left(\frac{\hat{k}}{n}\right)^{\frac{4+q}{q}} = \left(\frac{k_0}{n}\right)^{\frac{4+q}{q}} + \frac{4+q}{q} \left(\frac{k_0}{n}\right)^{\frac{4}{q}} \left(\frac{\hat{k}_1}{n}\right) + (s.o.). \qquad (28)$$

Substituting (28) into (27) and using (24), we obtain

$$\frac{4+q}{q} \left[\left(\frac{4}{q}\right) B_1 + \left(\frac{2}{q}\right) a_0\right] \left(\frac{k_0}{n}\right)^{\frac{4}{q}} \frac{\hat{k}_1}{n} - \frac{1}{2}\mathcal{C}_1 n^{-1} \left(\frac{k_0}{n}\right)^{\frac{1}{2}}$$
$$+ \left(\frac{2}{q}\right) \mathcal{C}_2 n^{-1/2} \left(\frac{k_0}{n}\right)^{\frac{2+q}{q}} + (s.o.) = 0.$$

Solving for $\hat{k}_1$, we obtain

$$\frac{\hat{k}_1}{n} = \frac{-\frac{1}{2}\mathcal{C}_1 n^{-1}\left(\frac{k_0}{n}\right)^{\frac{q-8}{2q}} + \left(\frac{2}{q}\right)\mathcal{C}_2 n^{-1/2}\left(\frac{k_0}{n}\right)^{\frac{q-2}{q}}}{\left(\frac{4+q}{q}\right)\left[\left(\frac{4}{q}\right)B_1 + \left(\frac{2}{q}\right)a_0\right]}. \tag{29}$$

Then

$$\frac{\hat{k}_1}{k_0} = \frac{-\frac{1}{2}\mathcal{C}_1 n^{-1}\left(\frac{k_0}{n}\right)^{-\frac{q+8}{2q}} + \left(\frac{2}{q}\right)\mathcal{C}_2 n^{-1/2}\left(\frac{k_0}{n}\right)^{-\frac{2}{q}}}{\left(\frac{4+q}{q}\right)\left[\left(\frac{4}{q}\right)B_1 + \left(\frac{2}{q}\right)a_0\right]}. \tag{30}$$

Recalling that $\hat{k}_1 = \hat{k} - k_0$, and using $k_0 = b_1 n^{4/(4+q)}$ we have

$$\frac{\hat{k} - k_0}{k_0} = O_p\left(n^{-1}\left(\frac{k_0}{n}\right)^{-\frac{q+8}{2q}} + n^{-1/2}\left(\frac{k_0}{n}\right)^{-\frac{2}{q}}\right) = O_p\left(n^{-\frac{q}{2(4+q)}}\right) \tag{31}$$

From (26) , we have that

$$\begin{aligned}
\hat{\lambda} &= \left(\frac{\hat{k}}{n}\right)^{\frac{2}{q}} A^{-1}c + n^{-1/2}A^{-1}b + (s.o.) \\
&= \left(\frac{k_0}{n} + \frac{\hat{k}_1}{n}\right)^{\frac{2}{q}} A^{-1}c + n^{-1/2}A^{-1}b + (s.o.) \\
&= \left[\left(\frac{k_0}{n}\right)^{\frac{2}{q}} + \frac{2}{q}\left(\frac{k_o}{n}\right)^{\frac{2}{q}-1}\frac{\hat{k}_1}{n}\right] A^{-1}c + n^{-1/2}A^{-1}b + (s.o.) \\
&= \lambda_0 + \frac{2}{q}\left(\frac{k_o}{n}\right)^{\frac{2}{q}-1}\left(\frac{\hat{k}_1}{n}\right) A^{-1}c + n^{-1/2}A^{-1}b + (s.o.) \qquad (32) \\
&= \lambda_0 + O_p(n^{-1/2}) \qquad (33)
\end{aligned}$$

since $(k_0/n)^{(2/q)-1}(\hat{k}_1/n) = O_p(n^{-1/2})$ from (29) and (25).

When $q \geq 5$,

$$CV(k,\lambda) = B_1\left(\frac{k}{n}\right)^{\frac{4}{q}} + B_2\left(\frac{1}{k}\right) + B_3(\lambda) + B_4(\lambda)\left(\frac{k}{n}\right)^{\frac{2}{q}} + D_1\left(\frac{k}{n}\right)^{\frac{6}{q}} + D_2\left(\frac{1}{k}\right)\left(\frac{k}{n}\right)^{\frac{2}{q}}$$
$$+ \quad D_3(\lambda)\left(\frac{k}{n}\right)^{\frac{4}{q}} + D_4(\lambda)\left(\frac{k}{n}\right)^{\frac{2}{q}} + D_5(\lambda) + (s.o.),$$

where $D_3(\lambda)$, $D_4(\lambda)$, and $D_5(\lambda)$ are polynomials in λ of order one, two, and three, respectively. Suppose for simplicity that $\lambda_1 = \lambda_2 = \cdots = \lambda_m = \lambda$.
Then we can write

$$CV(k,\lambda) = B_1\left(\frac{k}{n}\right)^{\frac{4}{q}} + B_2\left(\frac{1}{k}\right) + \bar{B}_3\lambda^2 + \bar{B}_4\lambda\left(\frac{k}{n}\right)^{\frac{2}{q}} + D_1\left(\frac{k}{n}\right)^{\frac{6}{q}} + D_2\left(\frac{1}{k}\right)\left(\frac{k}{n}\right)^{\frac{2}{q}}$$
$$+ \quad \bar{D}_3\lambda\left(\frac{k}{n}\right)^{\frac{4}{q}} + \bar{D}_4\lambda^2\left(\frac{k}{n}\right)^{\frac{2}{q}} + \bar{D}_5\lambda^3 + (s.o.),$$

where $\bar{B}_3, \bar{B}_4, \bar{D}_3, \bar{D}_4$, and $\bar{D}_5$ do not depend on k or λ.
Then we can derive the following first-order conditions for $\hat{k}$ and $\hat{\lambda}$:

$$\left(\frac{4}{q}\right)B_1\frac{1}{\hat{k}}\left(\frac{\hat{k}}{n}\right)^{\frac{4}{q}} - B_2\frac{1}{\hat{k}^2} + \left(\frac{2}{q}\right)B_4\hat{\lambda}\left(\frac{1}{\hat{k}}\right)\left(\frac{\hat{k}}{n}\right)^{\frac{2}{q}} + \left(\frac{6}{q}\right)D_1\frac{1}{\hat{k}}\left(\frac{\hat{k}}{n}\right)^{\frac{6}{q}}$$
$$+\left(\frac{2-q}{q}\right)D_2\frac{1}{\hat{k}^2}\left(\frac{\hat{k}}{n}\right)^{\frac{2}{q}} + \left(\frac{4}{q}\right)\bar{D}_3\hat{\lambda}\left(\frac{1}{\hat{k}}\right)\left(\frac{\hat{k}}{n}\right)^{\frac{4}{q}} + \left(\frac{2}{q}\right)\bar{D}_4\hat{\lambda}^2\left(\frac{1}{\hat{k}}\right)\left(\frac{\hat{k}}{n}\right)^{\frac{2}{q}} + (s.o.) = 0, \tag{34}$$

$$2\bar{B}_3\hat{\lambda} + \bar{B}_4\left(\frac{\hat{k}}{n}\right)^{\frac{2}{q}} + \bar{D}_3\left(\frac{\hat{k}}{n}\right)^{\frac{4}{q}} + 2\bar{D}_4\hat{\lambda}\left(\frac{\hat{k}}{n}\right)^{\frac{2}{q}} + 3\bar{D}_5\hat{\lambda}^2 + (s.o.) = 0. \tag{35}$$

Let $\hat{\lambda} = \lambda_0 + \hat{\lambda}_1$. Note that $\hat{\lambda}^2 = \left(\lambda_0 + \hat{\lambda}_1\right)^2 = \lambda_0^2 + 2\lambda_0\hat{\lambda}_1 + (s.o.)$ since $\hat{\lambda}_1 = o_p(\lambda_0)$ and $\left(\frac{\hat{k}}{n}\right)^a = \left(\frac{k_0}{n} + \frac{\hat{k}_1}{n}\right)^a = \left(\frac{k_0}{n}\right)^a + a\left(\frac{k_o}{n}\right)^{a-1}\left(\frac{\hat{k}_1}{n}\right) + (s.o.)$ using Taylor's Theorem and $\hat{k}_1 = o_p(k_0)$. Using these two equations, we can rewrite (34) and (35) as

$$\left(\frac{4}{q}\right) B_1 \frac{1}{k_0}\left(\frac{k_0}{n}\right)^{\frac{4}{q}} + \left(\frac{4-q}{q}\right)\left(\frac{4}{q}\right) B_1 \frac{1}{k_0^2}\left(\frac{k_0}{n}\right)^{\frac{4}{q}} \hat{k}_1 - B_2\left(\frac{1}{k_0^2} - \frac{2\hat{k}_1}{k_0^3}\right)$$
$$+ \left(\frac{2}{q}\right) B_4 \left[\lambda_0 + \hat{\lambda}_1\right]\left[\left(\frac{1}{k_0}\right)\left(\frac{k_0}{n}\right)^{\frac{2}{q}} + \left(\frac{2-q}{q}\right)\left(\frac{1}{k_0^2}\right)\left(\frac{k_0}{n}\right)^{\frac{2}{q}} \hat{k}_1\right]$$
$$+ \left(\frac{6}{q}\right) D_1 \left[\frac{1}{k_0}\left(\frac{k_0}{n}\right)^{\frac{6}{q}} + \left(\frac{6-q}{q}\right)\frac{1}{k_0^2}\left(\frac{k_0}{n}\right)^{\frac{6}{q}} \hat{k}_1\right]$$
$$+ \left(\frac{2-q}{q}\right) D_2 \left[\frac{1}{k_0^2}\left(\frac{k_0}{n}\right)^{\frac{2}{q}} + \left(\frac{2-2q}{q}\right)\frac{1}{k_0^3}\left(\frac{k_0}{n}\right)^{\frac{2}{q}} \hat{k}_1\right]$$
$$+ \left(\frac{4}{q}\right) \bar{D}_3(\lambda_0 + \hat{\lambda}_1)\left[\frac{1}{k_0}\left(\frac{k_0}{n}\right)^{\frac{4}{q}} + \left(\frac{4-q}{q}\right)\frac{1}{k_0^2}\left(\frac{k_0}{n}\right)^{\frac{4}{q}} \hat{k}_1\right]$$
$$+ \left(\frac{2}{q}\right) \bar{D}_4\left(\lambda_0^2 + \lambda_0\hat{\lambda}_1\right)\left[\frac{1}{k_0}\left(\frac{k_0}{n}\right)^{\frac{2}{q}} + \left(\frac{2-q}{q}\right)\frac{1}{k_0^2}\left(\frac{k_0}{n}\right)^{\frac{2}{q}} \hat{k}_1\right] + (s.o.) = 0. \tag{36}$$

$$2\bar{B}_3(\lambda_0 + \hat{\lambda}_1) + \bar{B}_4\left[\left(\frac{k_0}{n}\right)^{\frac{2}{q}} + \left(\frac{2-q}{q}\right)\frac{1}{k_0}\left(\frac{k_0}{n}\right)^{\frac{2}{q}} \hat{k}_1\right] +$$
$$\bar{D}_3\left[\left(\frac{k_0}{n}\right)^{\frac{4}{q}} + \left(\frac{4-q}{q}\right)\frac{1}{k_0}\left(\frac{k_0}{n}\right)^{\frac{4}{q}} \hat{k}_1\right] + 2\bar{D}_4(\lambda_0 + \hat{\lambda}_1)\left[\left(\frac{k_0}{n}\right)^{\frac{2}{q}} + \left(\frac{2-q}{q}\right)\frac{1}{k_0}\left(\frac{k_0}{n}\right)^{\frac{2}{q}} \hat{k}_1\right]$$
$$+ 3\bar{D}_3\left[\lambda_0^2 + 2\lambda_0\hat{\lambda}_1\right] + (s.o.) = 0. \tag{37}$$

Using (23) we can solve (37) to obtain $\hat{\lambda}_1 = O_p\left((1/k_0)(k_0/n)^{2/q}\hat{k}_1\right)$. Substituting this into (36), we obtain $\hat{k}_1 = O_p\left(k_0(k_0/n)^{2/q}\right)$, which implies that $\hat{k}_1/k_0 = O_p\left((k_0/n)^{2/q}\right)$. Recalling that $\hat{k}_1 \equiv \hat{k} - k_0$ and $k_0 = O\left(n^{4/(4+q)}\right)$, we have that $(\hat{k} - k_0)/k_0 = O_p\left(n^{-2/(4+q)}\right)$. Then, substituting into (32) , we obtain
$\hat{\lambda} - \lambda_0 = O_p\left(n^{-4/(4+q)}\right)$.

When $q = 4$,

$$CV(k,\lambda) = \quad B_1\left(\frac{k}{n}\right)^{\frac{4}{q}} + B_2\left(\frac{1}{k}\right) + B_3(\lambda) + B_4(\lambda)\left(\frac{k}{n}\right)^{\frac{2}{q}} + (nk)^{-1/2}\mathcal{C}_1 + n^{-\frac{1}{2}}\left(\frac{k}{n}\right)^{\frac{2}{q}}\mathcal{C}_2$$
$$+ n^{-\frac{1}{2}}C_3(\lambda) + \ D_1\left(\frac{k}{n}\right)^{\frac{6}{q}} + D_2\left(\frac{1}{k}\right)\left(\frac{k}{n}\right)^{\frac{2}{q}} + \quad D_3(\lambda)\left(\frac{k}{n}\right)^{\frac{4}{q}} + D_4(\lambda)\left(\frac{k}{n}\right)^{\frac{2}{q}}$$
$$+ D_5(\lambda) + (s.o.)$$

We can see that having added some more terms of the same order, our results from the case $q \geq 5$ still hold. Hence, we have for $q \geq 4$,

$$(\hat{k} - k_0)/k_0 = O_p\left(n^{-2/(4+q)}\right) \qquad \text{and} \qquad \hat{\lambda} - \lambda_0 = O_p\left(n^{-4/(4+q)}\right).$$

Proof of Theorem 4

Let $\hat{g}_0(x)$ and $\hat{f}_0(x)$ denote $\hat{g}(x)$ and $\hat{f}(x)$, respectively, evaluated at $k = k_0$ and $\lambda = \lambda_0$. We can write $\hat{g}_0(x) - g(x) = (\hat{g}_0(x) - g(x))\hat{f}_0(x)/\hat{f}_0(x) \equiv \hat{m}_0(x)/\hat{f}_0(x)$, where $\hat{m}_0(x) = (\hat{g}_0(x) - g(x))\hat{f}_0(x)$

We can show that $E[\hat{m}_0(x)] = f(x)\left[\mu_k(x)\,(k_0/n)^{2/q} + \lambda'\mu_l(x)\right] + o\left((k_0/n)^{2/q} + \|\lambda\|\right)$, where

$$\mu_k(x) = \quad c_w(c_0 f(x))^{2/q}\left[f(x)tr[\nabla^2 g(x)]/2 + \nabla f(x)'\nabla g(x)\right],$$

$\mu_l(x)$ is an $m \times 1$ vector whose s-th element is given by

$$\mu_{ls}(x) = \quad \sum_{z_d \in D} \mathbf{1}_s\left(z^d, x^d\right)\left[g\left(x^c, z^d\right) - g(x)\right] f\left(x^c, z^d\right).$$

We can also show that $Var[\hat{m}_0(x)] = c_0 d_w \sigma^2(x) f^2(x)\frac{1}{k_0} + o\left(\frac{1}{k_0}\right)$. We can verify that the conditions for Liapunov's central limit theorem hold. Then we have that

$$
\begin{aligned}
& k_0^{1/2}\left[\hat{g}_0(x)-g(x)-\mu_k(x)\left(k_0/n\right)^{2/q}-\lambda'\mu_l(x)\right] \\
= \; & k_0^{1/2}\left[\hat{m}_0(x)-\left(\mu_k(x)\left(k_0/n\right)^{2/q}+\lambda'\mu_l(x)\right)\hat{f}_0(x)\right]/\hat{f}_0(x) \\
= \; & k_0^{1/2}\left[\hat{m}_0(x)-\left(\mu_k(x)\left(k_0/n\right)^{2/q}+\lambda'\mu_l(x)\right)f(x)\right]/f(x)+o_p(1) \\
\xrightarrow{d} \; & N\left(0,c_0 d_w \sigma^2(x)\right). \qquad (38)
\end{aligned}
$$

We now let $\hat{g}_{\hat{\gamma}}(x)$ and $\hat{f}_{\hat{\gamma}}(x)$ denote, respectively, $\hat{f}(x)$ and $\hat{g}(x)$ evaluated at $k=\hat{k}$ and $\lambda=\hat{\lambda}$. We can write $\hat{g}_{\wedge}(x)-g(x)=\left[\hat{g}_{\hat{\gamma}}(x)-g(x)\right]\hat{f}_{\hat{\gamma}}(x)/\hat{f}_{\hat{\gamma}}(x)\equiv\hat{m}_{\hat{\gamma}}(x)/\hat{f}_{\hat{\gamma}}(x)$, where $\hat{m}_{\hat{\gamma}}(x)=\left[\hat{g}_{\hat{\gamma}}(x)-g(x)\right]\hat{f}_{\hat{\gamma}}(x)$. Using the same arguments as in Racine and Li (2004), we can show that

$$
\begin{aligned}
& \hat{k}^{1/2}\left[\hat{g}_{\hat{\gamma}}(x)-g(x)-\left(\mu_k(x)\left(k_0/n\right)^{2/q}+\lambda'\mu_l(x)\right)\right] \\
= \; & k_0^{1/2}\left[\hat{g}_0(x)-g(x)-\left(\mu_k(x)\left(k_0/n\right)^{2/q}-\hat{\lambda}'\mu_l(x)\right)\right]+o_p(1).
\end{aligned}
$$

Thus by (38), we have that

$$
\hat{k}^{1/2}\left[\hat{g}_{\hat{\gamma}}(x)-g(x)-\left(\mu_k(x)\left(\hat{k}/n\right)^{2/q}+\hat{\lambda}'\mu_l(x)\right)\right]\xrightarrow{d} N\left(0,c_0 d_w\sigma^2(x)\right).
$$

APPENDIX B

APPENDIX TO SECTION 3

B.1 Proof of Theorem 1

For simplicity, we assume that $h_1 = h_2 = \ldots = h_d = h$. We assume further that $k(0) = 0$, where $k(\cdot)$ is the kernel function. From Theorem 2.2 on page 117 of Su and Lu (2013), we have the following result:

Result 1

$$\hat{g}(x) - g(x) = h^2 b(x) + \frac{1}{NT_2 h^d} \sum_{j=1}^{N} \sum_{s=3}^{T} \epsilon_{js} c(x) K\left(\frac{x_{js} - x}{h}\right) + R_n(x), \qquad (1)$$

where the expressions for $b(x)$ and $c(x)$ can be derived from equations 2.16 and 2.17 of Su and Lu (2013). $R_n(x)$ is the remainder term which is defined by (1), i.e.,

$$R_n(x) = \hat{g}(x) - g(x) - h^2 b(x) - \frac{1}{NT_2 h^d} \sum_{j=1}^{N} \sum_{s=3}^{T} \epsilon_{js} c(x) K\left(\frac{x_{js} - x}{h}\right).$$

It is easy to see that $R_n(x)$ has a probability order uniformly (in x) smaller than

$$h^2 b(x) + \frac{1}{NT_2 h^d} \sum_{j=1}^{N} \sum_{s=3}^{T} \epsilon_{js} c(x) K\left(\frac{x_{js} - x}{h}\right).$$

Note that we do not give explicit expressions for $b(x)$ and $c(x)$ in (1) for two reasons: (i) The explicit definitions of $b(x)$ and $c(x)$ would require one define many quantities related a general local polynomial estimator, and operator related to a recursive estimation procedure. These will take too much spaces. (ii) The explicit expressions for $b(x)$ and $c(x)$ do not enter the leading term of our test statistics. We only need the term related to $b(x)$ is of the order $O(h^2)$ and that the term associated with $c(x)$ is of the order $O((Nh^d)^{-2})$.

Below we give the expressions for $b(x)$ and $c(x)$ for the case of a local linear estimation method. Readers interested in the detailed expressions for $b(x)$ and $c(x)$ for the general local polynomial are referred to Su and Lu (2013). For the local linear estimation method case, $b(x) = (1-\mathcal{A})^{-1}B_0(x)$, $B_0(x) = (\mu_2/2)h^2 \sum_{j=1}^{d} \partial^2 g(x)/\partial x_j^2$, $\mu_2 = \int k(v)v^2 dv$, $\mathcal{A}$ is a linear operator defined in (2.8) in Su and Lu (2013). And that $c(x) = c/f(x)$ (c is a constant).

Note that

$$
\begin{aligned}
\hat{\epsilon}_{it} &= y_{it} - y_{i,t-1} - (\hat{g}_{it} - \hat{g}_{i,t-1}) \\
&= (g_{it} + \nu_{it}) - (g_{i,t-1} + \nu_{i,t-1}) - (\hat{g}_{it} - \hat{g}_{i,t-1}) \\
&= (\nu_{it} - \nu_{i,t-1}) - [(\hat{g}_{it} - g_{it}) - (\hat{g}_{i,t-1} - g_{i,t-1})] \\
&= \epsilon_{it} - \hat{\eta}_{it}, \qquad (2)
\end{aligned}
$$

where

$$
\hat{\eta}_{it} = (\hat{g}_{it} - g_{it}) - (\hat{g}_{i,t-1} - g_{i,t-1}). \qquad (3)
$$

Thus

$$
\begin{aligned}
\hat{\epsilon}_{it}\hat{\epsilon}_{it-2} &= (\epsilon_{it} - \hat{\eta}_{it})(\epsilon_{it-2} - \hat{\eta}_{i,t-2}) \\
&= \epsilon_{it}\epsilon_{,it-2} - \epsilon_{it}\hat{\eta}_{i,t-2} - \hat{\eta}_{it}\epsilon_{i,t-2} + \hat{\eta}_{it}\hat{\eta}_{i,t-2}, \qquad (4)
\end{aligned}
$$

Using equation (4) , we can write a new expression for I_N:

$$\begin{aligned} I_N &= \frac{1}{NT_3}\sum_{i=1}^{N}\sum_{t=4}^{T}\hat{\epsilon}_{it}\hat{\epsilon}_{i,t-2} \\ &= \frac{1}{NT_3}\sum_{i=1}^{N}\sum_{t=4}^{T}\epsilon_{it}\epsilon_{i,t-2} - \frac{1}{NT_3}\sum_{i=1}^{N}\sum_{t=4}^{T}\epsilon_{it}\hat{\eta}_{i,t-2} \\ &\quad - \frac{1}{NT_3}\sum_{i=1}^{N}\sum_{t=4}^{T}\hat{\eta}_{it}\epsilon_{i,t-2} + \frac{1}{NT_3}\sum_{i=1}^{N}\sum_{t=4}^{T}\hat{\eta}_{it}\hat{\eta}_{i,t-2} \\ &\equiv \quad A_{1N} + A_{2N} + A_{3N} + A_{4N} \ , \end{aligned} \tag{5}$$

where the definitions of A_{1N}, A_{2N}, A_{3N}, and A_{4N} should be apparent.

By the Lindeberg central limit theorem, we have that

$$\sqrt{NT_3}A_{1N} \xrightarrow{d} N\left(0, \sigma_\epsilon^4\right). \tag{6}$$

We show in Appendix B that

$$A_{2N} \quad = \quad O_p\left(N^{-1/2}h^2 + N^{-1}h^{-d/2}\right), \tag{7}$$

$$A_{3N} \quad = \quad O_p\left(N^{-1/2}h^2 + N^{-1}h^{-d/2}\right), \tag{8}$$

$$A_{4N} \quad = \quad O_p\left(h^4 + N^{-1}\right). \tag{9}$$

By assumption A1 (xiii) we know that $A_{jN} = o_p(N^{-1/2})$. Combining (6) with (9), we have shown that

$$\frac{\sqrt{NT_3}I_N}{\sigma_\epsilon^2} \xrightarrow{d} N(0,1)$$

under H_0.

Similarly, one can show that $\hat{\sigma}_\epsilon^2 = \frac{1}{NT_3}\sum_{i=1}^{N}\sum_{t=4}^{T}\epsilon_{it}^2 + o_p(1) \xrightarrow{p} \sigma_\epsilon^2$. Hence, Theorem 1 follows.

B.2 Appendix B: Proofs of (7) to (9)

Lemma 1 *Under conditions given in the statement of Theorem 1, we have*

$$A_{2N} = O_p\left(N^{-1/2}h^2 + N^{-1}h^{-d/2}\right).$$

Note that $\epsilon_{it}\hat{\eta}_{i,t-2} = \epsilon_{it}\left[(\hat{g}_{i,t-2} - g_{i,t-2}) - (\hat{g}_{i,t-3} - g_{i,t-3})\right]$. Thus, we can write A_{2N} as

$$\begin{aligned}
A_{2N} &= \frac{1}{NT_3}\sum_{i=1}^{N}\sum_{t=4}^{T}\epsilon_{it}\hat{\eta}_{i,t-2} \\
&= \frac{1}{NT_3}\sum_{i=1}^{N}\sum_{t=4}^{T}\epsilon_{it}(\hat{g}_{i,t-2} - g_{i,t-2}) - \frac{1}{NT_3}\sum_{i=1}^{N}\sum_{t=4}^{T}\epsilon_{it}(\hat{g}_{i,t-3} - g_{i,t-3}) \\
&\equiv A_{2N,1} + A_{2N,2}\ ,
\end{aligned}$$

where the definitions of $A_{2N,1}$ and $A_{2N,2}$ should be apparent.

We first consider $A_{2N,1}$. For convenience, we will use b_{it} and c_{it} to denote $b(x_{it})$ and $c(x_{it})$, respectively. Using (1) we can write $A_{2N,1}$ as follows:

$$\begin{aligned}
A_{2N,1} = &\ \frac{1}{NT_3}\sum_{i=1}^{N}\sum_{t=4}^{T}h^2\epsilon_{it}b_{i,t-2} + \frac{1}{N^2T_1T_3h^d}\sum_{i=1}^{N}\sum_{t=4}^{T}\sum_{j=1}^{N}\sum_{s=2}^{T}\epsilon_{it}\epsilon_{js}c_{i,t-2}K_{js,(i,t-2)} + (s.o.) \\
\equiv &\ A_{2N,11} + A_{2N,12} + (s.o.),
\end{aligned}$$

where $K_{js,it}$ denotes $K\left((x_{js} - x_{it})/h\right)$. The notation $A_N = B_N + (s.o.)$ means that B_N is the leading term of A_N, and $(s.o.)$ denote terms having probability orders smaller than that of B_N.

Below we first consider $A_{2N,11}$. Using the fact that the data is cross-sectionally independent, we have that $E\left[A_{2N,11}^2\right] = N^{-2}h^4 O(N) = O\left(N^{-1}h^4\right)$. It follows that

$$A_{2N,11} = O_p\left(N^{-1/2}h^2\right).$$

We next consider $A_{2N,12}$. Note that

$$\begin{aligned} A_{2N,12} \equiv & \frac{1}{N^2 T_1 T_3 h^d} \sum_{i=1}^{N} \sum_{t=4}^{T} \sum_{s=2}^{T} \epsilon_{it}\epsilon_{is} c_{i,t-2} K_{is,(i,t-2)} \\ & + \frac{1}{N^2 T_1 T_3 h^d} \sum_{i=1}^{N} \sum_{j\neq i}^{N} \sum_{t=4}^{T} \sum_{s=2}^{T} \epsilon_{it}\epsilon_{js} c_{i,t-2} K_{js,(i,t-2)} \\ & \equiv B_{1N} + B_{2N} , \end{aligned} \tag{10}$$

where the definitions of B_{1N} and B_{2N} should be apparent. We first consider B_{1N}. Note that under H_0 we can write B_{1N} as

$$\begin{aligned} B_{1N} &= \frac{2}{N^2 T_1 T_3 h^d} \sum_{i=1}^{N} \sum_{t=4}^{T} \epsilon_{it}^2 c_{i,t-2} K_{(i,t),(i,t-2)} \\ & + \frac{2}{N^2 T_1 T_3 h^d} \sum_{i=1}^{N} \sum_{t=4}^{T} \sum_{s=2,s\neq t}^{T} \epsilon_{it}\epsilon_{is} c_{i,t-2} K_{(is),(i,t-2)} \\ &\equiv B_{1N,1} + B_{1N,2}, \end{aligned} \tag{11}$$

where the definition of $B_{1N,1}$ and $B_{1N,2}$ should be apparent.

Note that we can write $B_{1N,1}$ as $B_{1N,1} = \frac{2}{N^2} \sum_{i=1}^{N} \vartheta_i$, where

$\vartheta_i = \frac{1}{T_1 T_3 h^d} \sum_{t=4}^{T} \epsilon_{it}\epsilon_{i,t-1} c_{i,t-2} K_{(i,t-1),(i,t-2)}$. The ϑ_i's are i.i.d, and we can show that ϑ_i has finite a mean and variance of order h^{-d}. It follows that $\sum_{i=1}^{N} \vartheta_i$ has mean of order N and variance of order Nh^{-d}. Thus we have that $\sum_{i=1}^{N} \vartheta_i = O_p\left(N + N^{1/2}h^{-d/2}\right) =$

$O_p(N)$ and $B_{1N,1} = O_p(N^{-1})$. Similarly, one can show that $B_{1N,2} = O_p(N^{-1})$. Combining these results, we have that

$$B_{1N} = O_p\left(N^{-1}\right).$$

We next consider B_{2N}. Note that we can write B_{2N} as a degenerate second-order U-statistic:

$$\begin{aligned} B_{2N} \equiv & \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t=4}^{T}\sum_{s=2}^{T}\frac{1}{T_1T_3h^d}\epsilon_{it}\epsilon_{js}c_{i,t-2}K_{js,(i,t-2)} \\ = & \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}H_{N,ij} \\ = & \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{1}{2}(H_{N,ij}+H_{N,ji}) \\ = & \frac{2}{N^2}\sum_{i=1}^{N}\sum_{j>i}^{N}\bar{H}_{N,ij}\ , \end{aligned}$$

where the definition of H_{ij} should be apparent and $\bar{H}_{N,ij}(H_{N,ij}+H_{N,ji})/2$ is a symmetrized version of $H_{N,ij}$. Then we have that

$$E\left[B_{2N}^2\right] = \frac{2}{N^2}\,E\left[\bar{H}_{N,ij}^2\right].$$

We can show using a Taylor expansion and a change of variables that $E\left[\bar{H}_{N,ij}^2\right] = O\left(h^{-d}\right)$. Thus we have that $E\left[B_{2N}^2\right] = O(N^{-2}h^{-d})$, which implies that $B_{2N} = O_p\left(N^{-1}h^{-d/2}\right)$.

Then we have that $A_{2N,12} = O_p\left(N^{-1}h^{-d/2} + N^{-3/2}h^{-d} + N^{-1}\right) = O_p\left(N^{-1}h^{-d/2}\right)$, and thus $A_{2N,1} = O_p\left(N^{-1/2}h^2 + N^{-1}h^{-d/2}\right)$. It is clear that we can use similar arguments to show that $A_{2,2N}$ is of the same probability order. Thus, we have that

$$A_{2N} = O_p\left(N^{-1/2}h^2 + N^{-1}h^{-d/2}\right).$$

Lemma 4 *Under conditions given in the statement of Theorem 1, we have*

$$A_{3N} = O_p\left(N^{-1/2}h^2 + N^{-1}h^{-d/2}\right).$$

By using the same method as in the proof of lemma 1, one can easily prove that A_{3N} has the same probability order as that of A_{2N}. We therefore omit the proof of lemma 4 here.

Finally, we derive the probability order of A_{4N} in the next lemma.

Lemma 5 *Under conditions given Theorem 1, we have*

$$A_{4N} = O_p\left(h^4 + N^{-3/2}h^{-2d} + N^{-2}h^{-d} + N^{-1}\right).$$

Using (3) we have that

$$
\begin{aligned}
A_{4N} \equiv & \quad \frac{1}{NT_3}\sum_{i=1}^{N}\sum_{t=4}^{T}\hat{\eta}_{it}\hat{\eta}_{i,t-2} \\
= & \quad \frac{1}{NT_3}\sum_{i=1}^{N}\sum_{t=4}^{T}\left[(\hat{g}_{it}-g_{it})-(\hat{g}_{i,t-1}-g_{i,t-1})\right]\left[(\hat{g}_{i,t-2}-g_{i,t-2})-(\hat{g}_{i,t-3}-g_{i,t-3})\right] \\
\equiv & \quad \frac{1}{NT_3}\sum_{i=1}^{N}\sum_{t=4}^{T}(\hat{g}_{it}-g_{it})(\hat{g}_{i,t-2}-g_{i,t-2}) - \frac{1}{NT_3}\sum_{i=1}^{N}\sum_{t=4}^{T}(\hat{g}_{it}-g_{it})(\hat{g}_{i,t-3}-g_{i,t-3}) \\
& - \quad \frac{1}{NT_3}\sum_{i=1}^{N}\sum_{t=4}^{T}(\hat{g}_{i,t-1}-g_{i,t-1})(\hat{g}_{i,t-2}-g_{i,t-2}) \\
& + \frac{1}{NT_3}\sum_{i=1}^{N}\sum_{t=4}^{T}(\hat{g}_{i,t-1}-g_{i,t-1})(\hat{g}_{i,t-3}-g_{i,t-3}) \\
\equiv & \quad A_{4N,1}-A_{4N,2}-A_{4N,3}+A_{4N,4}.
\end{aligned} \tag{12}
$$

We first consider $A_{4N,3}$. Since $\hat{g}(\cdot)$ satisfies the assumption in (1). Define $N_j = N-j$, then we can write

$$
\begin{aligned}
A_{4N,3} = & \ \frac{1}{NT_3}\sum_{i=1}^{N}\sum_{t=4}^{T}\left(h^2 b_{i,t-1}+\frac{1}{N_1T_1h^d}\sum_{j=1}^{N}\zeta_{(i,t-1),j}\right)\left(h^2 b_{i,t-2}+\frac{1}{N_1T_1h^d}\sum_{k=1}^{N}\zeta_{(i,t-2),k}\right) \\
= & \quad \frac{1}{NT_3}\sum_{i=1}^{N}\sum_{t=4}^{T}h^4 b_{i,t-1}b_{i,t-2}+\frac{1}{NN_1T_1T_3h^{d-2}}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=4}^{T}b_{i,t-1}\zeta_{(i,t-2),k} \\
+ & \quad \frac{1}{NN_1T_1T_3h^{d-2}}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=4}^{T}b_{i,t-2}\zeta_{(i,t-1),j} \\
+ & \quad \frac{1}{NN_1^2T_3T_1^2h^{2d}}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{t=4}^{T}\zeta_{(i,t-1),j}\zeta_{(i,t-2),k} \\
\equiv & \quad A_{4N,31}+A_{4N,32}+A_{4N,33}+A_{4N,34}\ ,
\end{aligned}
$$

where

$$
\zeta_{it,j} \equiv \sum_{s=2}^{T}\epsilon_{js}c_{it}K_{js,it}
$$

and the definitions of $A_{4N,31}, A_{4N,32}, A_{4N,33}$, and $A_{4N,34}$ should be apparent.

Note that

$$E[A_{4N,31}^2] = N^{-2}h^8 O(N^2) = O\left(h^8\right).$$

It follows that $A_{4N,31} = O_p\left(h^4\right)$.

We next consider $A_{4N,32}$. Note that

$$\begin{aligned} A_{4N,32} =& \frac{1}{NN_1T_1T_3h^{d-2}} \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=4}^{T}\sum_{s=2}^{T} b_{i,t-1}\epsilon_{js}c_{i,t-2}K_{js,(i,t-2)} \\ =& \frac{1}{NN_1T_1T_3h^{d-2}} \sum_{i=1}^{N}\sum_{t=4}^{T}\sum_{s=2}^{T} b_{i,t-1}\epsilon_{is}c_{i,t-2}K_{is,(i,t-2)} \\ +& \frac{1}{NN_1T_1T_3h^{d-2}} \sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t=4}^{T}\sum_{s=2}^{T} b_{i,t-1}\epsilon_{js}c_{i,t-2}K_{js,(i,t-2)} \\ \equiv& B_{3N} + B_{4N}\ , \end{aligned}$$

where the definitions of B_{3N} and B_{4N} should be apparent.

We first consider B_{3N}. It is straightforward to show that $E[B_{1N}^2] = \frac{1}{N^4h^{2d-4}}O(Nh^d) = O\left(\frac{1}{N^3h^{d-4}}\right)$. It follows that $B_{3N} = O_p\left(\frac{1}{N^{3/2}h^{d/2-2}}\right) = o_p(N^{-1})$.

We next consider B_{4N}. Note that we can write B_{4N} as a second-order U-statistic:

$$\begin{aligned} B_{4N} =& \frac{1}{NN_1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t=4}^{T}\sum_{s=2}^{T}\frac{1}{T_1T_3h^{d-2}} b_{i,t-1}\epsilon_{js}c_{i,t-2}K_{js,(i,t-2)} \\ =& \frac{1}{NN_1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{1}{2}(H_{ij,N}+H_{ji,N}) \\ =& \frac{2}{NN_1}\sum_{i=1}^{N}\sum_{j>i}^{N}\bar{H}_{ij,N}\ , \end{aligned}$$

where the definition of $H_{ij,N}$ should be apparent and $\bar{H}_{ij,N}$ is symmetrized version of $H_{ij,N}$.

Using the H-decomposition, we have that

$$\begin{aligned} B_{4N} = & \frac{2}{N}\sum_{i=1}^{N}\bar{H}_{i,N} + \frac{2}{NN_1}\sum_{i=1}^{N}\sum_{j>i}^{N}(\bar{H}_{ij,N} - \bar{H}_{i,N} - \bar{H}_{j,N}) \\ \equiv & B_{4N,1} + B_{4N,2} \ , \end{aligned} \tag{13}$$

where $\bar{H}_{i,N} \equiv E\left[\bar{H}_{ij,N}|z_i\right]$ with $z_i \equiv (x'_{i1}, ..., x'_{iT}, \epsilon_{i1}, ...\epsilon_{iT})'$. We can show that $B_{4N,1}$ has a probability order larger than that of $B_{4N,2}$.

Note that

$$B_{4N,1} = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s=4}^{T}\frac{1}{T_1T_3h^{d-2}}\epsilon_{it}E\left[b_{j,s-1}c_{j,s-2}K_{it,(j,s-2)}|x_{it}\right] \tag{14}$$

We can show using a Taylor expansion and a change of variables that

$$E\left[b_{j,s-1}c_{j,s-2}K_{it,(j,s-2)}|x_{it}\right] = h^d f(x_{it})E[b(x_{j,s-1})]c(x_{it}) + O_p\left(h^{d+2}\right) \ , \tag{15}$$

where, after multiplying by h^{-d-2}, the $O_p\left(h^{d+2}\right)$ terms are bounded in probability uniformly in $x_{it} \in \mathcal{D}$.

Using (14) and (15), we have that

$$\begin{aligned} B_{4N,1} = & \frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s=4}^{T}\frac{1}{T_1T_3h^{d-2}}\epsilon_{it}E\left[b_{js}c_{j,s-2}K_{it,(j,s-2)}|x_{it}\right] \\ = & \frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\frac{1}{T_1}\epsilon_{it}h^2 f(x_{it})E[b(x_{j,s-1})]c(x_{it}) + (s.o.) \\ \equiv & B_{4N,11} + (s.o.) \ , \end{aligned} \tag{16}$$

where the definition of $B_{4N,11}$ should be apparent. Note that $E[B_{2N,11}^2] = N^{-2}h^4O(N) = O\left(N^{-1}h^4\right)$. It follows that $B_{4N,11} = O_p\left(N^{-1/2}h^2\right)$. Thus, from (16) we have that $B_{4N,1} = O_p\left(N^{-1/2}h^2 + h^4\right)$. Thus, since $B_{4N,2}$ is of smaller order, we have that $B_{4N} = O_p\left(N^{-1/2}h^2 + h^4\right)$.

It follows that $A_{4N,32} = O_p\left(N^{-1} + N^{-1/2}h^2 + h^4\right) = O_p\left(N^{-1} + h^4\right)$. It is clear that we can show in a similar way that $A_{4N,33}$ is on the same order order as $A_{4N,32}$.

We next consider $A_{4N,34}$. We have that

$$\begin{aligned}
A_{4N,34} &= \frac{1}{NN_1^2T_3T_1^2h^{2d}}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{t=4}^{T}\sum_{s=2}^{T}\sum_{q=2}^{T}\epsilon_{js}\epsilon_{kq}c_{i,t-1}c_{i,t-2}K_{(i,t-1),js}K_{(i,t-2),kq} \\
&= \frac{1}{NN_1^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}H_{N,ijk} \\
&= \frac{1}{NN_1^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\bar{H}_{N,ijk} \\
&= \frac{1}{NN_1^2}\sum_{i=1}^{N}\bar{H}_{N,iii} + \frac{2}{NN_1^2}\sum_{i=1}^{N}\sum_{j>i}^{N}\bar{H}_{N,iij} + \frac{6}{NN_1^2}\sum_{i=1}^{N}\sum_{j>i}^{N}\sum_{k>j}^{N}\bar{H}_{N,ijk} \\
&\equiv B_{5N} + B_{6N} + B_{7N}\ ,
\end{aligned}$$

where the definitions of $H_{N,ijk}$, B_{5N},B_{6N}, and B_{7N} should be apparent and $\bar{H}_{N,ijk}$ is a symmetrized version of $H_{N,ijk}$.

We first consider B_{5N}. Note that we can write B_{1N} as

$$B_{5N} = \frac{1}{NN_1^2}\sum_{i=1}^{N}\xi_i + (s.o.)\ , \tag{17}$$

where $\xi_i \equiv (T_3T_1^2)^{-1}\sum_{t=4}^{T}\sum_{s=2}^{T}\sum_{q=2}^{T}\frac{1}{h^{2d}}\epsilon_{is}\epsilon_{iq}c_{i,t-1}c_{i,t-2}K_{(i,t-1),is}K_{(i,t-2),iq}$. The ξ_i's are i.i.d, and we can show that they have mean of order 1 and variance of order h^{-2d}.

It follows that $\sum_{i=1}^{N} \xi_i$ has mean of order N and variance of order Nh^{-2d}. Thus we have that $\sum_{i=1}^{N} \xi_i = O_p\left(N + N^{1/2}h^{-d}\right) = O_p(N)$. Thus, from (17) we have that $B_{1N} = O_p(N^{-2})$.

We next consider B_{6N}. Note that we can write B_{6N} as

$$B_{6N} = \frac{1}{N_1}\left[\frac{2}{NN_1}\sum_{i=1}^{N}\sum_{j>i}^{N}\tilde{H}_{N,ij}\right],$$

where the quantity in the brackets is second-order U-statistic that we denote it as U_{2N}. Using the U-statistic H-decomposition, we have that

$$\begin{aligned} U_{2N} = &\ E[\tilde{H}_{N,ij}] + \frac{2}{N}\sum_{i=1}^{N}\left[\tilde{H}_{N,i} - E[\tilde{H}_{N,ij}]\right] \\ &+ \frac{2}{NN_1}\sum_{i=1}^{N}\sum_{j>i}^{N}\left[\tilde{H}_{N,ij} - \tilde{H}_{N,i} - \tilde{H}_{N,j} + E[\tilde{H}_{N,ij}]\right] \\ \equiv &\ U_{2N,1} + U_{2N,2} + U_{2N,3}, \end{aligned} \tag{18}$$

where the definitions of $U_{2N,1}$, $U_{2N,2}$, and $U_{2N,3}$ should be apparent and $\tilde{H}_{N,i} \equiv E[\tilde{H}_{N,ij}|z_i]$. We can show that $U_{2N,1} = O(1)$ and that $U_{2N,2}$ and $U_{2N,3}$ are of smaller probability order. Thus we have that $B_{6N} = O_p(N^{-1})$.

Finally, we consider B_{7N}. Using the U-statistic H-decomposition, we can show that

$$B_{7N} = \frac{6N_2}{NN_1^2}\sum_{i=1}^{N}\sum_{j>i}^{N}\bar{H}_{N,ij} + (s.o.), \tag{19}$$

where $\bar{H}_{N,ij} = E\left[\bar{H}_{N,ijk}|z_i, z_j\right]$. We can show that

$$
\begin{aligned}
\bar{H}_{N,ij} &= \frac{2}{3}\sum_{t=2}^{T}\sum_{s=2}^{T}\frac{1}{T_1^2 h^{2d}}\epsilon_{it}\epsilon_{js}E\left[c_{k,q-1}c_{k,q-2}K_{it,(k,q-1)}K_{js,(k,q-2)}|x_{it},x_{js}\right] \\
&= \frac{2}{3}\sum_{t=2}^{T}\sum_{s=2}^{T}\frac{1}{T_1^2}\epsilon_{it}\epsilon_{js}f(x_{it})f(x_{js})c_{it}c_{js} + O_p(h^2) \ ,
\end{aligned}
$$

where the convergence is uniform in $(x_{it}, x_{js}) \in \mathcal{D}^2$. From (19) we have that

$$
\begin{aligned}
E\left[B_{7N}^2\right] &= \frac{36N_2^2}{N^2N_1^4}\sum_{i=1}^{N}\sum_{j>i}^{N}E\left[\bar{H}_{N,ij}^2\right] + (s.o.) \\
&= \frac{36N_2^2}{N^2N_1^4}O\left(N^2\right) + (s.o.) \\
&= O\left(N^{-2}\right).
\end{aligned}
$$

It follows that $B_7 = O_p\left(N^{-1}\right)$. Combining the results from above, we have that $A_{4N,34} = O_p\left(N^{-1}\right)$. Thus, we have that, under H_0, $A_{4N,3} = O_p\left(h^4 + N^{-1}\right)$.

We can show in a similar way that $A_{4N,1}$, $A_{4N,2}$, and $A_{4N,4}$ are all of probability order $h^4 + N^{-1}$. It follows that, under H_0,

$$
A_{4N} = O_p\left(h^4 + N^{-1}\right).
$$

B.3 Results of Monte Carlo Simulations

Table B.1: Rejection frequency for DGP 1,2,3 under case (i): i.i.d error (estimated size).

DGP 1		Estimated Sizes				
N	T	1%	5%	10%	20%	50%
50	5	0.0075	0.0435	0.0955	0.1890	0.4915
50	10	0.0085	0.0460	0.0970	0.1895	0.4855
100	5	0.0030	0.0415	0.1010	0.1965	0.4805
100	10	0.0100	0.0550	0.1090	0.2050	0.4915
200	5	0.0095	0.0480	0.0890	0.1845	0.4885
200	10	0.0140	0.0510	0.1050	0.1960	0.4960
DGP 2		Estimated Sizes				
N	T	1%	5%	10%	20%	50%
50	5	0.0065	0.0430	0.0940	0.1895	0.4885
50	10	0.0080	0.0460	0.0960	0.1870	0.4900
100	5	0.0030	0.0430	0.1000	0.1975	0.4775
100	10	0.0100	0.0540	0.1110	0.2035	0.4910
200	5	0.0090	0.0485	0.0890	0.1825	0.4865
200	10	0.0135	0.0505	0.1035	0.1940	0.4930
DGP 3		Estimated Sizes				
N	T	1%	5%	10%	20%	50%
50	5	0.0065	0.0430	0.0980	0.1890	0.4960
50	10	0.0090	0.0475	0.0940	0.1870	0.4885
100	5	0.0030	0.0420	0.1010	0.1980	0.4830
100	10	0.0115	0.0530	0.1095	0.2060	0.4920
200	5	0.0095	0.0480	0.0865	0.1835	0.4895
200	10	0.0135	0.0520	0.1020	0.1965	0.4980

Table B.2: Rejection frequency for DGP 1,2,3 under case (ii): AR(1) (estimated power).

DGP 1		Estimated Powers				
N	T	1%	5%	10%	20%	50%
50	5	0.0125	0.1100	0.2075	0.3635	0.6485
50	10	0.1015	0.3345	0.4855	0.6450	0.8580
100	5	0.0045	0.2195	0.3505	0.5020	0.7700
100	10	0.3300	0.6465	0.7655	0.8615	0.9590
200	5	0.1580	0.4570	0.5805	0.7240	0.8990
200	10	0.7740	0.9280	0.9610	0.9855	0.9965
DGP 2		Estimated Powers				
N	T	1%	5%	10%	20%	50%
50	5	0.0125	0.1110	0.2080	0.3625	0.6460
50	10	0.1020	0.3335	0.4875	0.6475	0.8565
100	5	0.0045	0.2185	0.3510	0.5045	0.7715
100	10	0.3305	0.6515	0.7650	0.8620	0.9595
200	5	0.1585	0.4550	0.5815	0.7220	0.9000
200	10	0.7735	0.9275	0.9610	0.9850	0.9965
DGP 3		Estimated Powers				
N	T	1%	5%	10%	20%	50%
50	5	0.0135	0.1090	0.2045	0.3655	0.6465
50	10	0.0995	0.3330	0.4860	0.6460	0.8555
100	5	0.0040	0.2195	0.3465	0.5015	0.7680
100	10	0.3270	0.6495	0.7635	0.8595	0.9575
200	5	0.1585	0.4575	0.5770	0.7235	0.9005
200	10	0.7750	0.9280	0.9610	0.9855	0.9965

Table B.3: Rejection frequency for DGP 1,2,3 under case (iii): MA(1) (estimated power).

DGP 1		Estimated Powers				
N	T	1%	5%	10%	20%	50%
50	5	0.1685	0.5815	0.7520	0.8720	0.9665
50	10	0.9025	0.9920	0.9965	0.9990	1.0000
100	5	0.1500	0.9270	0.9665	0.9895	0.9995
100	10	0.9985	1.0000	1.0000	1.0000	1.0000
200	5	0.9800	0.9985	1.0000	1.0000	1.0000
200	10	1.0000	1.0000	1.0000	1.0000	1.0000
DGP 2		Estimated Powers				
N	T	1%	5%	10%	20%	50%
50	5	0.1665	0.5810	0.7500	0.8725	0.9660
50	10	0.9010	0.9920	0.9960	0.9990	1.0000
100	5	0.1485	0.9270	0.9665	0.9895	0.9995
100	10	0.9985	1.0000	1.0000	1.0000	1.0000
200	5	0.9810	0.9985	1.0000	1.0000	1.0000
200	10	1.0000	1.0000	1.0000	1.0000	1.0000
DGP 3		Estimated Powers				
N	T	1%	5%	10%	20%	50%
50	5	0.1630	0.5870	0.7495	0.8735	0.9650
50	10	0.8970	0.9910	0.9960	0.9990	1.0000
100	5	0.1465	0.9270	0.9650	0.9880	0.9995
100	10	0.9985	1.0000	1.0000	1.0000	1.0000
200	5	0.9795	0.9995	1.0000	1.0000	1.0000
200	10	1.0000	1.0000	1.0000	1.0000	1.0000

Table B.4: Rejection frequency for DGP 2 under case (i)',(ii)' and (iii)'.

Case (i)'		Estimated Sizes				
N	T	1%	5%	10%	20%	50%
50	10	0.0085	0.0465	0.0955	0.1875	0.4915
100	10	0.0105	0.0585	0.1120	0.2075	0.4920
200	10	0.0140	0.0520	0.1045	0.1955	0.4965
Case (ii)'		Estimated Powers				
N	T	1%	5%	10%	20%	50%
50	10	0.1055	0.3520	0.5005	0.6560	0.8650
100	10	0.3550	0.6665	0.7715	0.8730	0.9610
200	10	0.8030	0.9375	0.9690	0.9875	0.9960
Case (iii)'		Estimated Powers				
N	T	1%	5%	10%	20%	50%
50	10	0.7210	0.9470	0.9745	0.9925	0.9985
100	10	0.9885	0.9995	0.9995	1.0000	1.0000
200	10	1.0000	1.0000	1.0000	1.0000	1.0000

Table B.5: Rejection frequency for DGP 2 under case (i)',(ii)' and (iii)': an asymptotic test.

Case (i)'		Estimated Sizes						
N	T	1%	5%	10%	20%	50%	mean	std
50	10	0.0280	0.0950	0.1570	0.2610	0.5660	-0.0207	1.1615
100	10	0.0335	0.1040	0.1750	0.2825	0.5760	-0.0121	1.1960
200	10	0.0270	0.0980	0.1685	0.2725	0.5615	-0.0311	1.1767
Case (ii)'		Estimated Powers						
N	T	1%	5%	10%	20%	50%	mean	std
50	10	0.2930	0.5160	0.6250	0.7370	0.8915	-1.9752	1.0839
100	10	0.5925	0.7775	0.8465	0.9155	0.9700	-2.8174	1.1230
200	10	0.9025	0.9665	0.9845	0.9930	0.9975	-4.0118	1.1075
Case (iii)'		Estimated Powers						
N	T	1%	5%	10%	20%	50%	mean	std
50	10	0.9225	0.9790	0.9910	0.9960	0.9990	-4.0755	1.0223
100	10	0.9995	0.9995	1.0000	1.0000	1.0000	-5.8311	1.0523
200	10	1.0000	1.0000	1.0000	1.0000	1.0000	-8.2893	1.0478

APPENDIX C

APPENDIX TO SECTION 4: PROOFS OF THEOREMS

Before we prove Theorem 4.3.1 we will prove a Lemma which will be used in the proof.

Lemma 1 *Under the same conditions as in Theorem 4.3.1, we have*

$$\max_{1 \le t \le n} |d_t - \gamma^T \gamma_0 d_{t,0}| = O_p(\eta_n), \tag{1}$$

where $d_{t,0} = \beta'(Z_t^T \gamma_0)$ *and* $\eta_n = b^2 + (log(n)/(nb^q))^{1/2})$.

Proof of Lemma 1

It is well established that on a bounded trimmed set (with boundary regions trimmed out) that nonparametric kernel estimator converges to the true unknown function uniformly with a rate $O_p(\eta_n)$, where $\eta_n = b^2 + (log(n)/(nb^q))^{1/2})$, i.e., $\max_{1 \le t \le n} |\tilde{g}(Z_t) - g(Z_t)| = O_p(\eta_n)$. Then it is easy to see that ($g_t = g(Z_t) = \beta(Z_t^T \gamma_0)$)

$$\begin{aligned} d_t &= \left[\sum_{s=1}^{n} (Z_{st}^T \gamma)^2 H_{b,st} \right]^{-1} \sum_{s=1}^{n} Z_{st}^T \gamma (g_s - g_t) H_{b,st} + O_p(\eta_n) \\ &= \left[\sum_{s=1}^{n} (Z_{st}^T \gamma)^2 H_{b,st} \right]^{-1} \sum_{s=1}^{n} Z_{st}^T \gamma Z_{st}^T H_{b,st} \gamma_0 d_{t,0} + O_p(\eta_n) \\ &= \gamma^T \gamma_0 \, d_{t,0} + O_p(\eta_n) \end{aligned} \tag{2}$$

uniformly in $1 \le t \le n$, where $\gamma^T = \left[\sum_{s=1}^{n} (Z_{st}^T \gamma)^2 H_{b,st}\right]^{-1} \sum_{t=1}^{n} Z_{st}^T \gamma Z_{st}^T H_{b,st}$ (it is easy to check that $\gamma^T \gamma = 1$), and the second equality follows from Taylor expansion $g_s = g_t + \beta'(Z_t' \gamma_0) Z_{s,t}^T \gamma_0 + O_p\left((Z_{s,t}^T \gamma_0)^2\right)$. This completes the proof of Lemma 1.

Proof of Theorem 4.3.1

$$
\begin{aligned}
\bar{\gamma} &= \left[\sum_{s=1}^{n}\sum_{t=1}^{n} d_t^2 Z_{st} Z_{st}^T H_{b,st}\right]^{-1} \sum_{s=1}^{n}\sum_{t=1}^{n} d_t Z_{st}(g_s - g_t) H_{b,st} + O_p(\eta_n) \\
&= \left[\sum_{s=1}^{n}\sum_{t=1}^{n} d_t^2 Z_{st} Z_{st}^T H_{b,st}\right]^{-1} \sum_{s=1}^{n}\sum_{t=1}^{n} d_t Z_{st} Z_{st}^T H_{b,st} \gamma_0 d_{t,0} + O_p(\eta_n) \\
&= \gamma_0 + \left[\sum_{s=1}^{n}\sum_{t=1}^{n} d_t^2 Z_{st} Z_{st}^T H_{b,st}\right]^{-1} \sum_{s=1}^{n}\sum_{t=1}^{n} d_t Z_{st} Z_{st}^T H_{b,st} [d_{t,0} - d_t]\gamma_0 + O_p(\eta_n) \\
&= \gamma_0 + (\gamma^T\gamma_0)^{-2} \left[\sum_{s=1}^{n}\sum_{t=1}^{n} d_{t,0}^2 Z_{st} Z_{st}^T H_{b,st}\right]^{-1} (\gamma^T\gamma_0) \sum_{s=1}^{n}\sum_{t=1}^{n} d_{t,0}^2 Z_{st} Z_{st}^T H_{b,st} [1 - \gamma^T\gamma_0]\gamma_0 \\
&\quad + O_p(\eta_n) \\
&= \gamma_0 + (\gamma^T\gamma_0)^{-1}(1 - \gamma^T\gamma_0)\gamma_0 + O_p(\eta_n),
\end{aligned} \tag{3}
$$

where the second equality follows from the Taylor expansion $g_s \approx g_t + d_{t,0} Z_{s,t}^T \gamma_0$, in the third equality we used $d_{t,0} = d_t + (d_{t,0} - d_t)$, the fourth equality follows from (2). Equation (3) can be re-written as $\bar{\gamma} = (\gamma^T\gamma_0)^{-1}\gamma_0 + O_p(\eta_n)$, i.e., $\bar{\gamma}$ equals a constant (scalar) time γ_0 plus a $o_p(1)$ term. By the normalization requirement that $\bar{\gamma}^T\bar{\gamma} = 1$ and $\gamma_0^T\gamma_0 = 1$, we obtain

$$\bar{\gamma} = \gamma_0 + O_p(\eta_n).$$

This completes the proof of Theorem 4.3.1. Note that $\bar{\gamma} = \gamma_0 + o_p(1)$ implies that

$d_t = d_{t,0} + o_p(1)$, as we show below. Using $\max_{1\le t\le n}|\tilde{g}(Z_t) - g(Z_t)| = O_p(\eta_n)$, where

$\eta_n = b^2 + (log(n)/(nb^q))^{1/2})$, it is easy to see that $(g_t = g(Z_t) = \beta(Z_t^T\gamma_0))$

$$
\begin{aligned}
d_t &= \left[\sum_{s=1}^{n}(Z_{st}^T\gamma)^2 H_{b,st}\right]^{-1}\sum_{s=1}^{n} Z_{st}^T\gamma(g_s - g_t)H_{b,st} + O_p(\eta_n) \\
&= \left[\sum_{s=1}^{n}(Z_{st}^T\gamma)^2 H_{b,st}\right]^{-1}\sum_{t=1}^{n} Z_{st}^T\gamma d_{t,0} Z_{st}^T\gamma_0 H_{b,st} + O_p(\eta_n) \\
&= d_{t,0} + \left[\sum_{s=1}^{n}(Z_{st}^T\gamma)^2 H_{b,st}\right]^{-1}\sum_{s=1}^{n} Z_{st}^T\gamma Z_{st}^T(\gamma_0-\gamma) d_{t,0} H_{b,st} + O_p(\eta_n) \\
&= d_{t,0} + E[(Z_{st}^T\gamma)^2|Z_t]^{-1}E[d_{t,0}Z_{st}^T\gamma Z_{st}^T|Z_t](\gamma_0-\gamma) + O_p(\eta_n), \qquad (4)
\end{aligned}
$$

where the second equality follows from theTaylor expansion $g_s = g_t + \beta'(Z_t'\gamma_0)Z_{s,t}^T\gamma_0 + O_p\left((Z_{s,t}^T\gamma_0)^2\right)$; the third equality follows from $\gamma_0 = \gamma + (\gamma_0 - \gamma)$; and the last equality follows from the standard kernel estimation result. Equation (4) implies that $d_t = \beta'(Z_t^T\gamma_0) + o_p(1)$ because $\bar{\gamma} - \gamma_0 = o_p(1)$ by Theorem 4.3.1.

Proof of Theorem 4.3.2

To prove Theorem 4.3.2, we need to modify $A_n = \sum_{s=1}^n\sum_{t=1}^n d_t^2 Z_{st}Z_{st}^T H_{b,st}$ to $A_n^\gamma = \sum_{s=1}^n\sum_{t=1}^n d_t^2 Z_{st}Z_{st}^T K_{h,st}^\gamma$.

$$
\begin{aligned}
A_n^\gamma &\overset{def}{=} \frac{1}{n^2}\sum_{s=1}^{n}\sum_{t=1}^{n} d_t^2 Z_{st}Z_{st}^T K_{h,st}^\gamma \\
&= n^{-1}\sum_{t=1}^{n} d_{t,0}^2 E\{Z_{st}Z_{st}^T K_{h,st}^\gamma|Z_t\} + O_p(\eta_n) \\
&= n^{-1}\sum_{t=1}^{n} d_{t,0}^2 E\{E[Z_{st}Z_{st}^T K_{h,st}^\gamma|Z_s^T\gamma_0, Z_t]|Z_t\} + O_p(\eta_n) \\
&= n^{-1}\sum_{t=1}^{n} d_{t,0}^2 f_\gamma(Z_t^T\gamma_0)(Z_t - \xi(Z_t^T\gamma_0))(Z_t - \xi(Z_t^T\gamma_0))^T + O_p(\eta_n) \\
&= E\left[d_{t,0}^2 f_\gamma(Z_t^T\gamma_0)(Z_t - \xi(Z_t^T\gamma_0))(Z_t - \xi(Z_t^T\gamma_0))^T\right] + O_p(\eta_n), \qquad (5)
\end{aligned}
$$

where $\xi(Z_t^T\gamma_0) = E(Z_t|Z_t^T\gamma_0)$ and $d_{t,0} = \beta'(Z_t^T\gamma_0)$.

$C_n = n^{-2}\sum_{s=1}^n\sum_{t=1}^n d_t Z_{st}(\tilde{g}_s - \tilde{g}_t - d_t Z_{st}^T\gamma_0)H_{b,st}$ is modified to $C_n^\gamma = n^{-2}\sum_{s=1}^n\sum_{t=1}^n d_t Z_{st}(\tilde{g}_s - \tilde{g}_t - d_t Z_{st}^T\gamma_0)K_{h,st}^\gamma$. The leading term C_{1n} is modified to

$$\begin{aligned} C_{1n}^\gamma &= \frac{1}{n^2}\sum_{s=1}^n\sum_{t\neq s}^n d_{t,0}Z_{st}(g_s - g_t - d_{t,0}Z_{st}^T\gamma_0)K_{h,st}^\gamma \\ &= \frac{1}{n(n-1)}\sum_{s=1}^n\sum_{t\neq s}^n d_{t,0}Z_{st}(g_s - g_t - d_{t,0}Z_{st}^T\gamma_0)K_{h,st}^\gamma + O_p(n^{-1}) \\ &\equiv C_{1n,0} + O_p(n^{-1}), \end{aligned} \tag{6}$$

where the $O_p(n^{-1})$ term comes from the $O(n^{-3})$ in $1/n^2 = 1/[n(n-1)] + O(n^{-3})$.

Note that $C_{1n,0}$ can be written as a second order U-statistic. Define

$$H_{n,st} = \{d_{t,0}Z_{st}[g_s - g_t - \beta'(Z_t\gamma_0)Z_{st}^T\gamma_0] + d_{s,0}Z_{ts}[g_t - g_s - \beta'(Z_s\gamma_0)Z_{ts}^T\gamma_0)]K_{h,st}^\gamma\}/2.$$

Then by the H-decomposition of a U-statistic, we have

$$\begin{aligned} C_{1n,0}^\gamma &= \frac{2}{n(n-1)}\sum_{s=1}^{n-1}\sum_{t>s}^n H_{n,st} \\ &= M + \frac{2}{n}\sum_{t=1}^n[H_{n,t} - M] + \frac{2}{n(n-1)}\sum_{s=1}^{n-1}\sum_{t>s}^n[H_{n,st} - H_{n,s} - H_{n,t} + M] \end{aligned} \tag{7}$$

where $M = E[H_{n,st}]$, $H_{n,t} = E[H_{n,st}|Z_t]$. It is straightforward to show that $M = h^2D + o(h^2)$, where D is a constant, and $H_{n,t} = h^2 D_t$, where $E(D_t) = D + o(1)$ and $Var(D_t) = O(1)$. The third term in the H-decomposition has a smaller order than the first two terms. Hence, we have

$$C_{1n,0}^\gamma = O_p(h^2 + h^2 n^{-1/2}) = O_p(h^2).$$

Note that $C_{1n}^{\gamma} = O_p(n^{-1/2})$ if one selects h such that nh^4 is bounded, $h \to 0$ and $nh \to \infty$ as $n \to \infty$.

A tedious proof leads to $C_n^{\gamma} - C_{1n}^{\gamma} = O_p(b^{\nu} + h^2 + n^{-1/2})$. This then implies that $C_n^{\gamma} = O_p(h^2 + n^{-1/2})$, which in turn implies that $\hat{\gamma} - \gamma_0 = O_p(h^2 + n^{-1/2})$ because $A_n^{\gamma} = O_p(1)$ (an exact order, i.e., $A_n \neq o_p(1)$). Hence,

$$\hat{\gamma} - \gamma_0 = [A_n^{\gamma}]^{-1} C_n^{\gamma} = O_p(b^{\nu} + h^2 + n^{-1/2}).$$

This completes the proof of Theorem 4.3.2.

One can also derive the asymptotic distribution of $\sqrt{n}(\hat{\gamma} - \gamma_0)$. It can be shown that the asymptotic variance of $\sqrt{n}(\hat{\gamma} - \gamma_0)$ comes from C_{2n}^{γ}. Alternatively, one can also use some bootstrap methods to do inferences.