ESSAYS IN MODERN ECONOMICS RESEARCH
IN HONOUR OF PROFESSOR GREGORY CHOW

现代经济学论文集 上册

献给邹至庄教授

主编 周林 陶志刚 谢丹阳 宋敏

格致出版社 上海人民出版社

ESSAYS IN MODERN ECONOMICS RESEARCH

IN HONOUR OF PROFESSOR GREGORY CHOW

现代经济学论文集 (上册)

献给邹至庄教授

主编 周林 陶志刚 谢丹阳 宋敏

格致出版社 上海人民出版社

编委及作者（按汉语拼音排序）

艾春荣 白聚山 陈百助 陈松年 陈智琦 褚雪浩 范延琴
方星海 符育明 洪　露 胡祖六 李稻葵 李华刚 李　山
林　平 马成虎 梅建平 倪晓光 丘东晓 宋　敏 宋顺锋
谭国富 陶志刚 王建新 王汝渠 王苏生 魏尚进 魏向东
吴延瑞 谢丹阳 谢　新 许定波 许小年 徐永胜 查　涛
张　处 张建波 张俊森 张俊喜 赵海英 郑　旭 周　林

序

在中国经济改革开放三十周年之际，我们一些在国内外从事经济学研究的学者、政府部门专家以及实业界人士合议编纂一部现代经济学研究论文集。本文集的作者都是在20世纪80年代中后期，通过美国普林斯顿大学邹至庄教授与原国家教委的合作项目赴美国、加拿大等国的著名大学留学并获得经济学博士学位。本论文集精选了每位作者在顶尖国际经济学研究书刊上发表的代表作，从一个侧面反映了80年代后来自中国的新一代经济学者在现代经济学研究方面取得的成就。我们在此衷心感谢邹至庄教授为我们提供的宝贵机会，更加感谢他为现代经济学在中国的推广和发展作出的重要贡献。同时，我们还要衷心感谢教育部高瞻远瞩，在改革开放初期选派优秀学生出国深造，不仅推动了现代经济学发展，而且为中国经济建设培养了大批经济专家。

邹至庄教授，英文名Gregory Chow，是世界著名的美籍华人经济学家，美国普林斯顿大学荣誉教授，美国哲学学会、世界计量经济学学会和美国统计学会名誉会员。邹教授于1951年取得康乃尔大学学士学位，并于1952年和1955年在芝加哥大学分别获得硕士和博士学位。毕业后邹教授曾在麻省理工学院、康乃尔大学、IBM等著名大学和研究机构任职，自1970年起任教于普林斯顿大学并担任计量经济学研究中心主任，长达三十余载。为了表彰他的贡献，该中心于2001年在邹教授正式退休之际更名为邹至庄计量经济研究中心。1960年邹至庄教授发表成名作“Tests of Equality between Sets of Coefficients in Two Linear Regressions”后以“邹氏检验”（Chow Test）著称于国际经济学界。该论文在发表后几年内便成为全球经济学界引用次数最高的论文之一。此后邹至庄教授又在动态经济学以及控制论在经济学中的应用方面作出了开创性的工作。至今邹教授已撰写了13部经济学方面的专著，并在经济学国际顶级学术期刊上发表论文200多篇。

邹至庄教授在20世纪60年代曾担任中国台湾地区的经济顾问，对台湾地区经济的腾飞作出了直接的贡献。自1980年起，邹至庄教授将他的主要精力转向中国大陆。他频繁地访问中国大陆，与国家领导人、政府经济及教育部门主管、各重点大学领导、经济学家、学生等各界人士广泛接触与交流，对中国大陆经济的发展和现代经济学在中国大陆的推广提出很多宝贵建议，并亲自主持了很多研究和教学项目。

邹至庄教授亲自主持的重要项目之一是从1985年至1987年国家教委举办的三届选派学生出国留学攻读经济学博士的考试。此项目亦以第二个“邹氏检验”而著称。邹教授是在参加了1984年由国家教委主办的暑期经济学讲学班后，萌发了直接帮助中国学生赴北美主要名校留学深造的想法。随后，他致信当时国家教委领导夏自强、王复孙、王泽农，与他们洽谈通过考试选拔学生出国攻读经济学之事。考试由邹教授亲自出题，考生由国家教委在全国各重点大学挑选，择优录取后再由邹教授推荐到北美名校经济系。1985年到1987年间先后有168位学生经邹教授的推荐申请到北美各名校留学。

历时三年的邹至庄项目对于现代经济学在中国的发展产生了深远的影响。有幸通过邹教授推荐出国深造的中国学生绝大部分成为现代经济学及金融学研究方面卓有建树的学者，并且很多学者回到国内（包括香港地区）的高校任教，竭力推动现代经济学在中国的教学与发展。除了学者以外，很多受益于邹至庄项目的留学生成为了在国家政府部门任职的

专家型领导，或是经济和金融实务界的风云人物。读者可以看到，本论文集的作者中聚集了各类人才的杰出代表。可惜受时间与精力的限制，我们没有将更多的当年参加邹至庄项目的留学生的优秀论文一一选入。

本论文集收录的40余篇文章，涉及经济学和金融学等各研究领域，许多文章发表在包括*American Economic Review*、*Econometrica*、*Journal of Political Economy*和*Review of Economic Studies*四本公认的世界顶尖经济学杂志（与*Quarterly Journal of Economics*并称为五大经济学杂志），以及*Journal of Economic Theory, Journal of Econometrics*、*Rand Journal of Economics*、*Journal of Finance*和*Journal of Financial Economics*等世界一流经济学和金融学研究期刊上。这些文章可以作为年轻的读者接触现代经济学及金融学研究的一个窗口，读者们可以从阅读这些文章的过程中体会到现代经济学研究的主要特点：理论严谨、方法创新、应用广泛。这些文章实际上也是我们对邹至庄教授主持的第二个“邹氏检验”交出的第二份答卷，并恳请邹教授评判我们是否达到了他对我们的殷切期望。当然，本论文集中的文章仅代表作者个人观点，若有不当之处，请方家学者批评指正。

最后我们感谢上海世纪出版集团格致出版社的何元龙先生、孙素青女士和本书的责任编辑钱敏女士，他们为本文集的出版提供了全力的支持。同时，感谢本文集所录用论文的各原出版机构给予我们版权方面的支持，以及香港大学经济及工商管理学院博士生倪娟、余林徽和巴晴在本文集的整理和编辑过程中付出的辛勤劳动。

周　林　陶志刚　谢丹阳　宋　敏

2008年4月

A Brief History of Chow's Program

When the editors of the book told me that this volume of reprinted articles was soon going to be published and asked me to write a short history of the Chow program, I was delighted and honored. The authors of these articles went through the program that I initiated to place Chinese graduate students to pursue a PhD degree in economics in US and Canadian universities in 1984-1988. Their amazing accomplishments were testified by these 42 outstanding articles covering the fields of theory, applied micro, finance, econometrics, macroeconomics and international economics.

Having lived in the United States beginning in 1948 when I had entered Cornell as a sophomore I first returned to mainland China in 1980 to teach econometrics at the invitation of the Chinese Academy of Social Science. Since 1980 I have returned to mainland China often, about twice a year on average. A most significant visit was in June-July 1984 to organize and teach a summer workshop on micro-economics at Peking University sponsored by the Ministry of Education. When I left Beijing for Shanghai and Hangzhou I thought of the idea of a program to help place graduate students from China to study economics at American and Canadian universities. I immediately sent a telegram (before the existence of email) to Mr.Wang Zenong, Head of Bureau in the Minister of Education in charge of economics and legal education, and Mr.Wang Fusun, Head of the Bureau in charge of foreign affair, to suggest this idea. They were to be responsible for selecting the students based on a test on mathematics (analysis, modern algebra and probability) and a test on economics (with questions provided by me and based on my book, *The Chinese Economy,* Harper & Row, 1985 but available in June 1984) . About three days later when I arrived in Guangzhou visiting Zhongshan University, a telegram from Wang Fusun was awaiting me with a positive answer.

Early in January 1985, Wang Zenong sent me a list of eighty two students who had obtained 60 percent or better in both tests. In the Economics Department at Princeton, we had a list of over 60 US and Canadian universities to which we sent folders on our new PhDs as possible job candidates. I simply used this list to send a letter concerning our program. This letter was signed also by two to three other economists whom I had invited to teach in the summer workshops that I had organized.

I tried to send the letter to two to three universities for each student. It was not easy to determine which schools to select for each student. Without better information I used the total score from the two tests to rank the students. Those with higher ranks were recommended to higher-ranked schools. There was at least one safety school for each student. Only 62 students came in the fall of 1985. There were a few more who were accepted but did not get financial aid. The Ministry of Education only provided international travel. In a few cases, the Ford Foundation provided partial support. We are indebted to the American and Canadian universities for financing the graduate education of so many Chinese students in this program. To the extent that I failed to place the best students to the top schools because of the lack of information other than the total score, my failure was compensated by the resourcefulness of the good students who were able to transfer to better schools later on once they had an opportunity to get started in a graduate school in North America.

The program went on for three years and was terminated in 1988 by the Ministry of Education. It became clear to people in the Ministry that many students had no intention to return to China to teach. This prediction was right, at least in the short run. To me, however, training talented young Chinese was a good thing in itself whether the beneficiary intends to return to China or not. I have not returned to China to teach on a full time basis myself and I am in no position to render judgment on others. When the problem of brain drain from Taiwan was discussed in the late 1960s or early 1970s I recalled my good friend and thesis adviser Arnold Harberger of the University of Chicago saying, "Whose welfare are we talking about when we discuss the welfare of Taiwan affected by the brain drain?" Are we talking about the people living in Taiwan, or the people born in Taiwan including those who choose to live abroad? Such a viewpoint may not be popular among some Chinese who have a high sense of collective welfare for those living in China. In my case, I also believe in free choice in the sense of the book *Free to Choose* by Milton and Rose Friedman. More to the point is that almost all of the authors of this volume have contributed significantly to the advance of modern economics as a discipline and modern economics education in China in various ways, whether they reside in the US for most of the time or not. This volume is an outstanding example since it will affect the way economists in China think and do research.

I am very proud of the authors and other Chow testers and I am grateful that I was able to contribute to the development of their career. What I did was simple and was not much energy consuming as explained earlier in this short history. It is their tremendous effort and talent that have accounted for their success. All of us concerned with the development of economic research, economic education and the Chinese economy in general look forward to their future contributions for years to come.

Congratulations to all authors and especially the four editors-in-chief for the publication of this volume of essays.

Gregory Chow（邹至庄）

Princeton, NJ, USA

April 2008

作者简介

艾春荣，华中科技大学本科毕业，在美国麻省理工大学获得博士学位，现为佛罗里达大学经济系教授，并兼任上海财经大学统计系系主任。

白聚山，南开大学本科毕业，在美国加州大学伯克利分校获得博士学位，现为纽约大学经济系教授。

陈百助，复旦大学本科毕业，在美国罗切斯特大学获得博士学位，现为南加州大学商学院教授。

陈松年，复旦大学本科毕业，在美国普林斯顿大学获得博士学位，现为香港科技大学经济系讲座教授和新加坡国立大学经济系教授。

陈智琦，南京大学本科毕业，在加拿大西安大略大学获得博士学位，现为加拿大卡尔顿大学经济系教授。

褚雪浩，杭州大学本科毕业，在美国加州大学尔湾分校获得博士学位，现任职于南佛罗里达大学。

范延琴，吉林大学本科毕业，在加拿大西安大略大学获得博士学位，现为美国范德比尔特大学经济系教授。

方星海，清华大学本科毕业，在美国斯坦福大学获得博士学位，现任上海市人民政府金融服务办公室主任。

符育明，上海交通大学本科毕业，在加拿大英属哥伦比亚大学获得博士学位，现任教于新加坡国立大学。

洪露，复旦大学本科毕业，在美国明尼苏达大学获得博士学位，现为美国罗尤拉大学商学院教授。

胡祖六，清华大学本科毕业，在美国哈佛大学获得博士学位，现任高盛集团亚洲董事总经理。

李稻葵，清华大学本科毕业，在美国哈佛大学获得博士学位，现为清华大学金融系讲座教授。

李华刚，中山大学本科毕业，在美国加州大学洛杉矶分校获得博士学位，现任职于瑞士信贷投资银行。

李山，清华大学本科毕业，在美国麻省理工学院获得博士学位，曾任中银国际总裁，现为三山公司的合伙人。

林平，山东大学本科毕业，在美国明尼苏达大学获得博士学位，现为香港岭南大学经济系教授。

马成虎，山东大学本科毕业，在加拿大多伦多大学获得博士学位，现为厦门大学王亚南经济研究院教授。

梅建平，复旦大学本科毕业，在美国普林斯顿大学获得博士学位，现为长江商学院教授。

倪晓光，清华大学本科毕业，在美国明尼苏达大学获得博士学位，现为密苏里大学经济系教授。

丘东晓，中山大学本科毕业，在加拿大英属哥伦比亚大学获得博士学位，现为香港大学经济及工商管理学院教授。

宋敏，浙江大学本科毕业，在华中科技大学获得硕士学位，在美国俄亥俄州立大学获得博士学位，现任教于香港大学经济及工商管理学院。

宋顺锋，北京大学本科毕业，在美国加州大学尔湾分校获得博士学位，现为美国内华达大学经济系教授。

谭国富，中国矿业大学本科毕业，在华中科技大学获得硕士学位，在美国加州理工大学获得博士学位，现为美国南加州大学经济系教授，并兼任上海财经大学工商管理学院院长。

陶志刚，复旦大学本科毕业，在美国普林斯顿大学获得博士学位，现为香港大学经济及工商管理学院教授。

王建新，清华大学本科毕业，在美国西北大学获得博士学位，现任教于澳大利亚南新威尔士大学商学院。

王汝渠，复旦大学本科毕业，在美国波士顿大学获得博士学位，现为加拿大皇后大学经济系讲座教授。

王苏生，南开大学本科毕业，在加拿大多伦多大学获得博士学位，现任教于香港科技大学经济系。

魏尚进，复旦大学本科毕业，在美国加州大学伯克利分校获得博士学位，现为美国哥伦比亚大学商学院讲座教授。

魏向东，中山大学本科毕业，在英国伯明翰大学获得博士学位，现为香港岭南大学经济系教授。

吴延瑞，安徽大学本科毕业，在南开大学获得硕士学位，在澳大利亚阿德莱德大学获得博士学位，现任教于西澳州大学商学院。

谢丹阳，武汉大学本科毕业，在美国芝加哥大学获得博士学位，现为香港科技大学经济系教授。

谢新，武汉大学本科毕业，在美国哥伦比亚大学获得博士学位，现从事私人基金投资工作。

许定波，武汉大学本科毕业，在美国明尼苏达大学获得博士学位，现为中欧国际工商学院教授。

许小年，中国人民大学硕士研究生毕业，在美国加州大学戴维斯分校获得博士学位，现为中欧国际工商学院教授。

徐永胜，兰州大学本科毕业，在美国图兰大学获得博士学位，现为佐治亚州立大学经济系教授。

查涛，成都工学院本科毕业，在美国明尼苏达大学获得博士学位，现为美联储亚特兰大分部高级研究员。

张处，华东师范大学本科毕业，在复旦大学获得硕士学位，在美国芝加哥大学获得博士学位，现任教于香港科技大学金融系。

张建波，南开大学本科毕业，在美国宾西法尼亚州立大学获得博士学位，现任教于堪萨斯大学经济系。

张俊森，浙江大学本科毕业，在南开大学获得硕士学位，在加拿大麦柯玛斯特大学获得博士学位，现为香港中文大学经济系讲座教授。

张俊喜，南开大学本科毕业，在美国匹兹堡大学获得博士学位，现任教于香港大学经济及工商管理学院。

赵海英，天津大学本科毕业，在美国马里兰大学获得博士学位，现任中国投资有限责任公司资产配置与战略研究部总监，汇金公司非银部主任。

郑旭，南开大学本科毕业，在美国普林斯顿大学获得博士学位，现任教于上海交通大学经济系。

周林，复旦大学本科毕业，在美国普林斯顿大学获得博士学位，现为美国亚利桑那州立大学经济系讲座教授，并兼任上海交通大学经济学院院长。

目录

下册

Econometrica, Vol. 65, No. 4 (July, 1997), 933–963

A SEMIPARAMETRIC MAXIMUM LIKELIHOOD ESTIMATOR

BY CHUNRONG AI[1]

This paper presents a procedure for analyzing a model in which the parameter vector has two parts: a finite-dimensional component θ and a nonparametric component λ. The procedure does not require parametric modeling of λ but assumes that the true density of the data satisfies an index restriction. The idea is to construct a parametric model passing through the true model and to estimate θ by setting the score for the parametric model to zero. The score is estimated nonparametrically and the estimator is shown to be $\sqrt{N}$ consistent and asymptotically normal. The estimator is then shown to attain the semiparametric efficiency bound characterized in Begun et al. (1983) for multivariate nonlinear regression, simultaneous equations, partially specified regression, index regression, censored regression, switching regression, and disequilibrium models in which the error densities are unknown.

KEYWORDS: Kernel, semiparametric, nonparametric, asymptotic efficiency.

1. INTRODUCTION

A COMMON METHOD FOR ANALYZING data $z=(x, y)$ is to presume that the conditional density of y given x is given by a known relationship $q(y \mid x, \theta_0, \lambda_0)$,[2] where $\theta_0 \in \Theta$ represents the finite dimensional parameters and $\lambda_0 \in \Lambda$ represents the infinite dimensional parameters. After specifying λ_0 parametrically, the maximum likelihood (hereafter ML) method is often applied. The ML estimator (hereafter MLE) for θ_0 is consistent if λ_0 is specified correctly, but it is generally inconsistent if λ_0 is misspecified. The main objective of this paper is to extend the ML method to the case where λ_0 is not modeled parametrically.

To accomplish this objective, we restrict our attention to a class of models for which the density of the data satisfies an index restriction. The density of the data satisfies an index restriction when there exists a variable transformation $(v_1(z, \theta), v_2(x, \theta))$ of known form such that

$$(1.1) \qquad q(y \mid x, \theta_0, \lambda_0)=J(z, \theta_0) f[v_1(z, \theta_0) \mid v_2(x, \theta_0), \theta_0],$$

where $f[.\mid ., \theta]$ is the conditional density of $v_1(z, \theta)$ given $v_2(x, \theta)$ for arbitrary θ, and $J(z, \theta)$ is the known Jacobian of the transformation from $v_1(z, \theta)$ to y. The proposed estimator is a natural extension to a ML setting of the LS-based semiparametric estimators proposed by Robinson (1988a) and Ichimura and Lee

[1] I thank my advisor Daniel McFadden for stimulating discussions and Jonathan Hamilton, Hidehiko Ichimura, Lung-Fei Lee, Whitney Newey, David Sappington, Thomas Stoker, and Jeffrey Wooldridge for helpful comments. I am grateful to a co-editor and three anonymous referees for suggestions that improved the results and the presentation. All remaining errors are mine.

[2] Throughout the paper, densities of continuous random variables are with respect to Lebesgue measure, densities of discrete random variables are with respect to probability measure, and densities of mixed discrete/continuous random variables are with respect to the product of the probability and Lebesgue measure.

(1991). Specifically, the basic idea of the paper is to consider the "parametric model"

$$(1.2) \qquad q(y \mid x, \theta, f) = J(z, \theta) f[v_1(z, \theta) v_2(x, \theta), \theta] \quad \text{for} \quad \theta \in \Theta$$

where the functional form of the conditional density $f[.\mid .,.]$ is fixed and assumed to be known. If this form were indeed known, then ML estimation of θ_0 is the appropriate method. Since $f[.\mid .,.]$ is unknown, we estimate $f[.\mid .,.]$ nonparametrically using observations on $(v_1(z, \theta), v_2(x, \theta))$.

Equation (1.1) is satisfied by many familiar econometric models. It is satisfied by a partially specified regression model $\lambda_1[y] = x_1'\theta_{10} + \lambda_2[x_2'\theta_{20} + x_3] + u$, for example, where $\lambda_1[.]$ and $\lambda_2[.]$ are functions of unknown form and u is independent of x (the union of x_1, x_2, and x_3) and has an unknown density $\lambda_3[u]$. This follows because, with $\theta_0 = (\theta_{10}', \theta_{20}')'$ and $\lambda_0 = (\lambda_1, \lambda_2, \lambda_3)$, the conditional density of y given x is

$$(1.3) \qquad q(y \mid x, \theta_0, \lambda_0) = \lambda_3[\lambda_1[y] - x_1'\theta_{10} - \lambda_2[x_2'\theta_{20} + x_3]] * \mid \lambda_1'[y] \mid,$$

which is also the conditional density of $v_1 = y$ given $v_2 = (x_1'\theta_{10}, x_2'\theta_{20} + x_3)$, where $\mid . \mid$ denotes the absolute value and $\lambda_1'[y]$ denotes the derivative of $\lambda_1[y]$ with respect to y. Ichimura and Lee (1991) considered $\lambda_1[y] = y$. In their case, (1.3) can be viewed as the conditional density of $v_1 = y - x_1'\theta_{10}$ given $v_2 = x_2'\theta_{20} + x_3$. Robinson (1988a) considered the case: $\lambda_1[y] = y$ and $\theta_{20} = 0$. For his model, (1.3) can be viewed as the conditional density of $v_1 = y - x_1'\theta_{10}$ given $v_2 = x_3$. Thus, both Ichimura and Lee's model and Robinson's model satisfy (1.1).

A second model satisfying (1.1) is the selection model: $d = 1\{x_3'\theta_{30} > u_3\}$ and $y = d(x_1'\theta_{10} + u_1) + (1 - d)(x_2'\theta_{20} + u_2)$, where u (the union of u_1, u_2, and u_3) is independent of x (the union of x_1, x_2, and x_3) with unknown density $\lambda_0[u]$ and $1\{.\}$ denotes the indicator function. To see why (1.1) is satisfied in this case, define

$$q_1(a, b, \lambda_0) = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{b} \lambda_0[a, u_2, u_3]\, du_3 \right) du_2 \quad \text{and}$$

$$q_2(a, b, \lambda_0) = \int_{-\infty}^{+\infty} \left(\int_{b}^{+\infty} \lambda_0[u_1, a, u_3]\, du_3 \right) du_1.$$

With $\theta_0 = (\theta_{10}', \theta_{20}')'$, the conditional density of (y, d) given x is

$$q(d, y \mid x, \theta_0, \lambda_0)$$
$$= dq_1(y - x_1'\theta_{10}, x_3'\theta_{30}, \lambda_0) + (1 - d) q_2(y - x_2'\theta_{20}, x_3'\theta_{30}, \lambda_0),$$

which is also the conditional density of $v_1 = (d, d(y - x_1'\theta_{10}) + (1 - d)(y - x_2'\theta_{20}))$ given $v_2 = x_3'\theta_{30}$. When d is not observed, the conditional density of y given x is

$$q(y \mid x, \theta_0, \lambda_0) = q_1(y - x_1'\theta_{10}, x_3'\theta_{30}, \lambda_0) + q_2(y - x_2'\theta_{20}, x_3'\theta_{30}, \lambda_0),$$

which is also the conditional density of $v_1 = y - x_1'\theta_{10}$ given $v_2 = (x_1'\theta_{10} - x_2'\theta_{20}, x_3'\theta_{30})$. Note that, under the restrictions $x_3'\theta_{30} = x_1'\theta_{10} - x_2'\theta_{20}$ and $u_3 = u_2 - u_1$, the model is a disequilibrium model. Thus, the disequilibrium model satisfies (1.1). Also, note that, when the dependent variable is generated according to $y = d(x_1'\theta_{10} + u_1)$, the model is a binary selection model. For this model, the conditional density of (y, d) given x is also the conditional density of $v_1 = (d, d(y - x_1'\theta_{10}))$ given $v_2 = x_3'\theta_{30}$.

A third model satisfying (1.1) is a simultaneous equations model $H(y, x, \theta_0) = u$, where u is independent of x and has an unknown density $\lambda_0[u]$ and where $H(.)$ is a function of known form. The conditional density of y given x is $J(y, x, \theta_0) * \lambda_0[H(y, x, \theta_0)]$, where $J(y, x, \theta_0)$ is the Jacobian of the transformation from $H(y, x, \theta_0)$ to y. Equation (1.1) is satisfied since $\lambda_0[.]$ can be viewed as the density of $v_1 = H(y, x, \theta_0)$.[3]

Other examples of models satisfying (1.1) include: (i) duration models; (ii) limited dependent variable models in which the structural component is partially specified and the distribution of the stochastic component is completely unknown; and (iii) all these and other models in which the disturbances depend on the regressors only through a parametric function of the regressors.

Let $z_i = (x_i, y_i)$ for $i = 1, \ldots, N$ denote sample observations and let $m(z, \theta, f) = \partial \ln[q(y \mid x, \theta, f)]/\partial\theta$. Denote the sample score for model (1.2) by $S_N(\theta) = \sum_{i=1}^{N} m(z_i, \theta, f)$. Then, if $f[.\mid.,.]$ were indeed known, standard arguments imply that the MLE, $\tilde{\theta}_N$, which solves $S_N(\tilde{\theta}_N) = 0$, is $\sqrt{N}$ consistent and asymptotically normal under regularity conditions (Hansen (1982)). Since $f[.\mid.,.]$ is unknown, $\tilde{\theta}_N$ is not feasible. We estimate $f[.\mid.,\theta]$ nonparametrically and then estimate $m(z, \theta, f)$ and $S_N(\theta)$. Let $\hat{S}_N(\theta)$ denote the estimate of $S_N(\theta)$. We show that the estimator, $\hat{\theta}_N$, which solves $\hat{S}_N(\hat{\theta}_N) = 0$, has the same asymptotic distribution as $\tilde{\theta}_N$. Thus, $\hat{\theta}_N$ is asymptotically as efficient as $\tilde{\theta}_N$.

Since $\hat{\theta}_N$ is as efficient as the (infeasible) MLE $\tilde{\theta}_N$, the asymptotic efficiency of $\hat{\theta}_N$ (in the sense of attaining the semiparametric efficiency bound characterized in Begun et al. (1983)) can be ascertained by checking the efficiency of $\tilde{\theta}_N$. We show that $\tilde{\theta}_N$ attains the semiparametric efficiency bound for the model

$$(1.4) \qquad \left\{q(y \mid x, \theta, \tilde{f}) \text{: for all } \theta \in \Theta \text{ and } \tilde{f} \in \mathscr{F}\right\},$$

where $\mathscr{F} = \{$all $f[v_1 \mid v_2, \theta]$ satisfying: (i) $f[v_1 \mid v_2, \theta] \geq 0$ and (ii) $\int f[v_1 \mid v_2, \theta]\,dv_1 = 1\}$.[4] However, model (1.4) is generally less restrictive than the model of interest:

$$(1.5) \qquad \{q(y \mid x, \theta, \lambda) \text{: for all } \theta \in \Theta \text{ and } \lambda \in \Lambda\}.^5$$

[3]Since multivariate nonlinear regression models are special cases of simultaneous equations models, (1.1) is also satisfied by nonlinear regression models.

[4]$\mathscr{F}$ is a set of conditional densities.

[5]Note that the forms of $q(y \mid x, \theta, f)$ and $q(y \mid x, \theta, \lambda)$ are not necessarily the same; $q(y \mid x, \theta, f)$ is given by (1.2).

Thus, $\tilde{\theta}_N$ generally does not attain the semiparametric efficiency bound for (1.5). Despite a lack of a general result, "parametric models" satisfying (1.1) are constructed for all of the classes of models described above, and the (infeasible) MLE $\tilde{\theta}_N$ based on these "parametric models" is shown to attain the semiparametric efficiency bound for (1.5).

It is interesting to note that

$$(1.6) \qquad J(z,\theta)f[v_1(z,\theta) \mid v_2(x,\theta),\theta_0]$$

is also a "parametric model" satisfying (1.1), where $f[.\mid.,\theta_0]$ is the conditional density of $v_1(z,\theta_0)$ given $v_2(x,\theta_0)$. Since $f[.\mid.,\theta_0]$ is unknown, ML estimation of model (1.6) is infeasible. However, suppose that one has available a $\sqrt{N}$ consistent estimator $\bar{\theta}_N$ of θ_0. Then one could form a nonparametric estimate $\bar{f}$ (say) of $f[.\mid.,\theta_0]$ using observations $(v_1(z_i,\bar{\theta}_N), v_2(x_i,\bar{\theta}_N))$. Using trimming and the s_N-approximation described below, ML estimation of model (1.6) can be performed and the MLE $\hat{\hat{\theta}}_N$ can be obtained. But, $\hat{\hat{\theta}}_N$ cannot improve upon the efficiency of $\hat{\theta}_N$ regardless of the efficiency property of the initial $\sqrt{N}$-consistent estimator $\bar{\theta}_N$. This follows because: (i) $\bar{f}$ is a consistent estimate of the true value of the nuisance parameter $\bar{f}(.\mid.)$ of (1.4) (thus $\hat{\hat{\theta}}_N$ is just a semiparametric estimator of (1.4)), and (ii) $\hat{\theta}_N$ attains the semiparametric efficiency bound for (1.4).

The econometric literature on semiparametric estimation is large. The papers most closely related to our approach are those by Gallant and Nychka (1987), Severini and Wong (1991), Newey (1990), Klein and Spady (1993), and Lee (1989, 1990). Unlike the approach here, the first three sets of authors replace λ_0 with a sieve and estimate θ_0 together with the sieve parameters by either ML estimation or GMM. None of these studies, however, has given explicit conditions on the sieve so that their estimators are $\sqrt{N}$ consistent. The last two sets of authors, on the other hand, construct "parametric models" to satisfy (1.1) for discrete response models. Our approach is an extension of theirs to more general semiparametric models.

Other related studies include Bickel (1982), Manski (1984), Newey (1989), Robinson (1988a, 1991) and Ichimura and Lee (1991). While Robinson (1991) and Newey (1989) proposed best estimators for simultaneous equations models, Bickel and Manski studied adaptive estimation of these models. For a partially specified linear regression model, Robinson (1988a) proposed an estimator that attains Chamberlain's (1992) bound under homoskedastic errors, whereas Ichimura and Lee suggested a least squares procedure for an index regression model. Our estimation procedure can lead to efficient estimators for all these models under independent errors and, thereby, extends these studies.

The paper is organized as follows. Section 2 describes a nonparametric estimation scheme. Section 3 derives asymptotic results for the proposed estimator. Section 4 constructs efficient estimators for all of the example models above, and Section 5 concludes the paper. Technical proofs are relegated to the Appendix.

2. NONPARAMETRIC ESTIMATION

We first describe a scheme for estimating $S_N(\theta)$. Decompose $f[.\mid ., \theta]$ according to

$$f[v_1(z,\theta)\mid v_2(x,\theta),\theta]=f_1[v(z,\theta),\theta]/f_2[v_2(x,\theta),\theta],$$

where $f_1[.,\theta]$ denotes the joint density of $v(z,\theta)=(v_1(z,\theta), v_2(x,\theta))$ and $f_2[.,\theta]$ denotes the marginal density of $v_2(x,\theta)$ for arbitrary θ. Substituting the expression for $f[.\mid ., \theta]$ gives $m(z,\theta,f)=m_1(z,\theta,f_1,f_2)m_2(z,\theta,f_1,f_2)$, where

$$\begin{aligned} m_1(z,\theta,f_1,f_2) = {} & \frac{d\ln[J(z,\theta)]}{d\theta} f_1[v(z,\theta),\theta] f_2[v_2(x,\theta),\theta] \\ & + \frac{df_1[v(z,\theta),\theta]}{d\theta} f_2[v_2(x,\theta),\theta] \\ & - f_1[v(z,\theta),\theta]\frac{df_2[v_2(x,\theta),\theta]}{d\theta}, \end{aligned}$$

$m_2(z,\theta,f_1,f_2)=1/(f_1[v(z,\theta),\theta]f_2[v_2(x,\theta),\theta])$, and $d/d\theta$ denotes the total derivative with respect to θ. Next, we estimate $f[.\mid ., \theta]$ by estimating $f_1[.,\theta]$ and $f_2[.,\theta]$. We estimate $f_1[.,\theta]$ and $f_2[.,\theta]$ by the kernel method described in Silverman (1986).

It is most convenient to assume that the elements of $v(z,\theta)$ are continuous, though kernel methods can handle the case of both discrete and continuous elements.

ASSUMPTION 1: *$v^k(z,\theta)$, the kth element of $v(z,\theta)$, is a continuous random variable for each k and θ.*[6]

Let m and m_2 denote the dimension of $v(z,\theta)$ and $v_2(x,\theta)$, respectively. Let $K_1[v]$ and $K_2[v_2]$ be known symmetric functions and let h_1 and h_2 be bandwidths. Then, kernel estimators for $f_1[v,\theta]$ and $f_2[v_2,\theta]$ are

$$\hat{f}_1[v,\theta]=\frac{1}{Nh_1^m}\sum_{j=1}^{N}K_1\left[\frac{v-v(z_j,\theta)}{h_1}\right] \quad \text{and}$$

$$\hat{f}_2[v_2,\theta]=\frac{1}{Nh_2^{m_2}}\sum_{j=1}^{N}K_2\left[\frac{v_2-v_2(x_j,\theta)}{h_2}\right].$$

Let $\hat{f}_1$ and $\hat{f}_2$ denote $\hat{f}_1[.,\theta]$ and $\hat{f}_2[.,\theta]$ respectively. Then, $m_1(z,\theta,\hat{f}_1,\hat{f}_2)$ is an estimate of $m_1(z,\theta,f_1,f_2)$.

Note that $m_2(z,\theta,f_1,f_2)$ is unbounded when $f_1[.,.]$ is not bounded away from zero.[7] When this happens, $m_2(z,\theta,\hat{f}_1,\hat{f}_2)$ does not converge to

[6] This assumption is not essential for the results below. Extensions to the case of both discrete and continuous elements are straightforward. See the discussion following Theorem 1.

[7] For instance, in the simultaneous equations example, the density of $H(z,\theta)$ is not bounded away from zero because the support of $H(z,\theta)$ is $(-\infty,+\infty)$.