经济与金融高级研究丛书

Advanced Studies in Economics and Finance

丛书主编　邹恒甫

Editor in Chief　Heng-fu Zou

计量经济学：半参数计量经济学方法

Econometrics: Semiparametric Econometric Methods

艾春荣　陈小红　著

Chunrong Ai and Xiaohong Chen

北京大学出版社

Peking University Press

经济与金融高级研究丛书
Advanced Studies in Economics and Finance
丛书主编　邹恒甫
Editor in Chief　Heng-fu Zou

计量经济学：半参数计量经济学方法

Econometrics: Semiparametric Econometric Methods

艾春荣　陈小红　著
Chunrong Ai and Xiaohong Chen

北京大学出版社
Peking University Press

图书在版编目(CIP)数据

计量经济学：半参数计量经济学方法/艾春荣、陈小红著．－北京：北京大学出版社，2000

ISBN 7-301-04564-6

Ⅰ．半…　Ⅱ．①艾…　②陈…　Ⅲ．计量经济学　Ⅳ．F224.0

中国版本图书馆 CIP 数据核字(2000)第 61865 号

书　　　名：**计量经济学：半参数计量经济学方法**
著作责任者：艾春荣　陈小红
责 任 编 辑：梁鸿飞
标 准 书 号：ISBN 7-301-04564-6/F·346
出　版　者：北京大学出版社
地　　　址：北京市海淀区中关村北京大学校内　100871
网　　　址：http://cbs.pku.edu.cn/cbs.htm
电　　　话：出版部 62752015　发行部 62754140　编辑部 62753121
电 子 信 箱：zpup@pup.pku.edu.cn
排　版　者：兴盛达打字服务社　82614608
印　刷　者：北京大学印刷厂印刷
发　行　者：北京大学出版社
经　销　者：新华书店
787 毫米×1092 毫米　16 开本　14.5 印张　240 千字
2000 年 10 月第 1 版　2000 年 10 月第 1 次印刷
定　　　价：30.00 元

经济与金融高级研究丛书
出版前言

中国的经济学和金融学研究如何走向世界?这是一个值得探讨的问题。中国经济学者素以刻苦求知、真诚报国为荣。在国内,自改革开放以来经济学者的突出贡献已深得国人认同;在海外,中国的经济学者同样做出了可喜的成绩。但是,国内经济学者的学术成果得到国际上认可的为数寥寥,而海外中国经济学者所取得的学术成果在国内也鲜为人知。同时,国际经济学者的学术成果在国内的传播也很有限。凡此种种,原因当然是多方面的,其中之一是学术传播与交流上的障碍。这些障碍的存在造成彼不知我,我亦不知彼,国内经济学者的学术研究难以走向世界,国际经济学者和海外中国经济学者的学术研究难以走进中国这样一种尴尬的局面。不言而喻,在全球经济一体化趋势主导世界潮流的今天,这种状况不利于中国经济和中国经济学的发展。

随着改革开放的一步步深化,中国经济与世界经济日益接轨。世界各国经济学者对中国经济发展和中国经济研究的兴趣和热情有增无减。海内外中国经济学者的拳拳报国之心也日益高涨。科学无国界,学术交流也无国界。我们相信,学者们的热情与努力将冰释学术交流中的所有障碍。因此,在经济全球化的今天,在经济腾飞指日可待的中国,这套《经济与金融高级研究丛书》的出版是时代的要求,更是我们的历史使命。

ZJ06/24

本套丛书将尽可能全面地收录国际经济学者特别是中国经济学者在国际上已获得公认的学术成果。每部著作将基本保留其最初发表在国际刊物上的原貌(或其创作的原貌)，由作者按研究专题编纂成书。此举一方面是为了让更多的国人了解这些学者的研究成果，或者至少感知一下国际经济学者和海内外中国经济学者在国际主流经济学发展进程中所迈出的坚实的步伐，从而激励更多的青年学子求知问道；另一方面也是为了使世界各国的经济学者对中国经济学者的研究成果有更多和更全面的了解，或者至少感知到中国的经济学研究并非固步自封置身世界之外，而是与世界同步与潮流并进的。知己知彼，互相交流，这对于繁荣学术是有百利而无一弊的。北京大学出版社真诚地希望更多的海内外学者向我们赐稿，并给我们批评、建议，以助于这项造福世人的学术文化传播事业。

北京大学出版社

经济和金融高级研究丛书
编者说明

本丛书收录世界各国经济学者特别是海内外中国经济学者从事当代经济学和金融学理论研究和实际研究的前沿成果。就某一专题或者多个专题，作者既可以把已经发表的论文收集成册，也可以修辑整理成一部或多部专著。收集成册的公开发表的论文一律保持其发表时的各刊物排版印制的原貌，以方便读者查寻援引；尚未公开发表的论文则一律保持其创作原貌，以供读者参考。

本丛书主编同时还与海内外众多学者合作主办英文学术刊物 *Annals of Economics and Finance*。此刊物出版尚未发表的至少具有一些原创性的经济学和金融学（英文）论文。如有兴趣借此刊物宣布自己学术思想的学人，敬请寄论文给：

北京大学光华管理学院 *Annals of Economics and Finance* 主编
——邮编 100871

或者

武汉大学高级研究中心 *Annals of Economics and Finance* 主编
——邮编 430072

但愿此丛书和杂志能促进中国经济学者与世界各国经济学者的学术交流，促进中国经济学和金融学研究走向世界主流。

邹恒甫

于北京大学

Acknowledgments

The Peking University Press wishes to thank the following who have kindly given permission for the use of copyright material.

Blackwell Publishers for articles: 'A Semiparametric Maximum Likelihood Estimator', *Econometrica*, **65**(4): 933 - 963, 1997; 'Sieve Extremum Estimators for Weakly Dependent Data', *Econometrica*, **66**(2): 289 - 314, 1998.

Elsevier Science Publishers for articles: 'Estimation of Some Partially Specified Nonlinear Models', *Journal of Econometrics*, **76**: 1 - 37, 1997; 'Consistent Hypothesis Testing in Semiparametric and Nonparametric Models for Econometric Time Series', *Journal of Econometrics*, **91**: 373 - 401, 1999. 'A Semiparametric Efficiency Bound of a Disequilibrium Model without Observed Regime', *Journal of Econometrics*, **62**: 143 - 163, 1994.

Overseas Publishers Association for article: 'An Improved Estimator for Models with Randomly Missing Data', *Nonparametric Statistics*, **7**: 331 - 347, 1997.

前　言

计量经济学与统计学的差别在于计量经济学的模型是以经济学理论为基础的。经济学理论描述经济变量之间的关系,却没有提供这种关系的具体函数形式,也没有提供计量经济学模型中误差项(error term)的统计分布形式。在实际应用中,由于真实的函数和统计形式未知,研究人员通常假设线性的函数形式和正态分布。这种假设虽然简便,却容易导致错误的模型和错误的统计分布结果。因此,如何建立正确的计量经济模型是计量经济学家面临的一大难题。

解决这个难题有两个途径。第一个途径是使用非参数模型(nonparametric model)。所谓非参数模型是不对未知的函数和统计分布形式作任何具体的假定。比如说,我们想知道价格(x)对需求(y)的影响,我们不必要设定 y 和 x 之间的函数关系,也不必假定 y 的条件分布形式(对每一个 x 值)。我们可以设定,对每一个 x 值,y 的条件均值是一个未知的函数 $f(x)$。然后,我们应用非参数统计方法去估计 $f(x)$。非参数统计方法的优点是它的稳健性(Robustness),永远不会错误地估计 $f(x)$。但非参数统计方法有两个弱点。第一个弱点是它要求大量的数据。这一点在有些应用中很难满足。第二个弱点是 $f(x)$的估计值收敛到 $f(x)$的速度缓慢,令人很不满意。

矫正第二个弱点的办法是研究由 $f(x)$延伸出来的参数。比方说,一个政府部门的决策人员或一个公司销售部门的经理更感兴趣的是价格的变化如何影响整个市场的需求。价格的变化对某一位消费者的需求的影响可以用一阶导数$\frac{\mathrm{d}f(x)}{\mathrm{d}x}$来表达,而价格的变化对整个市场的平均需求的影响用平均导数(average derivative)来表达。如何正确地估计平均导数并使得估计值达到参数估计值的收敛速度是20世纪80年代中期的一个重要的研究课题。包维尔(Powell),斯多克(Stoker)和瓦森(Watson)等人提出了用非参数的方法估计 $f(x)$,然后用平均的方法估计平均导数。当时非参数统计方法被引进到计量经济学不久,计量经济学家知道 $f(x)$的估计值的收敛速度很慢,因而推断平均导数估计值的收敛速度也很慢。包维尔等人的贡献在于说明平均导数估计值的收敛速度可以达到参数估计值的收敛速度。这一研究结果对后来半参数计量经济学的研究起了推动的作用。

第二个途径是使用半参数模型(semiparametric model)。所谓半参数模型是指对经济变量关系作部份的设定。以上面的需求函数为例，我们可以设定需求函数为一指数形式(index)：$f(x)=F(x\beta)$，这里 $F(\cdot)$是一未知函数。半参数模型优于非参数模型之处在于它把高维的未知函数问题(例如，$f(x)$是一个高维的函数，如果 x 的维数很高)变为一个低维的问题(通常 $x\beta$ 的维数低于 x 的维数)，因而对数据样本的需求不是那么强烈。这在某种程度上矫正了非参数模型的第一个弱点。目前，半参数模型是微观计量经济学里的一个热门研究课题。本书收录了我们已经发表或即将发表的论文，共8篇。希望能对国内的统计学和计量经济学研究人员了解这一发展方向有所帮助。

为了便于读者了解半参数方法发展的历史和现状，我们将收录的论文按照历史的相关性来排列。因此，排在前面的文章不一定比排在后面的文章先发表，也不一定比排在后面的文章更重要。只是它们研究的问题与半参数领域早期的研究相关而已。

回到上面消费函数指数形式 $f(x)$：$F(x\beta)$的例子。如何估计未知参数 β 而不对未知函数$F(\cdot)$作任何的设定？包维尔等人提出的平均导数估计值可以用来估计 β。但平均导数估计值有一个致命的弱点，它要求 x 变量之间互相独立。也就是说，x 不应同时含有收入和收入的平方两项。这极大地限制了平均导数估计值的应用。本书第一篇文章修正了平均导数估计值的这一弱点。

与平均导数研究差不多同一时期，计量经济学家对所谓的部份线性模型也投入了相当的精力。所谓部份线性模型是指对 $f(x)$部分设定 $f(x)=x_1\beta+F(x_2)$，这里 $F(\cdot)$是未知函数。在这一方面的代表作是罗宾森(Robinson)发表在 *Econometrica* 上的文章。本书收入的第二篇文章把罗宾森的方法推广到非线性模型。

本书收录的第三篇文章把半参数的方法应用到失落数据(missing data)问题。

在对半参数估计值的一致性(consistency)和收敛速度进行研究的同时，计量经济学家对半参数估计值的有效性(efficiency)也进行了探讨。衡量一个半参数估计值是否有效需要一个准则。这个准则就是半参数估计值的上界。计算半参数估计值的上界是一个十分困难，又吃力不讨好的工作。本书收录的第四篇文章对某一特定的模型计算了半参数估计值的上界。

知道了半参数估计值的上界后，一个最具挑战性的问题是怎样找到一个半参数估计值，达到这一上界。在相当长的一段时间里，这一直是道难题。本书收录的第五篇文章对这一问题给出了部份答案，而第六篇文章给出了全面的和满意的答案。

半参数计量经济学方法，虽然是由微观计量经济学家最先研究的，但也仍然适

用于时间序列数据。本书收录的第七篇文章是把半参数方法推广到时间序列的最系统的研究之一。半参数方法还有一个应用，就是用来对计量经济学模型进行统计推断(statistical inference)。近来在这一方面的研究正方兴未艾。大多数文章只讨论一些个别模型的推断问题，缺少一般性的、系统的研究。本书收录的最后一篇文章朝着系统性的研究跨出了第一步。我们正在从事着这一系统的研究，希望在不久的将来能够全面地、满意地解决半参数统计推断问题。

最后，我们对邹恒甫教授和北京大学出版社对出版本书所给予的协助表示感谢，对我们的老师、同学和同事所给予的指导和鼓励致以谢意。

艾春荣　陈小红

2000 年 5 月

Contents

1

A MODIFIED AVERAGE DERIVATIVES ESTIMATOR

Chunrong Ai

Department of Economics, University of Florida
Gainesville, FL 32611, USA

Key Words and Phrases: Nonparametric estimation; Functional dependency; Average derivatives estimator.

JEL Classification: C2, C4.

ABSTRACT

We extend the average derivatives estimator to the case of functionally dependent regressors. We show that the proposed estimator is consistent and has a limiting normal distribution. A consistent covariance matrix estimator for the proposed estimator is provided.

1. INTRODUCTION

Brillinger (1983) proposed a least squares estimator for the model $E\{y|x\}=F(x'\beta_0)$, where $F(.)$ is an unknown function. He showed that, when the regressors x are normally distributed, a least squares regression of y on x yields a consistent estimator for some multiple of β_0. Since then, his estimator has been extended to the case where the joint distribution of x may be nonnormal. Ruud (1986) and Li (1991), for example, replaced the normal distribution with the one satisfying some symmetry condition, while Powell et al. (1989) and Härdle and Stoker (1989) only required continuous and functionally independent regressors.[1] A common attraction of all those studies is that their estimators are computationally simple, requiring no numerical search. However, the condition that the regressors are functionally independent is too restrictive. In practice, higher order terms such as squared and interaction terms are commonly used to model nonlinearity. The regressors often are functionally dependent. Also, the single index specification of the conditional mean of y is restrictive. In some applications, the conditional mean of y may have a semilinear form (see examples below). The main objective of this paper is to present a computationally simple estimator for the case where the regressors may be functionally dependent and the conditional mean of y has a semilinear form:

$$E\{y|x,z\} = z'\alpha_0 + F(x'\beta_0) \text{ for some } \alpha_0 \in \mathcal{A} \text{ and } \beta_0 \in \mathcal{B}, \quad (1.1)$$

where $x'\beta_0$ and $z'\alpha_0$ are known parametric parts but F(.) is an unknown function.

The extension to the semilinear form is important since many econometric models are included as special cases of (1.1). For example, the bivariate censored regression model: $y = 1\{x'\beta_0 - u\}(z'\alpha_0 + \varepsilon)$, with (ε, u) independent of x and z, satisfies (1.1) with $F(x'\beta_0) = E\{\varepsilon|u < x'\beta_0\}$. The binary choice model: $y = 1\{x'\beta_0 - \varepsilon > 0\}$, with ε independent of x, satisfies (1.1) with $\alpha_0 = 0$ and F(.) as the cumulative distribution function of ε. The disequilibrium model with unknown regime: $y = \mathrm{Min}\{z'\alpha_0 + \varepsilon, v'\gamma_0 + u\}$, with (u, ε) independent of (z,v), also satisfies (1.1) with some function F(.) and $x'\beta_0 = z'\alpha_0 - v'\gamma_0$. Other models satisfying (1.1) include selection models, ordered choice models, duration models, and disequilibrium models with observed regime [see Maddala (1983) for definitions of those models].

Our proposal of estimating α_0 and β_0 is analogous to the average derivatives estimator (hereafter ADE) proposed by Powell et al. (1989). The ADE is based on the derivatives: $\partial m(x,z)/\partial z = \alpha_0$ and $\partial m(x,z)/\partial x = F'(x'\beta_0)\beta_0$, where $m(x,z) = E\{y|x,z\}$. Let $\hat{m}(x,z)$ denote a nonparametric estimator of m(x, z) and $\{(x_i,z_i,y_i), i = 1, 2, ..., N\}$ denote a sample of observations. Then, the ADE for some scale multiple of β_0 is the sample average of the derivative $\partial\hat{m}(x_i, z_i)/\partial x$, and the ADE of α_0 is the sample average of the derivative $\partial\hat{m}(x_i, z_i)/\partial x$. Under the condition that the regressors are functionally independent and some additional conditions, Powell et al. showed that the ADE is $\sqrt{N}$ consistent.

The ADE may not be $\sqrt{N}$ consistent, however, when the regressors are functionally dependent. To see this, suppose that z and v are functionally independent scalars. Let $x = (z,v)'$. Then x and z are functionally dependent because $E\{z|x\} = z$. The joint density of x and z, denoted by f(x,z), is discontinuous and nondifferentiable with respect to x and z. Two methods are commonly used to estimate m(x,z): kernel and sieve. The kernel method estimates m(x,z) by estimating $h(x,z) = m(x,z)f(x,z)$ and f(x,z) respectively. Since neither h(.) nor f(.) is differentiable, the kernel estimators do not converge to h(.) and f(.) in probability as rapidly as required by Powell et al. (1989). Thus, the kernel-based ADE may not be $\sqrt{N}$ consistent. The sieve method approximates m(.) by a sieve such as a second order polynomial:

$$\lambda_0 + \lambda_1 z + \lambda_2 z^2 + \lambda_3 x_1 + \lambda_4 x_1^2 + \lambda_5 x_2 + \lambda_6 x_2^2 + \lambda_7 x_1 x_2 + \lambda_8 z x_1 + \lambda_9 z x_2,$$

where x_1 and x_2 denote the first and the second element of x and $\{\lambda_j, j = 1,2, ...,9\}$ are coefficients to be estimated by regressing y on those polynomial terms. The derivative $\partial m(x,z)/\partial x_1$ is then approximated by $\lambda_3 + 2\lambda_4 x_1 + \lambda_7 x_2 + \lambda_8 z$. To estimate the derivative $\partial m(x,z)/\partial x_1$, we need estimates of λ_3, λ_4, λ_7, and λ_8. Unfortunately, since the regressor x_1 is the same as z, only the coefficients: λ_0, λ_5, λ_6, $(\lambda_1 + \lambda_3)$, $(\lambda_7 + \lambda_9)$, and $(\lambda_2 + \lambda_4 + \lambda_8)$ can be estimated. Thus, the sieve method can not even estimate $\partial m(x,z)/\partial x_1$.

2. ESTIMATOR

Clearly, the ADE needs to be modified so that it accounts for the functional dependency explicitly. To do so we express the regressors as some measurable and known transformations of deep regressors w: $x = x(w)$ and $z = z(w)$. Equation (1.1) now writes

$$E\{y|w\} = z(w)'\alpha_0 + F(x(w)'\beta_0). \tag{1.1}'$$

Suppose that the regressors w are continuous and functionally independent, and that F(.), x(.), and z(.) are differentiable with respect to their own arguments. With $m(w) = E\{y|w\}$ and differentiating both sides of equation (1.1)' with respect to the j-th component of w, we obtain:

$$m_j(w) = z_j(w)'\alpha_0 + F'(x(w)'\beta_0)*x_j(w)'\beta_0, \tag{1.2}$$

where the subscript j denotes the first derivative with respect to the j-th component of w and F′(.) denotes the first derivative of F(.) with respect to its argument.

Ichimura and Lee (1991, Lemma 2) showed that a necessary condition for identifying α_0 and β_0 is the exclusion restriction that one element of x whose coefficient is nonzero (and thus can be normalized to one) is excluded from z. Without loss of generality, suppose that the first element of x satisfies such exclusion restriction. Furthermore, suppose that the first element of x is identical to the first element of w.

ASSUMPTION 0: *(i) F(·), x(·), and z(·) are differentiable with respect to their own arguments; (ii) all elements of w are functionally independent and continuous; and (iii) the first element of w is identical to the first element of x, which has a nonzero coefficient and is functionally independent of z and other elements of x.*

Under these conditions, we have: $x_{11}(w) = 1$, $x_{1j}(w) = 0$ for $j = 2, ..., k$, $z_1(w) = 0$, and $x_{21}(w) = 0$, where $x_{1j}(w)$ and $x_{2j}(w)$ denote the derivative of the first and all other elements of x with respect to the j-th element of w respectively. Normalizing the coefficient on the first element of x to one: $\beta_0 = (1, \beta_{20}')'$, equation (1.2) gives $m_1(w) = F'(x(w)'\beta_0)$. Substituting this expression into (1.2) for $j > 2$, we obtain:

$$m_j(w) = z_j(w)'\alpha_0 + m_1(w)*x_{2j}(w)'\beta_{20}, j = 2, ..., k. \tag{1.3}$$

Equation (1.3) can be estimated by applying a pooled least squares regression.

To apply the pooled least squares regression, we need to estimate m(w). We estimate m(w) by the kernel method. Let $\{(w_i,y_i), i = 1, ..., N\}$ denote a sample of independent and identically distributed observations. Let K(.) denote a kernel function and let h_N denote a bandwidth satisfying $h_N \to 0$ as $N \to +\infty$. Let f(w) denote the density function of w and define g(w) = m(w)f(w). Then, m(w) = g(w)/f(w). We estimate $m(w_i)$ by estimating $g(w_i)$ and $f(w_i)$ respectively; and we estimate $g(w_i)$ and $f(w_i)$ by the Rosenblatt-Parzen kernel density estimators:

$$\hat{g}(w_i) = \frac{1}{(N-1)h_N^k}\sum_{s\neq i}^{N} y_s K((w_i - w_s)/h_N);\ \hat{f}(w_i) = \frac{1}{(N-1)h_N^k}\sum_{s\neq i}^{N} K((w_i - w_s)/h_N).$$

The estimator of $m(w_i)$ is $\hat{m}(w_i) = \hat{g}(w_i)/\hat{f}(w_i)$ and the estimator of $m_j(w_i)$ is $\hat{m}_j(w_i)$, the first derivative of $\hat{m}(w)$ with respect to the j-th element of w. Let θ_0 denote the union of α_0 and β_{20} and define $X_j(w) = (z_j(w)', m_1(w)^* x_{2j}(w)')'$. Define $Y(w) = (m_2(w), ..., m_k(w))'$ and $X(w) = (X_2(w), ..., X_k(w))$. Equation (1.3) writes $Y(w) = X(w)'\theta_0$. Let $\hat{Y}(w_i)$ and $\hat{X}(w_i)$ denote the estimators of $Y(w_i)$ and $X(w_i)$ respectively, obtained by substituting $\hat{m}_j(w_i)$ for $m_j(w_i)$. The proposed estimator of θ_0 is:

$$\hat{\theta} = \{\sum_{i=1}^{N} 1\{w_1 \in W_1\}\hat{X}(w_i)\hat{X}(w_i)'\}^{-1}\{\sum_{i=1}^{N} 1\{w_i \in W_1\}\hat{X}(w_i)\hat{Y}(w_i)\},$$

where $\mathbf{W}_1$ is a strict subset lying in the interior of **W** (the support of w). The indicator function $1\{w_i \in \mathbf{W}_1\}$ is used for trimming purpose.

3. ASYMPTOTIC RESULTS

We now derive the asymptotic results of the proposed estimator. Write:

$$\sqrt{N}(\hat{\theta} - \theta_0) = \{N^{-1}\sum_{i=1}^{N} 1\{w_i \in W_i\}\hat{X}(w_i)\hat{X}(w_i)'\}^{-1} * N^{-1/2}\sum_{i=1}^{N} 1\{w_i \in W_1\}\hat{X}(w_i)\varepsilon_i,$$

where $\varepsilon_i = \hat{Y}(w_i) - Y(w_i) - [\hat{X}(w_i) - X(w_i)]'\theta_0$ denotes the nonparametric estimation error of $Y(w_i)$ and $X(w_i)$. Substituting the expression for ε_i, we obtain:

$$N^{-1/2}\sum_{i=1}^{N} 1\{w_i \in W_1\}\hat{X}(w_i)\varepsilon_i = N^{-1/2}\sum_{i=1}^{N} 1\{w_i \in W_1\}\hat{X}(w_i)[\hat{Y}(w_i) - Y(w_i)]$$

$$-N^{-1/2}\sum_{i=1}^{N} 1\{w_i \in W_1)\hat{X}(w_i)[\hat{X}(w_i) - X(w_i)]'\theta_0.$$

Under some sufficient conditions, we show that:

$$N^{-1}\sum_{i=1}^{N} 1\{w_i \in W_1\}[\hat{X}(w_i)\hat{X}(w_i)' - X(w_i)X(w_i)'] = o_p(1), \tag{3.1}$$

$$N^{-1/2}\sum_{i=1}^{N} 1\{w_i \in W_1\}[\hat{X}(w_i)[Y(w_i) - \hat{Y}(w_i)] - s_1(w_i)u_i] = o_p(1),$$

$$N^{-1/2}\sum_{i=1}^{N} 1\{w_i \in W_1\}[\hat{X}(w_i)[\hat{X}(w_i) - X(w_i)]'\theta_0 - s_2(w_i)u_i] = o_p(1),$$

where $u_i = y_i - m(w_i)$, $s_1(w_i) = \sum_{j=2}^{k}\{\partial[X_j(w_i)f(w_i)]/\partial w_{ji}\}/f(w_i)$, w_{ji} denotes the j-th element of w_i, $s_2(w_i) = \sum_{j=2}^{k}\{\partial[X_j(w_i)X_{2j}(w_i)'\beta_{20} * f(w_i)]/\partial w_{1i}\}/f(w_i)$. Combining the above results, we obtain

$$N^{-1/2}\sum_{i=1}^{N} 1\{w_i \in W_1\}\hat{X}(w_i)\varepsilon_i = -N^{-1/2}\sum_{i=1}^{N} 1\{w_i \in W_1\}s(w_i)u_i + o_p(1), \tag{3.2}$$

where $s(w_i) = s_1(w_i) + s_2(w_i)$. We then invoke the Weak Law of Large Numbers to obtain

$$N^{-1}\sum_{i=1}^{N} 1\{w_i \in W_1)X(w_i)X(w_i)' = E\{1\{w \in W_1\}X(w)X(w)'\} + o_p(1) \tag{3.3}$$

and the Lyapunov central limit theorem to obtain:

$$N^{-1/2}\sum_{i=1}^{N} 1\{w_i \in W_1\}s(w_i)u_i \xrightarrow{d} N(0, V), \tag{3.4}$$

where $\xrightarrow{d}$ denotes convergence in distribution, $V = E\{1\{w \in \mathbf{W}_1\}s(w)s(w)'\sigma^2(w)\}$ and $\sigma^2(w) = E\{u^2|w\}$. Hence $\sqrt{N}(\hat{\theta} - \theta_0)$ is asymptotically normally distributed with mean zero and covariance matrix $\Omega = [E\{1\{w \in W_1\}X(w)X(w)'\}]^{-1} * V * [E\{1\{w \in W_1\}X(w)X(w)'\}]^{-1}$.

To estimate the asymptotic covariance matrix Ω. Let $\hat{s}(w_i)$ denote the estimator of $s(w_i)$, obtained by substituting $\hat{f}(w_i)$ for $f(w_i)$ respectively. We estimate V and Ω by:

$$\hat{V} = N^{-1}\sum_{i=1}^{N} 1\{w_1 \in W_i\}\hat{s}(w_i)\hat{s}(w_i)'(y_i - \hat{m}(w_i))^2,$$

$$\hat{\Omega} = [N^{-1}\sum_{i=1}^{N} 1\{w_i \in W_1\}\hat{X}(w_i)\hat{X}(w_i)']^{-1}\hat{V}\,[N^{-1}\sum_{i=1}^{N} 1\{w_i \in W_1\}\hat{X}(w_i)\hat{X}(w_i)']^{-1}.$$

To ensure the validity of (3.4), we need the following conditions:

ASSUMPTION 1. $E\{1\{w \in W_1\}X(w)X(w)'\}$ *is nonsingular.*

ASSUMPTION 2. (i) $\{(w_i, y_i)$ *for* $i = 1, 2, \ldots, N\}$ *are i.i.d.; and (ii)* $E\{1\{w \in W_1\}s(w)s(w)'\sigma^2(w)\}$ *exists and is finite.*

ASSUMPTION 3. W_1 *is compact.*

ASSUMPTION 4. $\sigma^2(w)$ *is continuous in w.*

ASSUMPTION 5. $x(w)$, $z(w)$, *and* $m(w)$ *are continuous in w, and have up to order of* $p+1$ *derivatives with respect to w, where* $p > k+2$.

ASSUMPTION 6. $f(w)$ *is bounded away from zero on* W_1 *and has up to order of* $p+1$ *derivatives with respect to w, where* $p > k+2$.

ASSUMPTION 7. The kernel function $K(u)$ *is symmetric with compact support and for some* $p > k+2$ *satisfies:*

(i) $\int K(u)du = 1$;

(ii) $\int u_1^{l_1} u_2^{l_2} \ldots u_k^{l_k} K(u)du = 0$ *for all* $l_1 + l_2 + \ldots l_k < p$

(iii) $\int u_1^{l_1} u_2^{l_2} \ldots u_k^{l_k} K(u)du \neq 0$ *for some* $l_1 + l_2 + \ldots + l_K = p$

ASSUMPTION 8. The bandwidth h_N *satisfies: (i)* $h_N \to 0$ *as* $N \to +\infty$; *(ii) for some small* $\varepsilon > 0$, $N^{1-\varepsilon} h_N^{2(k+2)} \to +\infty$; *and (iii)* $Nh_N^{2p} \to 0$.

Under Assumptions 0 - 8, we show that equations (3.1) - (3.4) hold.

Theorem 1. Under Assumptions 0 - 8, we show: $N^{1/2}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Omega)$ and $\hat{\Omega} = \Omega + O_p(1)$.

The proposed estimator can be easily extended to multiple equations and multiple indexes models. The multiple equations and/or multiple indexes models can be derived from a multinomial discrete choice model, or from a selection model in which the dependent variables are limited in ranges according to a multinomial discrete choice model [see Maddala (1983) for definitions of those models]. To illustrate the extension, consider a two equations system:

$$m_1(w_1,v_1) = E\{y_1|w\} = z_1(w_1)'\alpha_{10} + F(v_1 + x_1(w_1)')\beta_{10}),$$

$$m_2(w_2,v_2) = E\{y_2|w\} = z_2(w_2)'\alpha_{20} + G(v_2 + x_2(w_2)'\beta_{20}),$$

where w is the union of (w_1, w_2, v_1, v_2). Suppose that v_1 is functionally independent of w_1, and that v_2 is functionally independent of w_2. Furthermore, suppose that the regressors w_1 are functionally independent and the regressors w_2 are functionally independent. We then have

$$\partial m_1(w_1,v_1)/\partial v_1 = F'(v_1 + x_1(w_1)'\beta_{10}),\ \partial m_2(w_2,v_2)/\partial v_2 = G'(v_2 + x_2(w_2)'\beta_{20}).$$

Eliminating F and G, we obtain:

$$\partial m_1(w_1, v_1)/\partial w_1 = [\partial z_1(w_1)/\partial w_1']'\alpha_{10} + \partial m_1(w_1, v_1)/\partial v_1 * [\partial x_1(w_1)/\partial w_1']'\beta_{10},$$