Non-parametric Estimation Methods for Instrumental Variables and Sample Selection: Theory and Applications

by

Mitali Das

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 1998
[June 1998]

Author.....
Department of Economics
May 08, 1998

Certified by.....
Whitney K. Newey
Professor of Economics
Thesis Supervisor

Certified by.....
Jerry A. Hausman
Professor of Economics
Thesis Supervisor

Accepted by.....
Peter Temin
Chairman, Department Committee on Graduate Students

Non-parametric Estimation Methods for Instrumental Variables and Sample Selection: Theory and Applications

by

Mitali Das

Submitted to the Department of Economics
on May 08, 1998, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This dissertation consists of three chapters that combine new theoretical results on the convergence, consistency and asymptotic normality of non-parametric estimators with applications in health economics. The first chapter is devoted exclusively to the econometric theory underlying non-parametric instrumental variables estimation. Chapter II (joint with Whitney Newey), presents the theorems relating to non-parametric estimation of the sample selection model along with an application that studies the effects of restrictions in abortion care on the birthweight of babies. In Chapter III, the focus in on implementing the nonparametric instrumental variables estimators in a study which documents how the growth in Health maintenance Organizations (HMOs) over the last decade has inversely affected the county cesarean-section rate.

I begin with Chapter I which studies non-parametric estimation of the model

$$y = m(x, d) + \epsilon$$

where m is unknown and is the object of estimation; the variables d are assumed to be correlated with the error term ϵ, and d are either dummy variables or discrete variables with finite support; x is uncorrelated with ϵ. I present the non-parametric instrumental variables estimators for such a model, and discuss how the discrete nature of d is critical in deriving the identification condition. The identification condition in turn suggests the natural estimators of the model. The key result underlying each of the proposed estimators is a simple transformation which allows us to parametrize the endogenous component of a non-parametric model. This tranform is also shown to be useful in other non-parametric models (without endogenous regressors) where it is typically assumed that the regressors are continuously distributed. Allowing for discrete variables requires substantial modification in the asymptotic theory unless the suggeted transform is used. Finally, it is shown how an important application of the proposed estimators is to the non-parametric estimation of flexible panel data models. The current set of estimators for these models relies on an analogue of the differencing technique proposed for the linear model which also differences out the time-invariant regressors. Non-parametric IV is suggested as the natural fix to this problem, yielding the non-parametric analogue to Hausman and Taylor (1981).

In Chapter II we consider non-parametric estimation of the sample selection model of Heckman (1974), by allowing for a non-parametric selection rule as well as a non-parametric primary regression model. The estimator we develop significantly generalizes existing sample selection estimators which are all *semi-parametric* since some part of their model is parametrically specified.

We impose no assumptions on the regression functions, and leave the joint distribution of the error terms unspecified. The only restriction we impose is the additivity of the error terms in each step; a weak restriction. Using power series or polynomial spline series approximations, we show that interesting functionals of our estimator are $\sqrt{n}$-consistent, and asymptotically normal. We present an application which studies the effect of medical, prenatal and abortion services available to pregnant women on the distribution of the birthweights of their newborns (a selected outcome), also reporting corresponding estimates from parametrically specified and semi-parametric models to facilitate comparison between the older and newer methods. We show that the proposed estimator is easily computable, yields non-trivial differences from parametric models, and obviates any incorrect parametric assumptions that are a potential drawback of existing estimators.

Chapter III takes on a more applied tone, extending the body of research on the demand-inducing behaviour of physicians. Apart from the ethical issues that arise, induced demand for those medical services which produce the same outcome as less costly substitues is an important policy issue in the ongoing national debate on health care reform. I identify one such medical category - - that of the delivery of babies - - and illustrate through an induced-demand model, and empirical evidence, how increased managed care activity results in altering the compositio n of deliveries from the more highly-reimbursed cesarean-section towards natural childbirths. Using a variety of linear, non-linear and non-parametric specifications, I show that while demand inducement plays a role in the c-section rate, its effect is swamped out by the contemporaneous increase in managed care activity that effectively induces demand for naturals. Further, this compositional change does not come with associated changes in the "quality " of births (for *e.g,* the neo-natal infant mortality rate, complications at birth, Apgar scores). The results obtained in this paper suggest that on average, a 10% increase in HMO penetration rates results in a 1.8% decrease in cesarean-section rates, translating into net cost savings of $11 billion at 1992 cesarean rates and cesarean reimbursement schedules. The natural implication for policy design is that policies which favor the growth of managed care will result in first-order cost savings from this medical service.

Thesis Supervisor: Whitney K. Newey
Title: Professor of Economics

Thesis Supervisor: Jerry A. Hausman
Title: Professor of Economics

Acknowledgments

Finally Done! I need to thank and acknowledge several people without whose help writing this tome in finite time would have been near impossible. My biggest debt of gratitude goes to Whitney Newey and Jerry Hausman, unarguably the best advisors I could have ever hoped for, who helped me every step of the way, making thesis-dom less daunting. With much love, admiration and respect, I dedicate this thesis to them.

Whitney should be bestowed with sainthood for heroically putting up with everything from my never-ending worries about finishing the thesis and my paranoia of being unemployed, to my insecurities and childish complaints. He provided simple advice, plenty of pep and made The Call. His wizardry in econometric theory is unmatched and I never had to feel embarrassed for asking inane questions that he answered over and over (and over) again, for habitually showing up to his office 5 days a week or for the occasional teary meltdown. He has been advisor, friend, mentor and parent, and in the last anxious days of the job market, I felt tremendously fortunate that there was someone who had more faith in me than I did myself. Since I dont feel deserving to write a paper with him (Chapter II), I also thank him for putting me on the Newey welfare rolls.

I was an unabashed fan of Jerry's long before I came to MIT and continued to shamelessly idolize him through my years here. I don't believe I have ever met a person more generous with his knowledge, time, advice, or Chicago Bulls tickets; this generosity perhaps dwarfed only by Jerry's magnanimous gestures in funding D-league hockey teams, dinners for starving students and an occasional job market suit. In dire days, he even promised to find someone gullible enough to marry me. I have learnt more from Jerry than I have from anyone else at MIT; partly from classes, partly from our daily interactions and partly from the summers as his RA (where, I should mention, I made oodles of money). Jerry's knowledge of economics is unparalleled - - a statement he will cheerfully attest to, - - and his blunt advice, though very blunt at times, was timely and in fact, kind. I'll miss having him around and wait till he outbids to become Columbia's biggest catch yet.

My fellow grad-students - - truly MIT's best asset - - should be singled out for making life outside of MIT wonderful and for keeping me sane. Though dissertating (to coin one) was frustrating, humbling and discouraging at many times, we worked it out together to keep the atmosphere upbeat and friendly. In our third year, through magic and sorcery, all my friends ended up in Rm

201 and were instrumental in the shaping of Chapter I, especially James Ortega, Stuh Bergantino, Larry Muecke, Al Jousten, Chris Frois, Andy Repetto, Ken Yudaeva and Annetah. My very special thanks to Annette Vissing-Jørgensen who, while the rest of the sane world slept, worked the night shifts with me in year 4. Night in and night out she was the indispensable companion, semi-econometrician and friend who provided the much needed solace and humour to make the nights shorter. A fraction λ of this thesis is dedicated to her.

My office and the computer room became a hub for gossip that I enjoyed. In particular, I cannot thank Scott enough for smooth-talking me into ploughing along in Hilbert space when it looked particularly bleak, or for the enormous amounts of time he has spent helping me since he demanded pizza at 2 a.m. when he was being recruited. Jon Levin, Jon Guryan, Esther Duflo, Guido Kuersteiner, Jen Wu, Jeff Furman, John Kim, Paul Ellickson, Alberto Abadie, Peter Sørensen, Max, Fernando, Andrew Percey, Pablo, Mark Aguiar and Lucia Breirova provided memorable diversions when I was affixed to balder.mit.edu. Samiya and Osama were the best sources of negative productivity shocks for me, and made sure I got a ton of great meals while they were here.

I owe a measure of gratitude to Guido Kuersteiner who made the last year so memorable for me. He made it a point to see that I didn't develop tunnel vision writing the thesis, and provided sufficient diversions watching lame movies, immensely hurried dinners at the Mall, and chauffering me from here to there. His humorous (though sincere) efforts in trying to convince me that time-series was the way to go provided much laughter; and his friendship, support and advice during the harrowing job market days was therapeutic.

Finally, I thank my family for standing behind me in all my endeavours. My parents are in shock and disbelief that someone in the Das dynasty actually managed to earn a PhD! I owe them a lot more than I can begin to explain here. Never mind that they are still unsure of how to pronounce e-con-o-met-tricks, they have always been supportive of their youngest spoilt child and always believed that if I put my mind to it, I could get it done. I never outgrew hero-worshipping my father who taught me about perseverance and integrity through some of the darkest days of his own professional life, and always encouraged me to pursue everything I set my heart on - - Thank you.

Contents

Chapter 1

Non-parametric Estimation Methods for Instrumental Variables

1.1 Introduction

The recent surge in the semi-parametric and non-parametric approach to econometric estimation has extended the standard methods in several ways [Manski (1985, 1987), Han (1988), Newey (1990, 1994a), Andrews (1995) Powell (1987), Honoré (1992)]. One area which has received very little attention in this growing body of work is that of instrumental variables (I.V) estimation. It is useful to consider non-parametric instrumental variables estimation since the analysis of structural models arises very frequently in actual empirical work; relaxing the restrictions of parametric models may aid in avoiding the inconsistencies that arise from incorrect modelling assumptions. This paper presents new theory for the estimation of non-parametric structural models. Two sets of estimators are developed, and the corresponding large sample results are presented.

The research in non-parametric instrumental variables (NPIV) estimation is thus far restricted to Roehrig (1988), Newey and Powell (1989) and Newey, Powell and Vella (1997). Roehrig (1988) presented general conditions under which non-parametric systems of equations are identified. Newey and Powell (1988) weaken the specification of the stochastic component of Roehrig's model and consider the estimation of a system of equations, demonstrating that a non-parametric two-stage least squares estimator is ill-defined asymptotically; there is no distribution theory. In Newey, Powell and Vella (1997), the analysis avoids studying a system of equations, instead considering the identification and estimation of a non-parametric triangular simultaneous equations model.

The triangularity assumption, while restrictive, is an important feature of deriving their two-step NPIV estimator.

In this paper, I demonstrate that an analysis of non-parametric models in which the endogeneity stems from dummy (or more generally, discrete) endogenous variables provides additional structure that is extremely useful in deriving NPIV estimators.[1] The results presented have two main implications. First, it is shown that the "non-existence" result presented by Newey and Powell (1989) arises only if the endogeneity is associated strictly with continuously distributed variables. With dummy endogenous variables, the NPIV estimator is asymptotically well-defined for both, systems of equations as well as triangular specifications of simultaneous equations models. This result is encouraging, since most of the available microeconometrics data take on a discrete or dummy variable form. Second, much of the available theory for non-parametric models (without endogenous variables) relies on the assumption that the regressors are continuously distributed. A second result of this paper is that in models with discrete variables, a reformulation of the problem by exploiting the discrete nature of the data allows us to relax this assumption and expand the applicability of the existing estimators.

The essential result of this paper is that in any non-parametric model with dummy variables, a simple transform allows us to parametrize the noncontinuous component. This is always possible since the dummy variable may simply be 'pulled out' of the unknown function, which can then be evaluated at each of the two support points. It is shown that for models in which the endogeneity is derived strictly from dummy variables, this transform is critical in deriving the identification condition. The transform is as follows (for d denoting the vector of endogenous dummy variables and x a vector of exogenous continuous variables)

$$\begin{aligned} y &= m(x,d) + \epsilon \\ &\equiv \tilde{m}(x)[d] + \tilde{\tilde{m}}(x)[1-d] \\ &= \tilde{\tilde{m}}(x) + [\tilde{m}(x) - \tilde{\tilde{m}}(x)]d \\ &= \tilde{\tilde{m}}(x) + \breve{m}(x)d \end{aligned} \tag{1.1.1}$$

where $m(\cdot)$, $\tilde{m}(\cdot)$, $\tilde{\tilde{m}}(\cdot)$ and $\breve{m}(\cdot)$ are unknown and unrestricted functions and ϵ is the disturbance.

[1] For the rest of the paper, we will simply focus on the dummy variable case to simplify the exposition. All of the results extend simply to the discrete variable with finite support case.

This transformation always holds; therefore, even though the results below focus on the case where d is endogenous, the results may be equally applied to any non-parametric model with dummy variables. The latter exercise is useful, since much of the literature on nonparametric estimation is motivated by assuming continuously distributed variables. Allowing for noncontinuously distributed variables requires some modification in the asymptotic theory.

For most of the paper we will only consider the case where the endogeneity arises strictly from the dummy variables. One interesting hybrid case is treated in Section 4. The model we will focus on is a general simultaneous equation system where the conditional mean restrictions $E[\epsilon|x] = 0$ and $E[\nu|x] = 0$ hold:

$$\begin{aligned} y &= m(x_1, d) + \epsilon \\ d &= f(x_2, y) + \nu \end{aligned} \tag{1.1.2}$$

where $m(\cdot)$ or $f(\cdot)$ is the object of estimation, $x = [x_1\ x_2]$ is a vector of exogenous variables and $z = [z_1, x_1]$ is a vector of available instruments such that $dim(x_1 + z_1) \geq dim(y)$ and $dim(x_2 + x_1) \geq dim(d)$. In section II, we begin by showing how this order condition would be insufficient to identify the model if d was continuously distributed

I present the NPIV estimator for this model. The construction of the estimator resembles that of parametric two-stage least squares, obtaining an estimate of the endogenous dummy variable in the first step and using this estimate as an instrument in step two. The estimator presented here is based on series estimation, with a focus on power series. Other results include the mean-square convergence rate and asymptotic normality of the proposed estimator.

Two modifications of the model are also considered. First, we extend the analysis to that of discrete endogenous variables and present the corresponding NPIV estimator. Next, we consider a triangular model and present an estimator complementing that of Newey *et al* (1997), whose estimator is derived using continuity of the regressors as an assumption. As with parametric linear IV, the NPIV estimators here are equally applicable to endogeneity arising from equilibrium conditions, measurement error, or unobservable individual heterogeneity. In Section *IV* we show how the current set of estimators for non-parametric panel data models rely on the first-differencing method used for linear panel models, and result in differencing out the time-invariant component. NPIV estimation is proposed as a solution to this problem, providing the non-parametic analog to

the method proposed by Hausman and Taylor (1981).

In Section II we begin by defining the problems associated with non-parametric IV estimation of simultaneous equations models, and present the identification conditions for (1.2). Section III presents the NPIV estimator, along with convergence rates and asymptotic normality. In Section IV, we study two simple extensions of the model, and in Section V, an application of the estimator to the panel problem is presented. Section VI concludes.

1.2 Identification

To demonstrate why the discrete nature of d is critical in deriving the identification condition of (1.2), begin by assuming d is continuously distributed. Denote w for the set of all exogenous variables, *i.e*, $w = [x, z]$. In the analysis below, we will focus on m as the object of estimation. Obviously, an isomorphic exercise can be carried out to estimate $f(x_2, y)$. With the assumed conditional mean restrictions on the error terms, it follows that

$$E[\,y|w\,] = E[\,m(x_1, d)|w\,] = \int m(x_1, d) f(x_1, d|w) dw \tag{1.2.1}$$

where $f(\cdot|w)$ is a conditional density. This equation relates the reduced form of the model to the structural form. Since both $E[y|w]$ and $f(\cdot|w)$ can be estimated from the data, they can be treated as identified components of the equation, leaving $m(\cdot)$ as the only unknown function in equation (2.1).

In the form given above, for known $E[y|w]$ and $f(\cdot|w)$, and for unknown $m(\cdot)$, equation (2.1) represents an *integral equation of the first kind*, for which it is well-established that a continuous map from $E[y|w]$ to m need not exist [Newey and Powell (1988), Wahba (1979), Hárdle and Linton (1991)]. This problem arises precisely because the differential operator (*i.e*, the inverse operator from m to $E[y|w]$) is typically discontinuous. The implication is that for very small changes in $E[y|w]$, the solution to the integral equation may fluctuate widely. Since a primary condition for identification is a one-to-one and continuous map from the structure to the reduced form, m is unidentified in (2.1). Note that this problematic feature arises exactly because $m(\cdot)$ is unknown.

It now follows quite simply how a special case of the non-parametric model in (2.1) with the

endogenous variables (d) are dummies can be identified:[2]

$$\begin{aligned} E[y|w] = E[m(x_1,d)|w] &= \int m(x_1,d)f(x_1,d|w)dw \\ &= \int [\bar{\tilde{m}}(x_1) + \check{m}(x_1)d]f(x_1,d|w)dw \\ &= \int \bar{\tilde{m}}(x_1)f(x_1,d|w)dw + \int \check{m}(x_1)d\ f(x_1,d|w)dw \\ &\equiv \bar{\tilde{m}}(x_1) + \check{m}(x_1)\,E[d|w] \end{aligned} \tag{1.2.2}$$

thus avoiding the discontinuity of the differential operator that arises with the integral equation of the first kind.

Each of the two additive components can now be identified in the following way. Without loss of generality, assume the instrumental variables z are discrete, $\in \{0,1\}$ Then,

$$\begin{aligned} E[y|x,z=0] &= \bar{\tilde{m}}(x_1) + \check{m}(x_1)E[d|x,z=0] \\ E[y|x,z=1] &= \bar{\tilde{m}}(x_1) + \check{m}(x_1)E[d|x,z=1] \end{aligned} \tag{1.2.3}$$

and,

$$\frac{E[y|x,z=0] - E[y|x,z=1]}{E[d|x,z=0] - E[d|x,z=1]} = \check{m}(x_1) \tag{1.2.4}$$

The formula in (2.4) is an explicit expression for the additive component $\check{m}(x_1)$ in terms of the identified components $E[d|x,z=i]$ and $E[y|x,z=i]$, $i=0,1$. The key condition underlying the identification of $\check{m}(x_1)$ is that given x, the conditional expectation of d must vary non-trivially for different realizations of z, *i.e*, there be sufficient variation in the conditional mean of d:

$$Var[\,E(d|x,z)\,|\,x] > 0 \tag{1.2.5}$$

Below we show how exactly this condition is necessary to bound the relevant second moment matrix away from singularity so that once identification of the model is assumed, the proposed instrumental variables estimator will be well-defined.

The Wald representation in equation (2.4) underlying the identification of $\check{m}(x_1)$ suggests that

[2]It can similarly be shown how parametric simultaneous equations models avoid the non-identification problem shown above.

kernel estimation of the various components in the numerator and denominator is a straightforward method to isolate $\breve{m}$. In particular applications $\breve{m}$ may be of interest in inference, as it represents the additional effect on the conditional mean of y for the subset of observations for whom the dummy takes on the value 1. Examples of such applications abound in the program evlaluation literature where we might be interested, for example, in the effects on the wage effects of participation in the AFDC program, or the effects of being drafted to serve in war.

For a scalar u let $\kappa(u)$ denote a kernel function such that $K_h(u) = \frac{1}{\sigma}(\kappa(u/\sigma))$ and $\int \kappa(u)du = 1$, where σ denotes the bandwidth parameter. Next, define the multivariate kernel function, $K(u_1, \ldots, u_d) = \Pi_{j=1}^{d} \kappa(u_j)$. Let $i = 1, \ldots n$ denote the data point, and for d_{x_1} denoting $\dim(x_1)$ and d_z denoting $\dim(z)$, define the weights:

$$W_\sigma(\bar{x_1}, \bar{z}) = \frac{\Pi_{j=1}^{d_{x_1}} K_\sigma(\bar{x}_{1j} - x_{1ij}) \Pi_{k=1}^{d_z} K_\sigma(\bar{z}_k - z_{2ik})}{\hat{f}(\bar{x}_1, \bar{z})}$$

$$\hat{f}(x_1, \bar{z}) = \sum_{i=1}^{n} K_\sigma(\bar{x}_1, \bar{z}) \tag{1.2.6}$$

Then we have the Nadaraya (1964) and Watson (1965) kernel smoother,

$$\begin{aligned} \hat{E}(y_i|\bar{x}_1, \bar{z}) &= \sum_{i=1}^{n} W_\sigma(\bar{x}_1, \bar{z}) y_i \\ \hat{E}(d_i|\bar{x}_1, \bar{z}) &= \sum_{i=1}^{n} W_\sigma(\bar{x}_1, \bar{z}) d_i \end{aligned} \tag{1.2.7}$$

which we can evaluate at $\bar{z} = 0$ and $\bar{z} = 1$ and any value of x_1 to get the various components in equation (2.4). An estimate of $\breve{m}(x_1)$ can now be obtained by substituting each of these in equation (2.4).

1.2.1 Proposed Estimator

The estimation method proposed here is a new result on instrumental variables estimation. It is the non-parametric correspondence of the standard I.V. estimator, and does not appear to have been explored previously. New results are also derived for the convergence rate, and for asymptotic

normality. As noted earlier, while this correspondence is valid for dummy endogenous variables, it does not hold for a general model where the endogenous variables are continuously distributed, due to a basic singularity of the second moment matrix problem discussed in the previous section.

The model we have is

$$\begin{aligned} y &= m(x,d) + \epsilon \\ d &= f(x_2, y) + \nu \end{aligned} \tag{1.2.1.1}$$

Let $w = [x, z]$. Then, the reduced form for the equation that we focus on is :

$$E[y|w] = \tilde{\tilde{m}}(x_1) + \breve{m}(x_1)\, E[d|w] \tag{1.2.1.2}$$

As typically done in series estimation, we begin by assuming that $m(x,d)$ can be approximated by a linear combination of $K(n)$ functions of $[x\, d]$, denoted $p^K(x,d)$, such that $|m(x,d) - p^K(x,d)'\overline{\beta}|$ is of an order that shrinks as $K(n)$ gets large. The key problem is that the approximation of $m(x,d)$ by $p^K\hat{\beta}$, with $\hat{\beta} = (P'P)^- P'\tilde{y}$, and $P = [P_1(x_1,d), \ldots, P_n(x_1,d)]$ is inconsistent for $m(x,d)$ just as least squares would be for the linear model. For the linear model we would, instead, choose a set of instruments correlated with $[x_1, d]$ and uncorrelated with ϵ such as $[x_1, z]$, and obtain a matrix of coefficients by computing $[(x_1', z')(x_1, d)']^{-1}$ and $(x_1', z')'y$. Consider the extension of this basic idea to the model in (2.1.2).

First note, that for the transformed model in (2.2.2), the series approximation will be additive, consisting of a set of functions approximating $\tilde{\tilde{m}}(x_1)$, say $P(x_1)$ and a set of functions to approximate $\breve{m}(x_1) * d$ such as $P(x_1) * d$. Let $\Upsilon(w) = E(d|w)$ denote the optimal instrument, $i = 1, \ldots, n$ denote the observations, and $A(x_1, \Upsilon) = [P(x_1)\ P(x_1)\Upsilon]$ denote the matrix of functions of the

instruments with $P(x_1) = \begin{bmatrix} p^K(x_{11})' \\ \vdots \\ p^K(x_{1n})' \end{bmatrix}$ $P(x_1, \Upsilon) = \begin{bmatrix} p^K(x_{11})'\Upsilon_1 \\ \vdots \\ p^K(x_{1n})'\Upsilon_n \end{bmatrix}$ where the $p^K(\cdot)$ represent

a $K \times 1$ vector of functions and, let $A(x_1, d) = [P(x_1)\, P(x_1)d]$ denote a matrix of of functions

of the original set of variables with $P_d = \begin{bmatrix} p^K(x_{11})'d_1 \\ \vdots \\ p^K(x_{1n})'d_n \end{bmatrix}$ such that $A(x_1, \Upsilon)$ and $A(x_1, d)$ are conformable for matrix multiplication. Then, our non-parametric instrumental variables estimator of $m(x_1, d)$ is

$$\begin{aligned} \hat{m}(x_1, d) &= \hat{\beta}_0 p^K(x_1) + \hat{\beta}_1 p^K(x_1) d \\ \hat{\beta}^{IV} &= [\hat{\beta}_0' \ , \hat{\beta}_1'], \\ \hat{\beta} &= [\ A(x_1, \Upsilon) A(x_1, d)\ ]^- \ A(x_1, \Upsilon)\ 'y \end{aligned} \tag{1.2.1.3}$$

The I.V. estimator above is well-defined only if the matrix of second moments, $[\ A(x_1, \Upsilon) A(x_1, d)\ ]$, is invertible. Thus, it is worth analyzing the technical details that allows us to bound this matrix away from singularity; this analysis is also useful as a comparison against the I.V. estimator for an alternate model with continuously distributed random variables where the second moment matrix will *necessarily* get arbitrarily close to singularity and is thus asymptotically ill-defined. For this case, we find that if the model is identifed (by *assumption*, in equation (2.5)), the second moment matrix will be bounded away uniformly from zero.

Let $\Sigma_{A(\Upsilon)A(d)} = E[A(x_1, \Upsilon)_i' A(x_1, d)_i]$. Retaining the notation that $w = [x, z]$, we evaluate $\Sigma_{A(\Upsilon)A(d)} = E[A(x_1, \Upsilon)_i' A(x_1, d)_i]$

$$= \begin{bmatrix} E[p^K(x_1)p^K(x_1)'] & E[p^K(x_1)p^K(x_1)E[d|w]] \\ \\ E[p^K(x_1)p^K(x_1)'\Upsilon(w)] & E[p^K(x_1)p^K(x_1)'\Upsilon(w)E[d|w]] \end{bmatrix} \tag{1.2.1.4}$$

$$= E\left[\begin{bmatrix} 1 & \Upsilon(w) \\ \\ \Upsilon(w) & \Upsilon^2(w) \end{bmatrix} \otimes p^K(x_1)p^K(x_1)'\right] \tag{1.2.1.5}$$

We can restrict our focus to $\Upsilon = \begin{bmatrix} 1 & \Upsilon \\ \Upsilon & \Upsilon^2 \end{bmatrix}$; if we can establish that the smallest eigen-value $\lambda_{min}(\Upsilon)$ is bounded away from zero, the enitre matrix will have full column rank since standard conditions on $p^K(x_1)$ will be sufficient to bound $E[p^K(x_1)p^K(x_1)']$ from singularity.

We have that $|E\Upsilon| = \mathrm{E}_{x_1}[(w)^2)] - E_{x_1}[(w)]^2$, which is the conditional variance of $E(d|w)$. By (2.5), necessary for identification, there exists some δ, $0 < \delta < \infty$, such that $Var(E(d|w)|x) \geq \delta$. The additional step now is to show that $\lambda_{min}(\Upsilon) > 0$: denote $\lambda_1(w)$, $\lambda_2(w)$ for the vectors of eigen-values of Υ. Then, since the trace of Υ is the sum of its eigen-values, we have $1 + \Upsilon^2(w) = \lambda_1(w) + \lambda_2(w) > 0$. To rule out $\lambda_{min} = 0$, note that $\lambda_1 + \lambda_2 > 0$ and $|E\Upsilon| > 0$ imply that $\lambda_1\lambda_2 > 0$. Thus, $\lambda_1, \lambda_2 > 0$ and λ_{min} is bounded away from zero. Then,

$$\left| E \begin{bmatrix} 1 & \Upsilon(w) \\ \Upsilon(w) & \Upsilon^2(w) \end{bmatrix} \otimes p^K(x_1)p^K(x_1)' \right| \geq \lambda_{min} I \otimes E[p^K(x_1)p^K(x_1)'] \qquad (1.2.1.6)$$

which is bounded away from singularity under standard assumptions on $E[p^K(x_1)p^K(x_1)']$. So, the proposed I.V. estimator is well-defined asymptotically.

Bounding λ_{min} away from zero using the approach above is precisely the condition that breaks down in establishing the boundedness of the second moment matrix in I.V. estimation with continuously distributed random variables. Above, as in the standard parametric model, assuming that the model is identified is equivalent to assuming invertibility of the matrix of second moments, and ensures a one-to-one continuous map from the structure to the reduced form; this one-to-one map, as shown in Newey and Powell (1988) cannot hold in the general model where assuming identification is not equivalent to assuming a continous map from the structural to the reduced form. Thus, discrete endogeneous variables are special, and appear to be the only case for which the standard I.V. estimator has a non-parametric correspondence.

1.3 Estimation

The implementation of the instrumental variables estimator is done in two steps. Begin with the first step in which the optimal instrument $\Upsilon(w) = E(d|w)$ is computed. Least squares regression of

the dummy variable d on an $(L \times 1)$ sequence of functions, $r^L(w)$ yields the instrument $\hat{\Upsilon} = r^L(w)'\hat{\gamma}$ where $\hat{\gamma} = (R'R)^- R'd$, $R = (r^L(w_1), \ldots, r^L(w_n))$. Since the estimated object is a probability, we use a trimming function to bound it between 0 and 1. Note that the above method is exactly analogous to the linear IV method, and not least squares using the predicted values (or, "repeated least squares"). This is simply because projecting non-linear functions of d onto the subspace spanned by w is not equivalent to taking non-linear functions of a linear term projected onto w.

Step two uses the optimal instrument in constructing a set of functions in which each element of $A(x_1, d)$ is replaced by terms consisting of $\hat{\Upsilon}$. The estimation is no longer done by least squares; instead, the two components of the estimator are computed and combined to yield $\hat{\beta}^{IV}$ as described earlier:

$$\hat{m}(x_1, d) = \hat{\beta}_0 p^K(x_1) + \hat{\beta}_1 p^K(x_1) d$$

$$\tau[x, d] = \Pi_{j=1}^{2*d_{x_1}} \quad 1 \quad (a_j \leq x_{1j} \leq b_j) 1(0 \leq d_j \leq 1)$$

$$\hat{\beta}^{IV} = \left\{ \sum_{i=1}^{N} \tau[x_{1i}, \hat{d}_i] \; [\; A(x_{1i}, \Upsilon_i) A(x_{1i}, d_i) \;]' \right\}^{-1} \cdot \left\{ \sum_{i=1}^{N} \tau[x_{1i}, \hat{d}_i] \; A(x_{1i}, \Upsilon_i)' y_i \right\}$$

where a_j and b_j are either pre-specified bounds or are themselves estimated.

1.3.1 Convergence Rates and Asymptotic Normality

The mean square error convergence rates for the IV estimator of $m(x_1, d)$ in simultaneous equations models with dummy endogenous variables is presented below. The single contrasting feature of the mean square convergence rate derived here, from the rate derived for a triangular simultaneous equations model (*e.g*, Newey et al, 1997) is that the inclusion of all interaction terms between x_1 and d yields a higher order for the model studied here, and thus, a slower convergence rate.

It is well known that convergence rates of additive interaction models depend on the order of the model (where the order is bounded above by the dimension of the regressors in a fully non-parametric model, and is bounded below by one in a fully additive no-interactions model) (Andrews and Whang, 1991). Newey *et al* show that the correction for endogeneity in a triangular model is

simply including the residual from the reduced form in the primary equation; the correction term is additive. It is well known that additivity helps alleviate the curse of dimensionality in series estimation, since the estimation correction for additive models is to have no interactions between the additive sets of regressors. In contrast, the reduced form in (2.1.2), though additive, includes every interaction between the two two sets of variables (x_1 and d). Accordingly, the convergence rate for this estimator cannot attain Stone's bound. We now present the set of assumptions needed in order to prove the first lemma.

Assumption IV-1 $\{(y_i, w_i, x_{1i}, d_i)\}$, $(i = 1, \ldots, n)$ *i.i.d,* $Var(y|w)$ *is bounded and* $E[(y - m(x_1, d)^4|w]$ *is bounded.*

Assumption IV-2 *For every* K*, there is a non-singular constant matrix* D *such that for* $Q^K(x_1, d) = Dq^K(x_1, d)$, *i) the smallest eigenvalue of* $E[Q^K(x_1, d)Q^K(x_1, d)]$ *is bounded away from zero uniformly in* K *and ii) there is a sequence of constants* $\zeta(K)^2 K/n \to 0$ *as* $n \to \infty$.

Assumption IV-3 *There is some* α, β_K, *such that* $|m - q^K \beta_K| = O(K^{-\alpha})$ *as* $K \to \infty$

Assumption IV-4 *There exists a function* $D(m, \tilde{m})$ *which is linear in* m *such that for some* $(C, \delta) > 0$, *and all* $\tilde{m}$, $\overline{m}$, *with* $|\tilde{m} - m| < \delta$ *and* $|\overline{m} - m| < \delta$, *it is true that* $\|m - \tilde{m} - D(m - \tilde{m}; \tilde{m}\| \leq C(|m - \tilde{m}|)^2$ *and* $\|D(m; \tilde{m}) - D(m; \overline{m})\| \leq |m^z \| \tilde{m} - \overline{m}|$.

Note that, atypically, no assumption need be made requiring the endogenous component (*i.e*, the dummy variables) to be continuously distributed.

<u>**Lemma 1**</u> (*Mean-square error convergence Rate of the NPIV estimator*)

If Assumptions (IV-1)-(IV-4) are satisfied, $\dfrac{\zeta(K)^2 K}{n} \to 0$, *as* $n \to \infty$, *and* $\|\frac{A'\epsilon}{n}\| = O_p(\varepsilon_n)$ *for some* ε_n, *then,*

$$\int \tau[x_1, d] \; [\hat{m}(x_1, d) - m(x_1, d)]^2 dF(x_1, d) = O_p(2\varepsilon_n + \frac{K}{n} + K^{-2\alpha})$$

Proof: Appendix A

It is precisely the interaction between the sequence of functions approximating $\breve{m}(x_1)$ and d that results in this higher convergence rate relative to a triangular simultaneous equations model; the higher convergence rate is captured by the leading term.

Next, turn to asymptotic normality of $\hat{m}(x_1, d)$. Some additional notation for the asymptotic variance-covariance matrix of the estimated function is required prior to stating this result. The sequence of approximating functions for the equation $y = m(x_1, d) + \epsilon$ is given by $p^K(x_1, d)$, or equivalently $A(x_1, d)$ as shown above. Define

$$\Sigma_{A(\Upsilon),A(d)} = E[A(x_{1i}, \Upsilon_i)' A(x_{1i}, d_i)] \tag{1.3.1.1}$$

$$\Omega_{A(\Upsilon)\epsilon} = E[A(x_1, \Upsilon)\ A(x_1, \Upsilon)\ Var(y|x, z)\ ] \tag{1.3.1.2}$$

$$\Sigma_{A(d)} = E[\ A(x_{1i}, d_i)\ A(x_{1i}, d_i)'] \tag{1.3.1.3}$$

$$\Omega_{A(d)\epsilon} = E[A(x_{1i}, d_i)\ A(x_{1i}, d_i)' Var(y|x, z)] \tag{1.3.1.4}$$

$$\breve{H} = E[\tau(x_{1i}, d_i)\ r^L(w_i) \otimes A(x_{1i}, d_i)] \tag{1.3.1.5}$$

The asymptotic variance of $\hat{m}(x_1, d) = A(x_1, d)\ \hat{\beta}^{IV}$ is then given by:

$$\begin{aligned} V(\hat{m}(x_1, d)) &= A(x_1, d)\ V(\hat{\beta^{IV}})\ A(x_1, d) \\ &= A(x_1, d)\ \{\Sigma^{-1}_{A(\Upsilon),A(d)}\ [\Omega_{A(\Upsilon)\epsilon} + (\breve{H}\ \Sigma_{A(d)}\ \Omega^{-1}_{A(d)\epsilon}\ \Sigma_{A(d)}\ \breve{H}')]\Sigma^{-1}_{A(\Upsilon),A(d)}\}\ A(x_1, d)' \end{aligned}$$

Lemma 2 [*Asymptotic Normality of* $\hat{m}(x_1, d)$]

If Assumptions IV-1-IV-4 are satisfied and $\sqrt{n}K^{-s/r} \to 0$,

$$\sqrt{n}\ \hat{V}^{1/2}\ [\hat{m}(x_1,d) - m(x_1,d)] \xrightarrow{d} N(0,I)$$

Proof: Appendix A

1.4 Extensions

1.4.1 Triangular Simultaneous Equations Models

The first extension we discuss is to the triangular specification considered by Newey, Powell and Vella (1997), which is a slightly more restrictive specification than the model considered in (1.2), but one with much wider empirical applications. In their paper, Newey *et al* derive a two-step estimator in which an estimate of the residual from the first step is included in the primary equation as a regressor. An assumption required in deriving the large-sample properties of this estimator is that the regressors be continuously distributed, an assumption that excludes the dummy endogenous variables case. Here, we modify the theory using the transformation described earlier for the triangular model 2-step NPIV estimator to cover a wider class of models.

Consider the model:

$$\begin{aligned} y &= m(x_1,d) + \epsilon \\ d &= f(x) + \nu \end{aligned} \qquad (1.4.1.1)$$

where $x = [x_1\ x_2]$, $E[\nu|x] = 0$, and $E[\epsilon|x,\nu] = E[\epsilon|\nu]$. Conditioning on all regressors,

$$\begin{aligned} E(y|x,d) &= m(x_1,d) + E(\epsilon|x,d) \\ &= m(x_1,d) + E(\epsilon|x,\nu) \\ &= m(x_1,d) + E(\epsilon|\nu) \\ &= m(x_1,d) + \lambda(\nu) \end{aligned} \qquad (1.4.1.2)$$

where $\lambda(\cdot)$ is an unknown function. This observation, which forms the basis for the estimation strategy suggested by Newey *et al* indicates that the model may be estimated using approximating

functions for both m and λ, and no interactions between (x_1, d) and $\hat{\nu}$ to reflect the additive nature of the model. Although ν is unavailable, the additive structure of the model permits an estimate $\hat{\nu}$ from nonparametric estimation of step one.

Below we extend this model to the case where the (i) d are dummy endogenous and (ii) where both d and x_1 are endogenous. For the case where only the d are endogenous,

$$\begin{aligned} y &= m(x_1, d) + \epsilon \\ &= \tilde{m}(x_1) + \check{m}(x_1)d + \epsilon \end{aligned} \tag{1.4.1.3}$$

using the transformation from (1.1), so that conditioning only on the exogenous variables,

$$E(y|x) = \tilde{m}(x_1) + \check{m}(x_1)E(d|x) \tag{1.4.1.4}$$

which may be estimated by getting a nonparametric estimate of $E(d|x) = \hat{d} = r^L(w)\hat{\gamma}$ in the first stage and using this to construct an approximating sequence for (4.1.4). For example, an approximating sequence could consist of functions of x_1 and each of these functions multiplied by $\hat{d}$ which we denote $A^K[x_1, \hat{d}]$. Note: the estimation strategy here is analogous to two-stage least squares, but not the instrumental variables strategy suggested in the previous section. We have,

$$\hat{m}(x_1, d) = \hat{\beta} A^K[x_1, d] = \hat{\beta}_0 A_x^K(x_1) + \hat{\beta}_1 A_{x,d}^K[x_1 * d]$$

$$\tau[x, d] = \Pi_{j=1}^{2*d_{x_1}} \quad 1 \quad (a_j \leq x_{1j} \leq b_j) 1(0 \leq d_j \leq 1)$$

$$\hat{\beta} = \left\{ \sum_{i=1}^{N} \tau[x_{1i}, \hat{d}_i] \; [\; A(x_{1i}, \hat{d}_i) A(x_{1i}, \hat{d}_i)' \right\}^{-1} \cdot \left\{ \sum_{i=1}^{N} \tau[x_{1i}, \hat{d}_i] \; A(x_{1i}, \hat{d}_i)' y_i \right\}$$

where $A_x^K(\cdot)$ represents those functions in $A^K[x_1, d]$ consisting only of functions of x_1 and $A_{x,d}^K(\cdot)$ represents the rest. The consistency arguments and asymptotic normality results follow logically from those presented in Newey *et al* with the primary difference being that the assumption requiring the regressors to be continuously distributed (Assumption 3 for lemma 4.1 in Newey, Powell, Vella (1997)) can be relaxed to allow for non-continuously distrributed variables in the parametric

component. Then, Lemma 4.1 in Newey *et al* holds even if Assumption 3 does not; it is stated below.

For a random matrix Y, denote $\|Y\|_v = \mathbf{E}[\|Y\|^v]^{1/v}$ $\forall v < \infty$, $\|Y\|_\infty$ denote the infimum of constants C such that $\mathcal{P}$ $(\|Y\| < C) = 1$, and for a matrix D, let $\|D\| = [trace(D'D)]^{1/2}$. Let w denote the set of instrumental variables which includes those variables in x uncorrelated with θ, $\epsilon_i = y_i - m(x_{1i}, d_i)$, $\nu_{il} = d_i - f(x_i)$. Let $\mathcal{W} = \{w : \tau(w) = 1\}$. Denote r for the dimension of $[x_1, d]$ and a for the dimension of x.

Assumption NPV-1 $\{(y_i, x_i, d_i, \epsilon_i)\}$, $(i = 1, \ldots, n)$ *i.i.d,* $Var(y_i|x_{1i}, d_i)$ *and* $Var(d_i|x_{il})$ *bounded away from zero,* $E[\|\nu^4\||x]$ *is bounded, and* $E[\epsilon^4|x_1, d]$ *is bounded.*

Assumption NPV-2 *w is continuously distributed with density that is bounded away from zero on its support, and the support of w is a cartesian product of compact, connected intervals. The density of w is bounded away from zero on $\mathcal{W}$, and $\mathcal{W}$ is contained in the interior of the support of u.*

Assumption NPV-3 *$f(\cdot)$ is Lipschitz and continuously differentiable of order s_1 on the support of z, and $m(\cdot)$ is Lipschitz and continuously differentiable of order s on $\mathcal{U}$.*

Lemma 3

If Assumptions $T1 - T2$ are satisfied, the number of terms in K and L are fixed to be proportional, respectively, to $n^{-r/r+2s}$ and $n^{-a/z+2s_1}$ and $\frac{\zeta(K)^2 K}{n} \to 0$, $\frac{\upsilon(L)^2 L}{n} \to 0$ as $n \to \infty$,

$$\int \tau[x_1, d]\ [\hat{m}(x_1, d) - m(x_1, d)]^2 dF(x_1, d) = O_p(\frac{K}{n} + K^{-2s/r} + \frac{L}{n} + L^{-2s_1/a})$$

Proof: Appendix A

Next consider the case where both d and x_1 are endogenous. The system of equations we are now considering is:

$$\left\{ \begin{array}{c} y = m(x_1, d) + \epsilon \\ d = f(x_2) + \nu_d \\ x_1 = g(x_2) + \nu_x \end{array} \right\}$$

Again, the primary equation may be re-written as given in (4.1.3):

$$y = \tilde{\tilde{m}}(x_1) + \breve{m}(x_1)d + \epsilon \tag{1.4.1.5}$$

The strategy now is to condition on the exogenous variables as well as the continuously distributed variables, not the entire set of regressors as done for the standard triangular model. Then,

$$\begin{aligned} E(y|x) &= \tilde{\tilde{m}}(x_1) + \breve{m}(x_1)E(d|x) + E(\epsilon|x_2, \nu_x) \\ &= \tilde{\tilde{m}}(x_1) + \breve{m}(x_1)E(d|x) + E(\epsilon|\nu_x) \\ &= \tilde{\tilde{m}}(x_1) + \breve{m}(x_1)E(d|x) + \lambda(\nu_x) \end{aligned} \tag{1.4.1.6}$$

The estimation of such a model takes place in three steps. In one step, the reduced form for the continuously distributed endogenous variables is nonparametrically estimated and $\hat{\nu}_x = x_1 - \hat{g}(x_2) = x_1 - r^L(x_2)'\hat{\gamma}_x$ obtained; second, the reduced form for d is estimated to obtain an estimate $E(d|x_2) = r^L(x_2)'\hat{\gamma}_d = \hat{d}$. Each of these estimates is used in constructing the approximating sequence for the primary equation. Since the equation is additive in functions of x_1, $(x_1, \hat{d})$ and $\hat{\nu_x}$, the approximating sequence is composed of functions of each of these components and no interactions of the three sets. For example, assume the approximating sequence is a $K \times 1$ vector of functions with these three components, $A^K[x_1, \hat{d}, \hat{\nu}_x]$ We then have,

$$\hat{m}(x_1, d) = \hat{\beta}\, A^K[x_1, \hat{d}, \hat{\nu}_x] = \hat{\beta}_0\ A^K(x_1) + \hat{\beta}_1\ A^K[x_1 * \hat{d}] + \hat{\beta}_2\ A^K[\hat{\nu}_x]$$

$$\tau[x, d, \nu_x] = \Pi_{j=1}^{2*d_{x_1}} \quad \mathbf{1} \quad (a_j \leq x_{1j} \leq b_j)\ \mathbf{1}(c_j \leq \nu_{xj} \leq d_j)\ \ \mathbf{1}(0 \leq d_j \leq 1)$$

$$\hat{\beta} = \left\{ \sum_{i=1}^{N} \tau[x_{1i}, \hat{d}_i, \nu_{xi}] \; [\; A^K[x_{1i}, \hat{d}_i, \hat{\nu}_{xi}\;] A^K[x_{1i}, \hat{d}_i, \hat{\nu}_{xi}\;]' \right\}^{-1} \cdot \left\{ \sum_{i=1}^{N} \tau[x_{1i}, \hat{d}_i, \nu_{xi}] \; A^K[\; x_{1i}, \hat{d}_i, \nu_{xi}\;]' y_i \right\}$$

1.4.2 Discrete Endogenous Variables

Next consider the case where the endogeneity stems from discrete variables in the triangular simultaneous equation system above. Let x_1 represent continuous exogenous variables and d a discrete variable with finite support $[a_1, a_n]$, We proceed by rewriting (41) as:

$$\begin{aligned} y &= m(x_1, d) + \epsilon \\ &= m_1(x_1)\mathbf{1}_{\{d=a_1\}} + m_2(x_1)\mathbf{1}_{\{d=a_2\}} + \ldots + m_K(x_1)\mathbf{1}_{\{d=a_n\}} + \epsilon \end{aligned} \tag{1.4.2.1}$$

where $\mathbf{1}_{\{A\}}$ is the indicator function for the event A, and K is the number of points in supp(d). Then,

$$\begin{aligned} E(y|x) &= m_1(x_1)E(\mathbf{1}_{\{d=a_1\}}|x) + \ldots + m_K(x_1)E(\mathbf{1}_{\{d=a_n\}}|x) \\ &= m_1(x_1)(Pr(d=a_1)|x) + \ldots + m_K(x_1)Pr(d=a_n)|x) \end{aligned} \tag{1.4.2.2}$$

where the $(E(\mathbf{1}_{\{d=a_1\}}|x), \ldots, E(\mathbf{1}_{\{d=a_n\}}|x))$ can be estimated as a series of binary dependent variable models just as in the dummy endogenous variables case. Series estimation of m is then carried out by using interactions of functions of x_1 with each of the estimated probabilities. It may be difficult to identify each m_i function, $(i = 1, \ldots, K)$ separately, and this may not even be an interesting exercise to pursue. Re-writing the model as we have done above allows us to recover an estimate of m in a straighforward manner. Again, this exploits the discrete nature of the variable to treat the model as one in which the endogenous variable were parametric.

1.5 Application of the Proposed NPIV Estimator to the Estimation of Flexible Panel Data Models

One application of the proposed NPIV estimators is to the estimation of flexible panel data models of the kind recently proposed by Porter (1996). Each of the existing set of estimators for non-

parametric panel data models relies on an analogue of the differencing technique proposed for the linear model; these include Manski (1987), Abrevaya (1996) and Porter (1997). An unaddressed feature of these estimators is that differencing the model to annihilate the individual heterogeneity term also differences out the time-invariant component, just as in the linear panel model, even if m is nonlinear in the model $y_{it} = m(x_{it}, z_i) + \theta_i + \epsilon_{it}$. We begin by demonstrating this problem in the estimation of generalized panel models and then discuss how NPIV estimation can address this problem. Thus, this section provides the nonparametric analogue to the method proposed by Hausman and Taylor (1981). The analysis below is not restricted to models with dummy endogenous variables, although the discussion applies equally to such models.

Consider the panel model

$$y_{it} = m(x_{it}, z_i) + \theta_i + \epsilon_{it} \tag{1.5.1}$$

where $i(i = 1, \ldots, n)$ indexes the individuals, $t(t = 1, \ldots, T)$ the time periods, y_{it} is the dependent variable and x_{it} and z_i denote the time-variant and time-invariant variables respectively. Individual heterogeneity, which is assumed to be correlated with some subset of teh regressors, enters additively and is denoted θ_i, ϵ_{it} represents random noise and $m(\cdot, \cdot)$ is the unknown and unrestricted function that is the object of estimation.

Even though $m(\cdot, \cdot)$ is unknown and potentially nonlinear, and in general $m(x_{it}, z_i)$ - $m(x_{i,t-1}, z_i) \neq m(x_{it} - x_{i,t-1}, z_i - z_i = 0)$, the methods that have been used by researchers to estimate the model by differencing does result in partialling out z. For example, Manski (1987) and Abrevaya (1996) consider a special (semi-parametric) case of (5.1) where m is invertible and the model studied is : $y_{it} = m(x_{it}'\beta + z_i'\gamma + \theta_i + \eta_{it})$, which is estimated using rank-based methods. For this model it is apparent how first-differencing results in the loss of all time-invariant variables. Surprisingly, even the more general series and kernel based methods, which are the focus of Porter (1996)'s work, are not robust to the differencing approach. Consider series estimation of (5.1) which begins with the assumption that the unknown function $m(\cdot, \cdot)$ can be approximated by a $K \times 1$ vector of approximating functions, $p^K(x, z) = (p_{1K}(x, z), \ldots, p_{KK}(x, z))'$. For example, a power series approximation yields:

$$m(x_{it}, z_i) \approx p^K(x_{it}, z_i) = \sum_{l_1=0}^{L_1} x_{it}^{l_1}\beta_{l_1} + \sum_{l_2=1}^{L_2}\sum_{l_3=1}^{L_3} x_{it}^{l_2} z_i^{l_3}\gamma_{l_2,l_3} + \sum_{l_4=1}^{L_4} z_i^{l_4}\delta_{l_4}$$

which produces the associated differenced approximating functions:

$$\begin{aligned} q^K(x_{it}, x_{i,t-1}, z_i) &= p^K(x_{it}, z_i) - p^K(x_{i,t-1}, z_i) \\ &= \sum_{l=1}^{L_1}(x_{it}^{l_1} - x_{i,t-1}^{l})\beta_{l_1} + \sum_{l_2=1}^{L_2}\sum_{l_3=1}^{L3}(x_{it}^{l_2}z_i^{l_3}) - (x_{i,t-1}^{l_2}\dot{z}_i^{l_3})\gamma_{l_2,l_3} \end{aligned} \tag{1.5.2}$$

Clearly, irrespective of the underlying true form of $m(\cdot,\cdot)$, the series differencing approach yields no estimate of $\{\delta\}$, and thus, under the specification in (5.1), an estimate of the function $m(\cdot,\cdot)$, cannot be recovered. What is recovered, instead, is some time-varying component of m. Denoting the unknown function as $m(x_{it}, z_i) = m^x(x_{it}) + m^{x,z}(x_{it} * z_i) + m^z(z_i)$, differencing only allows an estimate of the time-varying component of the function, $\hat{m}^x(x_{it}) + \hat{m}^{x,z}(x_{it} * z_i)$.

To extend Hausman-Taylor (1981) to the estimation of this model, consider rewriting the equation as:

$$\tilde{y} = y - \hat{m}^x(x_{it}) - \hat{m}^{x,z}(x_{it}, z_i) = m^z(z_i) + \theta_i + \eta_{it} \tag{1.5.3}$$

where η_{it} is the new error term that arises from replacing m^x and $m^{x,z}$ with their estimated values. The idea is that if $m^z(z)$ can now be estimated consistently, we will have consistent estimators of each of the three components of $m(\cdot,\cdot)$ and can thereby construct an estimate of m. The advantage of differencing the model in lieu of applying the NPIV estimator directly is that the exogenous x are valid instruments for those z that are correlated with θ in the differenced equation.

Equation (5.3) can now be estimated by applying the appropriate NPIV estimator. If the z are dummies, the model can be transformed and the dummy variable estimator proposed in Sections 2 and 4 can be applied. If the equation is a part of a triangular simultaneous equation system, and the z are continuous variables, the estimator proposed by Newey *et al* (1997) may be applied.

1.6 Conclusion

This paper extends the body of work on non-parametric instrumental variables (NPIV) estimation by presenting the estimators for the set of models in which the endogeneity is either (1) strictly associated with dummy (or more generally, discrete) variables in nonparametric systems of simultaneous equations or triangular simultaneous equations systems or (2) associated with both dummy and continuously distributed variables in triangular simultaneous equations systems.

The essential results of this paper rest on the observation that for any non-parametric model, a simple transformation parametrizes the discrete component of the model. This observation is shown to be useful in two different ways. First, for any non-parametric sytems of simultaneous equations, the model cannot be identified if the endogneiety stems from continuously distributed variables. Newey and Powell (1998) link this problem to that of solving a linear integral equation of the first kind for which no unique solution exists. We demonstrate that where the endogeneity is strictly associated with dummy variables, rewriting the model so that the dummy variable is no longer a component of the unknown function avoids this problem so that the model is identified; the transformation lies at the heart of this identification condition. The proposed estimator, based on series estimation, is shown to be consistent and asymptotically normal.

For the estimator proposed by Newey, Powell and Vella (1997) for triangular simultaneous equations systems, we extend the applicability of the estimator to those models in which the endogeneity either derives solely from dummy variables or from both dummy and continuous variables. The estimator proposed by Newey *et al* requires continuously distributed variables. Relaxing this assumption expands the applicability of the estimator without affecting the convergence rate or the large sample properties.

There is a second benefit of applying the parametrizing transformation. For models without endogeneity, the existing theory requires the regressors to be continuously distributed in deriving the consistency and asymptotic normality results. This assumption is required only for those regressors that are an argument of the unknown function. Applying the transformation by 'pulling out' the discrete component of the model effectively parametrizes the discrete component for which the standard assumption can be relaxed, allowing the existing theory to cover models with dummy endogenous variables. These include program evaluation models which frequently arise in empirical applications.

Finally, we show that one application of the estimators proposed in the paper, as well as the estimator due to Newey *et al* is to the estimation of flexible panel data models. Each of the existing semi-parametric and non-parametric panel data estimators due to Manski (1987), Abrevaya (1996) and Porter (1996) relies on an analogue of the first-differencing technique proposed for the linear panel model and each of these suffers from an important drawback: they difference out the time-invariant component of the model along with the individual fixed-effect. Section 5 demonstrates this problem and discusses how NPIV estimation can be applied to the panel problem so that the time-invariant component may be estimated even if the model is differenced to begin with. Thus, we provide the non-parametric analogue to the method proposed by Hausman and Taylor (1981) for the linear panel model.

1.7 References

Abrevaya, J.I (1996). "Semiparametric Estimation Methods for Nonlinear Panel Data Models and Mismeasured Dependent Variables', MIT PhD.

Andrews, D.W.K (1991). "Asymptotic normality of series estimators for nonparametric and semi-parametric models," Econometrica 59, 307-345.

Andrews, D.W.K and Y.J Whang (1990). "Additive interactive regression models: circumvention of the curse of dimensionality," Econometric Theory 6, 466-479.

Hausman, J.A and William Taylor (1981). "Panel Data and Unobservable Individual Effects," Econometrica, v49 n6 Nov. 1981, pp. 1377-98.

Honore, Bo (1994). "Pairwise Difference Estimators of Censored and Truncated Regression Models," Journal of Econometrics, v64 n1-2 Sept.-Oct. 1994, pp. 241-78.

Manski, C.F. (1987). "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data, " Econometrica; v55 n2 March 1987, pp. 357-62.

Newey, W.K. (1990). "Efficient Instrumental Variables Estimation of Nonlinear Models," Econo-

metrica, 58 809-837.

Newey, W.K and J. Powell (1990). "Nonparametric Instrumental Variables Estimation", Working Paper, MIT Department of Economics.

Newey, W.K (1995). "Convergence Rates for series estimators", In Maddala, G.S., Phillips, P.C.B., Srinivasan, T.N (Eds.) *Statistical Methods of Economics and Quantitative Economics: Essays in Honor of C.R. Rao* , Blackwell, Cambridge, USA, pp. 254-275.

Newey, W.K. (1996). "Convergence Rates and Asymptotic Normality for Series Estimators," MIT Department of Economics.

Newey, W.K. (1997a). Unpublished lecture notes, MIT Department of Economics.

Newey, W.K, Powell J.L, Vella, F (1997). "Nonparametric Estimation of Triangular Simultaneous Equations Models," MIT, Department of Economics.

Newey, W.K. (1997b). "Convergence Rates and Asymptotic Normality for Series Estimators," Journal of Econometrics, 79, 147-168.

Porter, Jack (1996). "Nonparametric Regression Estimation for a Flexible Panel Data Model", MIT mimeo.

Roehrig, C (1988). "Conditions for Identification in Nonparametric and Parametric Models," Econometrica, 56, 433-477.

Stone C.J. (1982). "Optimal global rates of convergence for nonparametric regression." Annals of Statistics 10, 1040-1053.

1.8 Appendix A

Proof for Lemma 1

$y = \bar{\tilde{m}}(x_1) + \breve{m}(x_1)d + \epsilon = m(x_1, d) + \epsilon$. Assume that $m(\cdot)$ can be approximated by some $(K \times 1)$ functions $a^K(x_1, d)$. Let $a^K(x_1, \Upsilon)$ be another set of functions, so $y = a^K(x_1)\overline{\beta_1} + a^K(x_1)d\overline{\beta}_2 - m_K + m$. Let a_d denote $a^K(x_1, d)$ and a_Υ denote $a^K(x_1, \Upsilon)$. $A_d = (A_{d1}, \ldots, A_{dn})$ and $A_\Upsilon = (A_{\Upsilon 1}, \ldots, A_{\Upsilon n})$. Assume $\|\frac{A_d'\epsilon}{n}\| = O_p(\epsilon_n)$ for some ϵ_n. w denotes $[x_1, z]$ (z is the set of instruments).

Let $\hat{\beta}^{IV} = (A_\Upsilon' A_d)^- A_\Upsilon' y$

$= (A_\Upsilon' A_d)^- A_\Upsilon'(A_d\overline{\beta} + (m_o - m^K) + \epsilon) \Rightarrow (\hat{\beta}^{IV} - \beta) = (A_\Upsilon' A_d)^- A_\Upsilon'(m_o - m^K) + (A_\Upsilon' A_d)^- A_\Upsilon'\epsilon$

Denote $\Sigma_n = E\left(\frac{A_\Upsilon' A_d}{n}\right)$, $\quad \Sigma_{A_\Upsilon A_\Upsilon} = E\left(\frac{A_\Upsilon' A_\Upsilon}{n}\right)$.

We're interested in characterizing $\|\hat{\beta}^{IV} - \beta\|^2$:

$\|\hat{\beta}^{IV} - \beta\|^2 \leq \|(A_\Upsilon' A_d)^- A_\Upsilon'(m_o - m^K)\|^2 + \|(A_\Upsilon' A_d)^- A_\Upsilon'\epsilon\|^2$ Analyzing each of the terms separately:

$\|(A_\Upsilon' A_d)^{-1} A_\Upsilon'\epsilon\|^2 = \| \left(\frac{A_\Upsilon' A_d}{n}\right)^{-1} \frac{A_\Upsilon'\epsilon}{n}\|^2$

$= \left\| \Sigma_{A_d A_\Upsilon}^{-1} \frac{A_\Upsilon'\epsilon}{n} + \left[\left(\frac{A_\Upsilon' A_d}{n}\right)^{-1} - \Sigma_{A_d A_\Upsilon}^{-1}\right] \frac{A_\Upsilon'\epsilon}{n} \right\|^2$

$\leq \left\| \Sigma_{A_d A_\Upsilon}^{-1} \frac{A_\Upsilon'\epsilon}{n} \right\|^2 + \left\| \left(\frac{A_\Upsilon' A_d}{n}\right)^{-1} \left[\Sigma_{A_d A_\Upsilon} - \left(\frac{A_\Upsilon' A_d}{n}\right)\right] \Sigma_{A_d A_\Upsilon}^{-1} \frac{A_\Upsilon'\epsilon}{n} \right\|^2$ by the triangle inequality

$\leq \left\| \Sigma_{A_d A_\Upsilon}^{-1} \right\|^2 \left\| \frac{A_\Upsilon'\epsilon}{n} \right\|^2 + \left\| \left(\frac{A_\Upsilon' A_d}{n}\right)^{-1} \left[\Sigma_{A_d A_\Upsilon} - \left(\frac{A_\Upsilon' A_d}{n}\right)\right] \Sigma_{A_d A_\Upsilon}^{-1} \frac{A_\Upsilon'\epsilon}{n} \right\|^2$.

Den oting A for $\Sigma_{A_d A_\Upsilon}$, $\hat{Q}$ for $\left(\frac{A_\Upsilon' A_d}{n}\right)$) and R_n for $\left(\frac{A_\Upsilon'\epsilon}{n}\right)$ the above expression is

$= \|\Sigma_{A_d A_\Upsilon}^{-1}\|^2 \|R_n\|^2 + R_n'(Q')^{-1}(Q - \hat{Q})'\hat{Q}^{-1\prime}\hat{Q}^{-1}(-\hat{Q})Q^{-1}R_n$

Consider the second term in this expression first.

$R_n'(Q')^{-1}(Q - \hat{Q})'\hat{Q}^{-1\prime}\hat{Q}^{-1}(Q - \hat{Q})Q^{-1}R_n$

$\leq CR_n'(Q')^{-1}(Q - \hat{Q})'(Q - \hat{Q})(Q')^{-1}R_n$

which holds if (1)$\|Q - \hat{Q}\| = o_p(1)$ and (2)$\lambda_{min} \geq C$ where λ denotes an eigen-value; then, (1) and (2) $\Rightarrow$ Q invertible, and by Slutsky, $\hat{Q}$ invertible, or $\lambda_{min}(\hat{Q}) > C \Rightarrow \lambda_{min}(\hat{Q}'\hat{Q}) \geq C \Rightarrow \lambda_{max}(\hat{Q}^{-1\prime}\hat{Q}^{-1}) \leq C$.

Further, if $\lambda_{min}(Q) \geq C$ wpa 1, $\Rightarrow \lambda_{max}(Q^{-1}) = O_p(1)$ and $C \geq \lambda_{max}(\hat{Q}^{-1}\hat{Q}^{-1})$.

Then, by Cauchy-Schwartz, $CR_n'(Q')^{-1}(Q-\hat{Q})'(Q-\hat{Q})(Q')^{-1}R_n \leq C\|(Q-\hat{Q})\|^2\|Q^{-1}R_n\|^2$ $\leq o_p(1) + \|R_n\|^2\lambda_{max}(Q^{-1})$ (only if Q^{-1} is positive semi-definite, which is necessary to assume for identification).

By assumption, we have $\|R_n\| = O_p(\varepsilon_n)$. An explicit expression for ε_n can be derived as follows:

Analyzing $R_n = \frac{A_\Upsilon'\epsilon}{n}$: $E\|R_n\|^2 = E\frac{\epsilon' A_\Upsilon A_\Upsilon'\epsilon}{n} = E(trace(R_n'R_n)|w) =$

$E(trace\left(\frac{\epsilon' A_\Upsilon A_\Upsilon'\epsilon}{n}\right)|w) = trace\left(\frac{A_\Upsilon' A_\Upsilon}{n}E(\epsilon\epsilon'|w\right) = trace\left(\frac{A_\Upsilon' A_\Upsilon}{n}Var(\epsilon|w)\right) \leq C\,trace\left(\frac{A_\Upsilon' A_\Upsilon}{n}\right) =$

$C\,E\left(trace\frac{A_\Upsilon' A_\Upsilon}{n}|w\right) = C\,trace\,E\left(\frac{A_\Upsilon' A_\Upsilon}{n}|w\right) = C\frac{K}{n}$.

Therefore, by Markov's inequality, $\|R_n\|^2 = O_p(\frac{CK}{n})$. It follows that $E\|\frac{A_\Upsilon' A_d}{n}^{-1}\frac{A_\Upsilon'\epsilon}{n}\|^2$

$\leq \lambda_{max}(Q^{-1})^2 O_p(\frac{CK}{n}) + o_p(1)\,\lambda_{max}(Q^{-1})^2 O_p(CK/n)$

$= O_p(1)O_p(\varepsilon_n) + o_p(1)O_p(1)O_p(\frac{CK}{n}) = O_p(\varepsilon_n + \frac{CK}{n})$, thus, $\left\|\frac{A_\Upsilon' A_d}{n}^{-1}\frac{A_\Upsilon'\epsilon}{n}\right\| = O_p(\varepsilon_n + \frac{CK}{n})$.

Assume $sup_{x_1,d}|m(x_1,d) - \hat{m}(x_1,d)(\overline{\beta}| = O(K^{-\alpha})$.

Next, $\|\frac{A_\Upsilon' A_d}{n}^{-1} A_\Upsilon'(m-m_K)\|^2 = E(\|\left(\frac{A_\Upsilon' A_d}{n}\right)^{-1}{}'(m - A_d\overline{\beta})\|^2 =$

$\|\Sigma_{A_d A_\Upsilon}^{-1}\frac{A_\Upsilon'\epsilon}{n} + \left[\left(\frac{A_\Upsilon' A_d}{n}\right)^{-1} - \Sigma_{A_d A_\Upsilon}^{-1}\right] A_\Upsilon'(m - a^K\overline{\beta})\|^2$

$\leq \|\Sigma_{A_d A_\Upsilon}^{-1}\frac{A_\Upsilon'\epsilon}{n}\|^2 + \|\left(\frac{A_\Upsilon' A_d}{n}\right)^{-1}\left[\Sigma_{A_d A_\Upsilon} - \left(\frac{A_\Upsilon' A_d}{n}\right)\right]\Sigma_{A_d A_\Upsilon}^{-1} A_\Upsilon'(m - a^K\overline{\beta})\|^2$

$\leq \|\Sigma_{A_d A_\Upsilon}^{-1}\|^2\left\|\frac{A_\Upsilon'\epsilon}{n}\right\|^2 + \left\|\left(\frac{A_\Upsilon' A_d}{n}\right)^{-1}\left[\Sigma_{A_d A_\Upsilon} - \left(\frac{A_\Upsilon' A_d}{n}\right)\right]\Sigma_{A_d A_\Upsilon}^{-1} A_\Upsilon'(m - a^K\overline{\beta})\right\|^2$

(Denoting R_n for $\frac{A_\Upsilon'\epsilon}{n}$), $= \|\Sigma_{A_d A_\Upsilon}^{-1}\|^2\left\|\frac{A_\Upsilon'\epsilon}{n}\right\|^2 + \|\hat{Q}^{-1}(Q-\hat{Q})Q^{-1}R_n\|^2$

$\leq \|\Sigma_{A_d A_\Upsilon}^{-1}\|^2\left\|\frac{A_\Upsilon'\epsilon}{n}\right\|^2 + \|Q-\hat{Q}\|^2\|Q^{-1}R_n\|^2 = \|Q^{-1}\|^2\|R_n\|^2 + o_p(1)\lambda_{max}Q^{-1})^2\|R_n\|^2$.

Analyzing $\|R_n\|^2 = \left\|\frac{A_\Upsilon'(m - a^K\overline{\beta})}{n}\right\|^2 \leq \|\frac{A_\Upsilon}{n}\|^2\|m - a^K\overline{\beta}\|^2 = \|A_\Upsilon - \Sigma_{A_\Upsilon} + \Sigma_\Upsilon\|^2\|m - a^K\overline{\beta}\|^2 \leq$

$\|A_\Upsilon - \Sigma_{A_\Upsilon}\|^2 + \|\Sigma_{A_\Upsilon}\|^2\|m - a^K\overline{\beta}\|^2$ (where $\|A_\Upsilon - \Sigma_{A_\Upsilon}\|^2 = o_p(1)$ and Σ_{A_Υ} is finite)

$= C[(m - a^K\overline{\beta})'(m - a^K\overline{\beta})] = C(K^{-2\alpha})$.

So we have: $\|\left(\frac{A_\Upsilon' A_d}{n}\right)^{-1} A_\Upsilon'(m - A_d\overline{\beta})\|^2 = O_p(1)O_p(\varepsilon_n) + O_p(CK^{-2\alpha})$

and $\|\hat{\beta}^{IV} - \beta\|^2 = +O_p(\varepsilon_n + \frac{CK}{n}) + O_p(\varepsilon_n) + O_p(K^{-2\alpha})$

Now $\int[\hat{m}(x_1,d) - m(x_1,d)]^2 dF(x_1,d) = \int[a^K(x_1,d)'(\hat{\beta}^{IV} - \overline{\beta}) + a^K(x_1,d)(\overline{\beta}) - m(x_1,d)]^2 dF(x_1,d)$

$$\leq \|\hat{\beta}^{IV} - \beta\|^2 + \int [a^K(\overline{\beta} - m(x_1, d)]^2 dF(x_1, d)$$

$$= O_p(\varepsilon_n + \tfrac{K}{n}) + O_p(K^{-2\alpha} + \varepsilon_n) = O_p(2\varepsilon_n + \tfrac{K}{n} + K^{-2\alpha})$$

Proof of Lemma 2 :

Methods for the asymptotic normality of series estimators are presented in Andrews (1992, 1994) and Newey (1995, 1996, 1997a, 1997b). The proof below extend these since parts of the existing proofs require conditioning on the entire set of variables to obtain the needed result, which is not done here. Other modifications to account for the non-symmetric second moment matrix $A_\Upsilon' A_d$ are made. Adapting Newey (1995):

Let $F = Var(\hat{\beta}^{IV}) = V_K^{-1/2} = (\Sigma_{A_\Upsilon' A_d})^{-1} \Sigma_{' A_d \epsilon} (\Sigma_{A_\Upsilon' A_d})^{-1}$, where $\Sigma_{' A_d \epsilon} = E[a^K a^K [(y - \hat{m}(x_1, d))^2]$

$\|F\|^2 = trace(FF) = trace(V_K^{-1/2} V_K^{-1/2}) = C$ since $\Sigma_{A_\Upsilon' A_d}$ is finited and bounded (as necessary for identification), and $\Sigma_{' A_d \epsilon}$ is bounded by $Var(y|z)$ bounded.

Let $M = (m(x_{11}, d_1), \ldots, m(x_{1n}, d_n))$ and let $\overline{m}_K(x_1, d) = a^{K\prime} \beta_K$ following Assumption (IV-3). Let $\epsilon = Y - M$. Then,

$$\sqrt{n} V_K^{-1/2} = \sqrt{n} F[\hat{m} - m]$$

$$= \sqrt{n} F[\hat{m} - m - D(\hat{m}) + D(m)] + \left[F \Sigma_{A_\Upsilon' A_d} \frac{A_\Upsilon' \epsilon}{\sqrt{n}} \right] + \left[\sqrt{n} F \left(\frac{A_\Upsilon' A_d}{n}^{-1} - \Sigma_{A_\Upsilon' A_d} \right) \frac{A_\Upsilon' \epsilon}{n} \right] +$$

$$\left[\sqrt{n} F \left(\frac{A_\Upsilon' A_d}{n}^{-1} \right) \left(\frac{A_\Upsilon' (M - A_d \beta_K)}{n} \right] + [\sqrt{n} F [D(\overline{m} - D(m)]$$

newline which holds if $\hat{\beta}^{IV} \xrightarrow{p} \beta$, since $D(\hat{m}) - D(m) = (\hat{\beta}^{IV} - \beta)$.

By assumption (IV-3) the last term is equal to 0 wpa 1. Applying Cauchy-Schwarz,

$$\|\sqrt{n} F \left(\frac{A_\Upsilon' A_d}{n}^{-1} \right) \left(\frac{A_\Upsilon' (M - A_d \beta_K)}{n} \right) \|^2 = \sqrt{n} \| F \frac{A_\Upsilon' A_d}{n}^{-1} A_\Upsilon' (M - q^K \beta) \|^2 = \ newline$$

$$= \sqrt{n} \| F \Sigma_{A_\Upsilon' A_d}^{-1} \frac{A_\Upsilon (M -' A_d \beta_K)}{n} + \left[\left(\frac{A_\Upsilon A_d}{n} \right) - \Sigma_{A_\Upsilon A_d}^{-1} \right] \frac{A_\Upsilon' (M - A_d \beta_K)}{n} \|^2 \leq \sqrt{n} \|F\|^2 \| \Sigma_{A_\Upsilon A_d}^{-1} \frac{A_\Upsilon (M - A_d \beta_K}{n}$$

$$\left[\left(\frac{A_\Upsilon' A_d}{n}\right) - \Sigma_{A_\Upsilon A_d}^{-1}\right] \frac{A_\Upsilon'(M - A_d\beta_K)}{n}\|^2$$

by the triangle inequality, $\leq C\sqrt{n}\|\Sigma_{A_\Upsilon A_d}^{-1} \frac{A_\Upsilon(M-A_d\beta_K)}{n}\|^2 + \|\left(\frac{A_\Upsilon' A_d}{n}\right) - \Sigma_{A_\Upsilon A_d}^{-1} \frac{A_\Upsilon'(M-A_d\beta_K)}{n}\|^2$

$= \mathrm{C} \quad \lambda_{max}(A^{-1})^2 O_p(\sqrt{n}K^{-\alpha}) + o_p(1)\|R_n\|^2.$

$= \mathrm{C} \quad \lambda_{max}(A^{-1})^2 O_p(\sqrt{n}K^{-\alpha}) + o_p(1)[(m - a^K\bar{\beta})'(m - a^K\bar{\beta})] = o_p(1)$

Next, by Cauchy-Schwartz, $\sqrt{n}F\left(\frac{A_\Upsilon' A_d}{n}^{-1} - \Sigma_{A_\Upsilon A_d}\right)\frac{A_\Upsilon'\epsilon}{n}$

$\leq \sqrt{n}\|F\left(\frac{A_\Upsilon' A_d}{n}\right)^{-1}\left(\Sigma_{A_\Upsilon A_d} - \left(\frac{A_\Upsilon' A_d}{n}\right)\right)\Sigma_{A_\Upsilon A_d}^{-1}\frac{A_\Upsilon'\epsilon}{n}\|^2$

$\leq \sqrt{n}\|F\|^2\|\left(\frac{A_\Upsilon' A_d}{n}\right)^{-1}\left(\Sigma_{A_\Upsilon A_d} - \left(\frac{A_\Upsilon' A_d}{n}\right)\right)\|^2 \|\Sigma_{A_\Upsilon A_d}^{-1}\frac{A_\Upsilon'\epsilon}{n}\|^2$ o

$= \mathrm{C} \quad o_p(1)\lambda_{max}(\Sigma_{A_\Upsilon A_d}^{-1})^2 O_p(\varepsilon^2) = o_p(1)$

Denote $Z_{in} = F\Sigma_{A_\Upsilon A_d}\frac{A_\Upsilon'\epsilon}{\sqrt{n}}$. Since for each n, Z_{in} is i.i.d, Newey shows that $E[Z_{in}] = 0$, and we take a transform of the functions in A_Υ so that $\Sigma_i E[Z_{in}^2] = 1$. Then, $\Sigma_i Z_{in} \xrightarrow{d} N(0,1)$ and $F\Sigma_{A_\Upsilon A_d}\frac{A_\Upsilon'\epsilon}{\sqrt{n}} \xrightarrow{d} N(0,I)$. The remainder of Newey's (1995) proof for Theorem 2 holds, so $\sqrt{n}V_K^{-1/2}(\hat{m} - m) \xrightarrow{d} N(0,I)$. □

Proof of Lemma 3

Follows from the proof for Theorem 2 in Newey, Powell and Vella (1997). The convergence rate for $m(x_1, d)$ is given by

$$\int \tau(x_1,d)[\underline{\hat{m}}(x_1,d) - \underline{m}(x_1,d)]^2 dF(x_1,d) = O_p(K/n + K^{-2r/s} + L/n + L^{-2a/s_1})$$

If L is the number of functions used in the series approximation for the first step, it is apparent that the convergence rate takes the form of the sum of the 2 convergence rates: that of the second step alone (when the first did not have to be estimated) and the first step. From this it follows that

for a particular choice of K and L, the sum is simply the maximum of the two rates (White 1990):

$$\int \tau(w)[\hat{\underline{m}}^z(z) - \underline{m}^z(z)]^2 dF(u) = O_p(max\{n^{-2r/r+2s}, n^{-2a/a+2s_1}\})$$

Chapter 2

Nonparametric Estimation of the Sample Selection Model

2.1 Introduction

The specification and estimation of sample selection models has been one of the most active areas of research in the microeconometrics literature. Much of the interest in these models has emerged from the fact that sample selection is pervasive, arising in several econometric applications including topics in finance and labor economics, and is important for program evaluation and policy analyses. First studied by James Heckman (1976), the prototypical sample selection model required a complete specification of both the selection equation and the outcome equation, as well as the joint distribution of the error terms for consistent estimation. Recognizing that incorrect parametric modelling assumptions would lead to incorrect inference, a new literature has emerged with the intent of weakening and gradually eliminating the restrictions on the model originally proposed by Heckman to remove selectivity bias.

We contribute to this body of work by considering a non-parametric two-step estimator allowing for a non-parametric selection rule as well as a non-parametric outcome equation. The motivation for the work is twofold. First, sample selection is well known to be endemic in the evaluation of social welfare programs (*e.g.,* Currie et al (1996), Kane and Staiger(1996)) and in policy analyses (*e.g.,* the implications of a tax break on corporate investment behaviour) and is therefore a necessary tool in such studies where accurately estimated parameters may be especially important for inference. Thus, the estimator discussed here provides an alternative to those derived from parametrically-

specified sample selection models, which are commonly used in applied work (*e.g.*, Gronau 1976).

The second motivation is academic, and addresses the apparent asymmetry in the body of work on sample selection. Specifically, the modelling of sample selection following Heckman has focussed on weakening (or removing) parametric assumptions on either the sample selection rule or the outcome equation, but not both together. All the existing estimators are therefore *semi-parametric* since some aspect of these models are parametrically specified. Parametric assumptions on the functional form of a model, on the regression functions or on the selection correction term are not typically derived from some underlying economic theory, rendering parametric assumptions on any part of a model unduly restrictive and leaving estimators obtained from such models potentially inconsistent. The estimator we develop avoids specification errors due to *any* parametric assumptions by leaving the regression functions and the distribution of the pair of errors parametrically unspecified.

Sample selection models with weak or few parametric assumptions have been previously studied in the literature. The common theme among the proposed estimators in this body of work has been the assumption of an unknown joint distribution of the error terms and a focus on applying non-parametric regression methods to one equation (typically the selection rule) while leaving the other equation has fully specified. Amongst these, some estimators proceed directly with a non-parametric estimation of the parameters of interest, avoiding estimation of the unknown distribution altogether (Manski (1975), Han (1985), Cavanagh and Sherman (1993)) while others deal squarely with estimating the joint distribution along with the parameters of a (fully specified) selection equation using non-parametric methods; a non-parametric selection correction term is thus obtained, and substituted into a parametrically specified second-step; examples of this strategy include Coslett (1983) and Ichimura (1987). Estimation techniques have also been proposed for models in which the selection mechanism is non-parametric, while the outcome equation is a standard linear. These include Newey (1988), Powell and Ahn (1991), and Kyriazidou (1997). While Newey applies series estimation to the selection equation, and Powell and Ahn (1991) and Kyriazidou (1997) focus on kernel methods in order to derive the relevant correction term, and then proceed by estimating a partially linear outcome equation.

In all these works, the emphasis has been on removing the parametric assumptions on the sample selection rule, while little attention has been paid to the outcome equation. However, it appears that

if the objective is to dispense with parametric assumptions in order to reduce inconsistencies due to unknown functional forms, a *semi-parametric* treatment of the sample selection problem does not deal adequately with potential misspecification. Indeed, the gains in avoiding specification errors by weakening parametric restrictions on one part of the model may easily be lost by retaining parametric forms in other parts of the model. In this paper, we seek to avoid this problem by studying a model that exhibits no parametric assumptions at all, and only a weak restriction of additive errors in each stage.

The innovation of our work is in exploiting the additive structure of the model both in constructing the selection correction term in the first step and in non-parametric estimation of the outcome equation in the second. The selection correction term is shown to depend on the conditional mean of an indicator variable (which determines whether or not the dependent variable is observed), and is non-parametrically estimated using series approximations. The two-step estimator is then shown to achieve Stone's (1982) bound. This paper also discusses interesting linear functionals of the series estimate and derives the conditions under which the functionals are $\sqrt{n}$-consistent.

In the next section we review the classic censored selection problem, and present the sample selection model that is studied in this paper. In Section 3, we turn to a description of the two-step estimator, reviewing with it series estimation and presenting asymptotic properties of the estimator. Section 4 discusses functionals of the estimator, describes the construction of a consistent estimate of its asymptotic variance, and derives asymptotic normality results. We extend our analysis to a semi-parametric treatment of the model in Section 5. Following this, we present an application of our estimator, also reporting corresponding estimates from parametrically specified and semi-parametric models to facilitate comparison between the older and newer methods. Proofs follow.

2.2 Censored Sample Selection

Formal analysis of the classic (parametric) sample selection problem (e.g, Heckman, 1976) begins with the model

$$y_i = \delta_i[x_i'\beta_0 + \epsilon_i] \qquad i = 1, \ldots, n \tag{2.2.1}$$

where the dependent variable y is assumed to depend on regressors x, an unobservable error term ϵ, and the indicator δ which determines whether the dependent variable is censored or not, accordingly taking the value 0 or 1. The indicator variable in turn is assumed to depend on a binary variable d, a set of conditioning variables z and unobservable error terms η, through a standard binary response model:

$$d_i = \mathbf{1}\{z_i'\gamma_o + \eta_i > 0\} \tag{2.2.2}$$

where $\mathbf{1}(A)$ denotes the indicator function for the event A. Equation (2.2) may take on a latent variable interpretation if we assume that data on y is obtained only if another random variable v crosses a threshold, *i.e.*, y observed only if

$$\begin{aligned} v_i &> 0, \text{ where} \\ v_i &= z_i'\gamma_0 + \eta_i \end{aligned}$$

and y is unobservable when $v \leq 0$. We can then define

$$d_i = \begin{cases} 1, & \text{iff } v_i > 0 \\ 0, & \text{iff } v_i \leq 0 \end{cases}$$

yielding (2.2). Thus, the close relationship between the pair of equations (2.1) and (2.2) arises because $\delta_i = 1$ holds if and only if $d_i > 0$ is true (or, equivalently, if and only if $\eta_i > -z_i'\gamma_0$). Without making any assumptions about any corelations between the error pair (η_i, ϵ_i) however, least squares or instrumental variables estimation of (1) for the subset of observations with $\delta_i = 1$ is valid and yields consistent estimates of β_0. Sample selection, and consequently "selectivity bias", occurs when we assume that the distribution of ϵ_i is not independent of the distribution of η_i. Then, conditional on the vector z_i and the event that y_i is observed, equation (2.1) is misspecified unless a correction term is added to account for the non-random censoring. The selection correction term derived from (2.1) – (2.2) is well known to be

$$\begin{aligned} \mathbf{E}[\epsilon_i | z_i, \delta_i = 1] &= \mathbf{E}\{\epsilon_i | z_i, \eta_i > -z_i'\gamma_0\} \\ &= \theta(z_i'\gamma_0) \end{aligned}$$

for some function $\theta(\cdot)$, leaving as the correctly specified equation of interest

$$y_i = x_i'\beta_0 + \theta(z_i'\gamma_0) + \psi_i \tag{2.2.3}$$

where ψ_i is an error term that distributed independently of z_i and δ_i by construction. An alternative but equivalent representation of (2.3) which will be useful for the analysis in this paper is derived by noting the following (previously noted in Heckman and Robb (1990), and Ahn and Powell (1993)):

$$Pr(d_i = 1) = \mathbf{E}[d_i|z_i] = F_{-\eta}(z_i'\gamma_0) \tag{2.2.4}$$

implying, for an invertible cumulative distribution function F of $-\eta$,

$$F^{-1}\mathbf{E}[d|z] = z'\gamma_0 \tag{2.2.5}$$

so that we may express (2.3) as

$$y_i = x_i'\beta_o + \lambda(E(d_i|z_i)) + \psi_i \tag{2.2.6}$$

where $\lambda(\cdot)$ is the unknown composite transformation $\theta \circ F^{-1}$. For the estimation strategy considered in this paper the non-parametric formulation of (2.6) is the model of interest, highlighting that the appropriate correction involves an unknown function of the conditional mean of the *selection variable*, *d*. [1]

The analysis of (2.3) or (2.6) typically proceeds in accordance with the estimation technique chosen for the model. The popular parametric method follows Heckman's two-step strategy: (*i*) estimating the parameters of the selection rule (γ) to construct the parametric form of the correction term, and (*ii*) estimating the parameters of the primary regression model using instrumental variables or least squares estimation.[2] Three possible sources of specification error exist in this parametric estimation: mis-specification of the unknown joint distribution, mis-specification of the unknown regression function (or "single-index") in the selection equation and mis-specificaiton of

[1] In an semi-parametric study of sample selection, Ahn and Powell (1991) use the conditional mean of d in a similar correction term. The method outlined in their study, however, is unrelated to the approach taken here, instead resembling a partially linear model.

[2] For example, for a joint normal distribution of the error terms, as is commonly assumed, the explicit form of the correction term is shown to be proportional to the Inverse Mills' ratio.

the regression function in the outcome equation by *a priori* assuming linearity of the parameters β in x. Semi-parametric estimation of (2.3) or (2.6) avoids at least one of the sources of parametric error by leaving the joint distribution of the error terms unspecified. Some studies further weaken the possibility of parametric mis-specification by formulating a non-parametric selection rule, thereby deriving flexible correction terms that are devoid of parametric error (*e.g*, Powell (1987), Newey (1988), Powell and Ahn (1991)); these, however, leave open the possibility that the selection correction is substitued in a parametrically mis-specified outcome equation.

We reformulate the classic selection model to avoid every source of parametric error, by studying a model with no parametric assumptions and only one weak restriction that involves additivity of the error terms in each stage. Our focus is on the model[3]

$$\begin{aligned} y &= \mathbf{1}\{d=1\} \cdot (\, g_0(x) + \epsilon \,] \\ d &= \mathbf{1} \; (\, \pi_0(z) + \eta > 0 \,) \end{aligned} \tag{2.2.7}$$

with the following assumptions on the pair of errors:

$$\begin{aligned} E[\epsilon|z,\eta] &= E[\epsilon|\eta] \\ E[\eta|z] &= 0 \end{aligned} \tag{2.2.8}$$

$g_o(x)$ represents an unknown function of the (q x 1) vector of variables, x and $\pi_o(z)$ is the unknown transformation that relates the (q_1 x 1) vector z to the probability that data on y is observed. The conditional mean restrictions on the error terms match those commonly assumed in the literature on sample selection models; in particular, the correlation of (η,ϵ), which is the source of selectivity bias. Equation (2.7) represent a generalization of the common sample selection specification where it is typically assumed that $g_o(x)$ is linear in x and $\pi_o(z)$ is linear in z. Our more general formulation of the model leaves the correction term as $\mathbf{E}[\epsilon|\eta > -\pi(z)]$ which entails a non-trivial problem when neither the joint distribution is specified nor the form of π is known. This is the problem that we seek to address in our paper.

Additivity of the error terms forms an integral part of the model. For example, one could

[3]Since d and δ share a one-to-one relation, we formulate the model in terms of d alone, retaining this notation for the rest of the paper.

consider even more general forms of a selection model by specifying $y = \mathbf{1}\{d = 1\} \cdot [\, g_o(x), \epsilon \,]$ and similarly allowing for a multiplicative error in the selection rule. However, this specification greatly complicates the analysis without adding any extra insight into the non-parametric analysis of sample selection. Further, it is well known that imposing additivity leads to improved efficiency, improved convergence rates and lower asymptotic variances of estimators of the functionals of $g_o(x)$ (*e.g.*, Stone, 1985; Newey, Powell and Vella, 1997). We therefore restrict ourselves to the analysis of an additive non-parametric sample selection model, emphasizing that our results are especially useful for those applications in which the analyst expects the model to be specified as in (2.7).

We outline a two-step procedure to estimate the model. In the first step we employ a non-parametric regression of d on z in order to obtain an estimate of $\hat{d} = \hat{E}[d|z]$. The second step uses the estimated $\hat{d}$ for the correction term in a non-parametric regression of y on x and $\hat{d}$. While several non-parametric methods exist for this estimation, we focus on series estimation for both steps. We show that series estimators are particularly useful when the model is additive, as we have assumed.

2.3 Two-step Estimation

We begin by correctly specifying equation (7) under the assumed restrictions on the error terms. Using the notation $\theta_0(\pi(z)) = \mathbf{E}[\epsilon|\eta > -\pi_0(z)]$, evaluate, conditional on z, the expected value of y when it is observed:

$$\begin{aligned}
\mathbf{E}[y|z, d = 1] &= \mathbf{E}[g_o(x) + \epsilon|z, d = 1] && (2.3.1)\\
&= g_0(x) + \theta(\pi_0(z)) && (2.3.2)\\
&= g_0(x) + \lambda_0(E[d|z]) && (2.3.3)\\
&= h_0(w) && (2.3.4)
\end{aligned}$$

where $\lambda(\cdot) = \theta \cdot F_{-\eta}^{-1}(\cdot)$ and is the non-parametric correspondence of equation (2.6). Denote $w \equiv [x, e_d]$, where e_d represents $E[d|z]$. y is thus determined by an unknown function of x and an unknown function of e_d, and a non-parametric regression of y on $w \equiv \{x, e_d\}$ would suffice in uncovering the relation that we are interested in.[4] Since e_d is itself unknown, it is replaced by an

[4]Note that although $\lambda_0(E[d|z])$ is equivalently viewed as some unknown function, $\Psi_0(z)$, estimating the model in a single step as a function of (x, z) throws away information pertaining to d; further, $E[d|z]$ is estimated on a

estimate, $\hat{e}_d$, from a first step non-parametric regression of d on z. The estimation is therefore done in two steps: (i) the estimation of $\hat{e}_d$, followed by (ii) the estimation of $\hat{\mathbf{E}}[y|x,z]$.

The function of interest, $g_0(x)$ is one component of $h(w)$, and thus additional steps are required in order to isolate it from a non-parametric regression of y on $[x, e_d]$, which results only in an estimate of $h(w)$. This identification essentially relies on exclusion restrictions. One method of isolating $g_0(x)$ is to employ a partial means estimator (*e.g*, Newey, 1994) to integrate out the λ component from $\hat{h}(w)$; $g_0(x)$ is then identified up to some additive constant term. To be precise, let $\xi(e_d)$ denote the density of e_d. Then,

$$\int E[y|z, d = 1]\xi(e_d)de_d = g_0(x) + \int \lambda_0(e_d)\xi(e_d)de_d \tag{2.3.5}$$

(3.5) provides us with an estimate of $g(x)$ up to an additive constant (where the additive constant is $\int \lambda(e_d)de_d$.) Observe, however, that the partial means strategy to identifying $g(x)$ is valid if and only if integrating over e_d out does not result in integrating out $g(x)$ as well, *i.e.*, if and only if $g(x)$ can be held constant while e_d is changing. Since e_d is functionally related to z through the selection rule, this essentially requires that the set of variables that determine the probability of being in the selected sample include at least one variable that is not in x. Re-stated, the partial means approach to identifying $g(x)$ is valid only if the dimension of z is strictly larger than that of x, a condition similar to the standard order condition for instrumental variables estimation, and not unrelated to identification in parametric sample selection models.

The estimation strategy proposed here uses series estimation in each stage. The utility of considering a model additive in the errors is now apparent: First, without additivity, the correction term would involve estimating the marginal density of ϵ or equivalently, integrating out the x in order to obtain an expression for λ, adding an unnecessary complication. Second, with an additive model, series estimation in the second stage is simple to impose, involving a non-parametric regression of y on approximating functions of either x or $\hat{d}$ separately, but no combinations of the two. Further, as discussed earlier, this additivity has the attractive feature of improving efficiency and convergence rates of the series estimate as well as reducing the asymptotic variance of its functionals' (see Newey *et al*, 1997).

---

larger sample *ie*, the entire population, not just those for whom the outcome is observed, and thus, $E[d|z]$ must be computed on the whole sample.

We open our discussion of the two-step estimator by focussing on the first step. For series estimation of the selection rule, we first need to specify a $(L \times 1)$ sequence of functions that can approximate the underlying function $\pi(\cdot)$ by some linear combination of the functions, with the property that the approximation to π is arbitrarily close for large L. This underlying principle of series estimation is a simple restatement of the well-known Weierstrass theorem. Various series approximations exist, each with distinct virtues and each best suited to specific features of the data on hand. The choice of the approximating functions, r^L, and the dimension of L therefore depends on what the modeler may know *a priori* about the extent of outliers or the expected shape of the smoother, and whether the focus is on reducing multi-collinearity, improving convergence rates or in easing computation.

For some positive integer L, let $r^L(w) = (r_{1L}(w), \ldots, r_{LL}(w))'$ denote the sequence of approximating functions. Assume n observations, and let $d = [d_1, \ldots, d_n]'$. Then the non-parametric correction term we are interested in, $\hat{d}$, is obtained as follows:

$$\hat{\gamma} = \left\{ \sum_{i=1}^{N} r^L(z_i) r^L(z_i)' \right\}^{-1} \cdot \left\{ \sum_{i=1}^{N} r^L(z_i)' d_i \right\} \tag{2.3.6}$$

where $\hat{\gamma} \equiv (\hat{\gamma}_1, \ldots, \hat{\gamma}_L)'$ is the vector of least squares coefficients obtained from regressing d on $r^L(z)$. Since the primary regression model is given by $y = g_o(x) + \lambda(e_d) + \psi$, the second step may be estimated in a manner similar to that in step one, with appropriate changes to reflect that z is replaced by $w \equiv [x, \hat{e}_d]$ and that x and $\hat{e}_d$ enter the model separately, and additively.

Let $p^K(w) = (p_{1K}(w), \ldots, p_{KK}(w))'$ be the (1 x K) series of approximating functions for step two with the restriction that for any $k \in [1, K]$, $p_{kK}(w)$ is either a function of x or a function of d, but not any combination of the two; this step is the estimation correspondence of the fact that the model we study in equation (2.3) is additive in x and e_d. Before employing the estimated $\hat{e}_d$ in the non-parametric regression of the second stage, we recognize that the first step dependent variable is a probability which necessitates bounding the values of d. A convenient method of doing this is to use a trimming function which resrticts the domain of the data, considering only those values deemed plausible by the analyst, allowing the exclusion of large values of the data that appear to be outliers, and having the additional feature of facilitating the asymptotic results. Define

$$\tau(w) = \Pi_{j=1}^{2q} \mathbf{1}(a_j \leq x_j \leq b_j) \mathbf{1}(0 \leq d \leq 1) \tag{2.3.7}$$

where $q = \dim(x)$ and a_j and b_j are either pre-specified constants based on the analyst's knowledge of what values of x may represent outliers, or may themselves be estimated.

Let $\hat{w} = (x, \hat{e}_d')$, $\hat{p}^K = p^K(\hat{\tau}(\hat{w}))$ and $\hat{\tau} = \tau(\hat{w})$. The second step of the estimation is a regression of y on $p^K(w)$, taking into account the restriction imposed in (3.7), which gives us the series estimate of interest:

$$\hat{\beta} = \left\{ \sum_{i=1}^{N} p^K(\hat{w}_i) p^K(\hat{w}_i)' \right\}^{-1} \cdot \left\{ \sum_{i=1}^{N} p^K(\hat{w}_i)_i' \right\} \tag{2.3.8}$$

There is an important issue of identifying a constant term that arises in the estimation strategy just described. Let the first element of the series $p_{1K}(w)$ be equal to one; further, let the next g elements of the series (K_g) depend only on x and the remaining $K - K_g$ depend on the estimated $\hat{e}_d$. Consider, respectively, estimators of $g_0(\cdot)$ and $\lambda_0(\cdot)$

$$\hat{g}(x) = \hat{\alpha}_g + \sum_{i=2}^{K_g+1} \hat{\beta}_i * p_{iK}(x) \tag{2.3.9}$$

$$\hat{\lambda}(\hat{e}_d) = \hat{\alpha}_\lambda + \sum_{i=K_g+2}^{K} \hat{\beta}_i * p_{iK}(\hat{e}_d) \tag{2.3.10}$$

Clearly the estimate of $\hat{\beta}_1 = (\hat{\alpha}_\lambda + \hat{\alpha}_g)$, and the constant terms cannot be separately identified without additional restrictions. While in many applications, it is the local curvature of g as x changes that is of particular importance (in which case, identifying the constant may be unnecessary), in specific cases the *level* of g may be of specific interest. For example, we might be interested in predicting the exact number of participants in a social welfare program in order to estimate the costs of a program.

One approach to identifying the constant is to assume that the value of $\lambda(\cdot)$ is known for some $e_d = \bar{e}_d$. For example, in the classic wage-hours sample selection problem, we could assume that at the lowest observed value of z, the expected value of d is likely to approach zero; this accords with the standard parametric formulation of sample selection, where $\lambda(\cdot)$ is proportional to inverse of the Mills ratio, $\left[\frac{\phi(\cdot)}{\Phi(\cdot)}\right]$, which approaches zero at the tails; thus, one restriction could be to specify that for $z = z_{min}$ and $\bar{e}_d = E[d|z = z_{min}]$, $\lambda(\bar{e}_d) = 0$.

Then, α_λ is identified:

$$\overline{\lambda} - \sum_{i=K_g+2}^{K} \hat{\beta}_i * p_{iK}(\overline{e}_d)\,, \tag{2.3.11}$$

and α_g can be estimated as $\alpha_g = \hat{\beta}_1 - \alpha_\lambda$, yielding estimates of both of the components of $\hat{h}$ separately:

$$\hat{g}(x) = \hat{h}(x, \overline{e}_d) - \lambda(\overline{e}_d) \tag{2.3.12}$$

$$\hat{\lambda}(e_d) = \hat{h}(x, d) - \hat{h}(x, \overline{e}_d) + \lambda(\overline{e}_d) \tag{2.3.13}$$

We discuss this in more detail in section 4 when we study functionals of the regression estimate $\hat{h}$.

The two series estimators we consider in this paper are (1) power series and (2) smooth piecewise polynomials, or splines, with evenly spaced break-points. Power series provide especially good approprimations to smooth functions and are simple to compute but are adversely affected by outliers in the data. A typical power series is modelled as increasing powers of a single function such as

$$\hat{p}_{kK} = f^{k-1}. \qquad k = 1, \ldots, K; \qquad K = 1, 2, \ldots$$

for some function f. Note that the first term of the series is a constant. The function f that is used may be chosen according to the context of the model being studied, so for example, with no *a priori* knowledge about the shape of the function, a convenient choice may be to use $\hat{p}_{kK}(x) = x^{k-1}$, a standard polynomial series approximation. Newey (1988) suggests a number of appropriate choices in the context of sample selection; these include

$$\hat{p}_{kK}(x) = \mathcal{F}(\hat{\mu} + \hat{\sigma}(x)) \tag{2.3.14}$$

where $\mathcal{F}(\cdot) = \left[\frac{\phi(\cdot)}{\Phi(\cdot)}\right]$, $\phi(\cdot)$ is the standard normal density and $\Phi(\cdot)$ represents the standard normal distribution function; this example gives as its second term the inverse Mills' ratio, the correction term most commonly used in parametric estimation of the sample selection model. Another example may be to specify $\hat{p}_{kK}(x) = [1 - \Phi(\hat{\mu} + \hat{\sigma}x)]^{k-1}$: a series that could reduce the effect of outliers of x since it uses the bounded function $\Phi(\cdot)$.

Splines, or smooth piecewise polynomials with fixed joining points (or "knots") for the polynomial, are an alternate type of series whose distinct features are that they provide better approximations when outliers in the data are likely to be an issue and when the function being approximated is assumed to be discontinuous. The knots may be placed at any points where the analyst thinks the underlying function is changing rapidly and thus, applying knots requires some knowledge about the features of the data; an alternate approach is to place the knots at equidistant points. A typical mth degree spline for some univariate data z, with J known knot points, $j_1, \ldots, j_J$, may be expressed as

$$\mho_{jJ}(z) = \begin{cases} z^{j-1} & ,1 \leq j \leq m+1 \\ \mathbf{1}(z \geq j_i)(z - j_i)^m & ,1 \leq i \leq J \end{cases} \tag{2.3.15}$$

For multivariate z, the approximating functions are products of the the functions in the single variable case. Thus, for a cubic series with 10 knots, for example, the number of approximating functions grows from 14 in the univariate case, to 256 for a bivariate z. In the work here we restrict ourselves to evenly spaced knots (in the support of z, whose range must therefore be known) in order to facilitate the theoretical results. For exposition, consider the case where the support of z is on the interval $[-1, 1]$. Denote for a scalar c, $(c)_+ = \mathbf{1}(c > 0)\cdot)c$. Then, an mth degree spline with $J-1$ knots is defined to be a linear combination of:

$$\mho_{jJ}(z) = \begin{cases} z^{j-1} & ,1 \leq j \leq m+1 \\ ([z + 1 - 2(j-m-1)/J])_+^m & ,m+2 \leq j \leq m+J \end{cases} \tag{2.3.16}$$

Next, denote q_1 for the dimension of z, let $\mu = (\mu_1, \ldots, \mu_{q_1})$ denote a (1 x q_1) vector of non-negative integers, and let $(\mu(l))_{l=1}^{\infty}$ denote a sequence of such vectors. Then, for a set of vectors $\{\mu(l)\}$ with the restriction that $\mu_j(l) \leq (m + J)$ for each j and each l, the approximating series for z can be expressed as:

$$r_{kK}(z) = \Pi_{j=1}^{q_1} \mho_{\mu_j(k),J_j}(z_j), \ (k = 1, \ldots, K) \tag{2.3.17}$$

where J_j represents the number of knots for the jth component of z. Similarly, in the second step, spline approximations can be applied while imposing additivity by considering terms that depend only on $E[d|z]$ or on x but no combinations of the two.

2.3.1 Asymptotic Properties of the Two-step estimator

We now derive mean-square error and uniform convergence rates for our two-step estimator $\hat{h}$ as well as the function $\hat{g}$. We illustrate the degree to which the mean square convergence rate of the second step is affected because the first step has to be estimated, and discuss the relevance of imposing additivity for the convergence rate. The technical details needed in deriving the theorems of this section rely on various general results about convergence rates and asymptotic normality for series estimators due to Newey (1994b, 1995, 1996). Theorem 1 is the essential result of this section.

For a random matrix Y, denote $\|Y\|_v = \mathbf{E}[\|Y\|^v]^{1/v}$ $\forall v < \infty$, $\|Y\|_\infty$ denote the infimum of constants C such that $\mathcal{P}(\|Y\| < C) = 1$, and for a matrix D, let $\|D\| = [trace(D'D)]^{1/2}$.

Assumption 1 *The data,* $\{(y_i, z_i, d_i)\}$ *for* $i = 1, \ldots, n$, *is* i.i.d, *and* $Var(y|w)$ *and* $Var(d|z)$ *are bounded.*

Denote $\mathcal{W} = \{w : \tau(w) = 1\}$ for $w \equiv [x, e_d]$.

Assumption 2 *z is continuously distributed with density that is bounded away from zero on its support, and the support of z is a cartesian product of compact, connected intervals. The density of w is bounded away from zero on $\mathcal{W}$, and $\mathcal{W}$ is contained in the interior of the support of w.*

In an instrumental variables framework this assumption implies that the dimension of z is at least as large as that of x, a condition familiar from parametric instrumental variables identification. In the sample selection case, a similar result obtains: In order to identify g_o in the equation $y = g_o(x) + \lambda(e_d) + \psi$, x must be a strict subset of z so that at least one determinant of the probability that y is observed is not a determinant of y.

Assumption 3 *$\pi_o(z)$ is continuously differentiable of order s_1 on the support of z and $g_o(x)$ and $\lambda_o(\pi(z))$ are Lipschitz and continuously differentiable of order s on $\mathcal{W}$.*

The two derivative orders s_1 and s determine the rate of approximation for the conditional mean of interest, $h(w)$, by controlling the rate at which the series estimators (either power series or polynomial splines in our case) approximate g_o and λ. Under Assumption 3, the rate of approximation for $h(w)$ is $O(K^{-s/q})$. Recall that $q = \dim(x)$ and K represents the number of approximating functions for the second step.

The next assumption is a regularity condition needed to control the rates of growth of the number of terms in the approximating functions in the first step (L) and second step (K).

Assumption 4a *For power series, assume* $(K^3 + K^2L)\{(L/n)^{1/2} + L^{-s_1/q_1}\} \to 0$.

Assumption 4b *For splines,* $(K^2 + KL^{1/2})\{(L/n)^{1/2} + L^{-s_1/q_1}\} \to 0$.

Theorem 1

Under Assumptions (1) – (4)

$$(i) \quad \int \tau(w)[\hat{h}(w) - h_o(w)]^2 dF_o(w) = O_p(K/n + K^{-2s/q} + L/n + L^{-2s_1/q_1})$$

Further, for $c = 1/2$ *in the case of splines, and for c=1 in the case of power series*

$$(ii) \quad sup_{w\epsilon\mathcal{W}}|\hat{h}(w) - h_o(w)| = O_p(K^c[(K/n)^{1/2} + K^{-s/d} + (L/n)^{1/2} + L^{-s_1/q_1}]).$$

Proof: Appendix

There are two main implications of Theorem 1:
(*i*) the mean square convergence rate for the second step estimator, $\hat{h}(w)$, is the slower of the first-step optimal convergence rate and the second-step optimal rate when the first step does not have to be estimated, and this is precisely due to the exclusion of the interaction terms between x and $\hat{d}$.

As part (i) of Theorem 1 indicates, the convergence rate depends on the number of terms included in K and L; this term resembles the convergence rate obtained in Stone (1985), where it has been shown that optimality of the convergence rate depends on the manner in which K and L are chosen. Consider fixing the number of terms in K and L to be proportional, respectively, to $n^{q/(q+2s)}$ and $n^{q_1/(q_1+2s_1))}$; then, the convergence rate in part (i) reduces to

$$\int \tau(w)[\hat{h}(w) - h_o(w)]^2 dF_o(w) = O_p(max\{n^{-2s/(q+2s)}, n^{-2s_1/(q_1+2s_1)}\})$$

According to the discussion in Stone (1985), the optimal mean-square convergence rate in part (i) should be $n^{-2s/(q+2s)}$. Thus, implication (*i*) follows.

(*ii*) The second implication of Theorem 1 is that additivity in the regression function improves the convergence rate; without additivity a similar result would hold but q would be replaced by

$2q$ which will slow down the convergence rate of the estimator, and this is precisely due to the exclusion of the interaction terms between x and $\hat{d}$.[5]

Uniform convergence rates for $\hat{g}(x)$ follow from the conclusion of Theorem 1. In particular, for $\hat{h}(w) = \hat{g}(x) + \hat{\lambda}(\hat{e}_d)$, the restriction imposed on λ to identify the constant term fixes $\hat{\lambda}$, alowing an estimate of $\hat{g}(x)$ to be derived from $\hat{h}$ and thus yielding its convergence rate based on Theorem 1:

Theorem 2.1

For $\hat{g}(x) = \hat{h}(x, \overline{d})$ - $\overline{\lambda}$ and denoting W_1 as the coordinate projection of $\mathcal{W}$ on x

$$sup_{w_1 \epsilon W_1} |\hat{g}(x) - g_o(x)| = O_p(K^c[(K/n)^{1/2} + K^{-s/d} + (L/n)^{1/2} + L^{-s_1/q_1}]).$$

Proof: Appendix

The mean-square convergence rate of $\hat{g}(x)$ can also be derived from the conclusion in part (i) of Theorem 1 if the restriction on λ_0 used in identifying the constant term is a continuous functional of λ, such as, for example, a conditional expectation restriction; then, Theorem 1 holds with $\hat{h}$ and h_0 replaced by $\hat{g}$ and g_0. For the type of restriction considered here (a point restriction), however, the convergence rate of $\hat{g}$ does not simply follow, but the mean square convergence rate can be derived from a function $\underline{g}(x)$ which "partials out" the constant term.

Theorem 2.2

For $\overline{\tau} = E(\tau(w))$, define

$$\begin{aligned} \underline{g}_0(x) &= g_0(x) - \int \tau(w) g_0(x) dF_0(w) dw / \overline{\tau} \\ \underline{g}(x) &= \hat{g}(x) - \int \tau(w) \hat{g}(x) dF_0(w) dw / \overline{\tau} \end{aligned}$$

$$\text{Then, } \int \tau(w)[\underline{g}(x) - \underline{g}_o(x)]^2 dF_o(x) = O_p(K/n + K^{-2s/q} + L/n + L^{-2s_1/q_1})$$

Proof: Appendix

This completes our description of the two-step non-parametric sample selection estimator. An estimate of the conditional mean $E[y|w] = h_0(w)$ or of $g_0(x)$ may be of interest in various settings

[5]This would be true whenever the second step's convergence rate dominated in part (i). A more thorough discussion of a similar result is in Newey, Powell and Vella (1995).

that are frequently studied in applied work. These include studies of female hours of work (y) as a function of demographic and other variables as has been studied by Hausman (1984) and Mroz (1987), or analyses of *ex-post* wage income for participants of a social welfare program as examined in Meyer (1989)'s study of Unemployment Insurance (UI) or Yelowitz's (1994) analysis of marriage-rates and divorce-rates for AFDC participants, all of which involve sample selection issues. If the objective is to study other features of the model such as the marginal effect of the conditioning variables on the y, or the (local) concavity of convexity of the estimated conditional mean, we may do so by extending the above results. We discuss regression derivatives and other functionals in the following section.

2.4 Functionals of the two-step estimator

Often the regression estimate $\hat{h}(w)$ is of little intrinsic interest to the modeler, and the purpose of estimation is some numerical characteristic of $h(w)$. For example, one interesting functional that may be of interest in public policy or program evaluation is the issue of a price (tax or subsidy) change and the associated dead-weight loss; we would then focus our attention on some $\alpha = \int_a^b g_o(x)dx$. This functional has been studied by Hausman and Newey (1995) with an application to gasoline demand; in this application $g_0(x)$ is derived from the indirect utility function of an individual. Their framework can easily be extended to studying, for example, the deadweight loss associated with an increase in the marginal tax rates on a subset of the population, an application likely to involve sample selection issues.

Studying the implications of a proposed policy change is another example that applied work devotes much attention to. We might, for instance, be interested in the marginal change from a income tax reduction on labor supply participation such as that studied by Nada Eissa (1994), or the marginal effect on the birth-weight of babies following exogenous decreases in Medicaid funding of specific medical services for single mothers-to-be (which is the application studied in this paper). In the latter example, sample selection is an obvious problem since not all pregnancies result in births, making observed birth-weight of a newborn an outcome of a selected sample.

For such studies, the comparative statics of interest to us is given by a functional $a(h_0)$ which represents a linear map from h_0 to the reals, and encompasses a variety of functions that may be of interest; for example, point estimates of g. As previously discussed, such estimates can be derived

from the series estimate $h_0(w)$ when we restrict the value of λ_0 to be known as $\overline{\lambda}$ at some particular value of z. Then, the value of g at some particular $(\overline{x}, \overline{d})$ can be derived as:

$$g_0(\overline{x}) = a(h_0) \tag{2.4.1}$$

$$a(h_0) = h_0(\overline{x}, \overline{e}_d) - \overline{\lambda} \tag{2.4.2}$$

a natural estimate of which is obtained by "plugging in" the estimated value of $\hat{h}$ at the desired values $(\overline{x}, \overline{e}_d)$. We restrict our focus to linear functionals of $\hat{h}$ such as this, denoting the functional as $\hat{\theta} = a(\hat{h})$. Another linear functional is an average derivative estimator of $g_0(x)$ such as Stoker's (1986)

$$a(\hat{h}) = \int v(w)(\frac{\partial \hat{h}(w)}{\partial x})dw \tag{2.4.3}$$

where $v(w)$ is some weight function. The average derivative is similar to the functional in (4.2), but instead of measuring the strength of the relation between x and y at some particular x_i, it obtains an average effect over all the data points.

2.4.1 Asymptotics of Linear Functionals

We next state the conditions needed to derive the asymptotic behaviour of linear functionals of $\hat{h}$ and show that the functionals are asymptotically normal. The utility of studying the asymptotics is that with a consistent estimate of the asymptotic variance in hand, we may then do inference on the estimated functionals, $\hat{\theta}$. Our focus on linear functionals has the added advantage of fitting into a general class of two-step estimators for which Whitney Newey (1984) has established an explicit expression for the asymptotic variance. The class of estimators Newey studies is defined by the property that the first step first-order conditions are uncorrelated with the second step first-order conditions, a criterion satisfied here.[6] We show that $\sqrt{n}\hat{V}^{1/2}(\hat{\theta} - \theta_o) \xrightarrow{d} N(0, I)$, so that in large samples inference on $\hat{\theta}$ using the variance estimate $\hat{V}/n$ is valid. To characterize when $\sqrt{n}$-consistency holds, a further assumption needs to be made. In particular, $\sqrt{n}$-consistency is

[6]This holds since in the first step $\hat{d}$ is a function of z, and z is a conditioning variable in step two.

attained only if the following condition (the Reisz representation theorem) holds:

$$a(h) = E[\tau(w)\nu(w)h(w)] \tag{2.4.1.1}$$

such that $E[\tau(w)\|\nu(w)\|^2]$ is finite for some vector of functions $\nu(w)$. We will assume this condition to hold in deriving the asymptotic results for the functionals.

Validity of (4.1.1) implies that $\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow N(0, \overline{V})$ for a variance matrix $\overline{V}$. An exact expression for the variance in this case can be derived: denote $\mathcal{L}$ for $\{b(z)$: for a sequence α_L, such that $\{lim_{L\rightarrow\infty}E[\{b(z) - \tau^L(z)\alpha_L\}^2] = 0\}$. Further, let $\mathcal{K}$ denote $\tau(w)\nu(w)\delta h_0/\delta d'$, and let $\rho(z)$ be the matrix of projections of the elements of $\mathcal{K}$ on $\mathcal{L}$ in the space of random variables with finite variance. Then, if the reduced form (the equation $y = g_0(x) + \lambda(E[d|z]) + \psi$) is left unrestricted, $\mathcal{L}$ will typically comprise of all functions of z with finite variance, and $\rho(z)$ will be the expectation of $\mathcal{K}$ conditional on z. From this, the asymptotic variance in the $\sqrt{n}$-consistent case will follow as:

$$\overline{V} = E[\tau(w)\nu(w)\nu(w)'Var(y|w)] + E[\rho(z)Var(d|z)\rho(z)']$$

The variance of $\hat{\theta}$ in the more general case (where (4.1.1) does not hold and $\sqrt{n}$-consistency cannot be attained) is given below.

The specification of the selection rule has thus far been given a latent variable interpretation. An equivalent representation is to specify a non-parametric version of the linear probability model. In particular, note that the latent variable model gives $E[d|z] = F_{-\nu}(\pi(z))$. An alternate formulation would be to specify a version of the linear probability model: $d_i = k(z_i) + v_i$. But since this implies $E[d|z] = k(z)$, series estimation of either specification results in the same estimate of $E[d|z]$, *i.e.*, $k(\cdot) = F_{-\nu}(\pi(\cdot))$ and $v = d - E[d|z]$. Quite simply, in the non-parametric setting, a representation of the reduced form in either a latent variable framework, or in a non-parametric version of the linear probability model is equivalent; *i.e*, the latent variable representation is fully general. This equivalence is useful, since calculation of the asymptotic variance of the second-step estimator is greatly simplified using the non-parametric linear probability specification.

Define

$$\hat{Q} = \frac{\hat{P}'\hat{P}}{n}, \quad \hat{\Sigma} = \sum_i^n \frac{\hat{\tau}_i\hat{p}_i\hat{p}_i'[y_i - \hat{h}(\hat{w}_i)]^2}{n}$$

$\widehat{Q_1} = \mathbf{I}_s \otimes (\frac{R'R}{n})$, where the subscript 1 refers to step one and $\otimes$ denotes the Kronecker product.

$$\hat{\Sigma}_1 = \sum_i^n \frac{(\hat{\eta}_i'\hat{\eta}_i) \otimes (r_i r_i')}{n}, \quad \hat{H} = \sum_i^n \hat{\tau}_i [\frac{\partial \hat{h}(w_i)}{\partial \eta}]' \otimes \frac{\hat{p}_i r_i'}{n}$$

The variance of $\hat{\theta}$ is then simply A $\{Var(\hat{h})\}A'$, or,[7]

$$\hat{V} = A\{\hat{Q}^{-1}[\hat{\Sigma} + \hat{H}\hat{Q}^{-1}\hat{\Sigma}^{-1}\hat{H}']\hat{Q}^{-1}\}A' \tag{2.4.1.2}$$

We use this estimator of the variance in deriving the asymptotic normality of $\hat{\theta}$, to which we turn now. For this result, we need the regularity conditions listed below.

Assumption 5 *$\sigma^2 = Var(y|w)$ is bounded away from zero. $\mathbf{E}[\psi^4|w]$ is bounded, and $\mathbf{E}[\|\eta\|^4|w]$ is bounded. Also, $h_o(w)$ is twice continuously differentiable in d with bounded first and second derivatives.*

Assumption 6 *There exists $\nu(w)$ and β_K such that $\mathbf{E}[\tau(w)\|v(w)\|^2] < \infty$, $a(h_o) = \mathbf{E}[\tau(w)\nu(w)h_o(w)]$, $a(p_{kK}) = \mathbf{E}[\tau(w)\nu(w)p_{kK}(w)]$, $\mathbf{E}[\tau(w)\|\nu(w) - \beta_K p^K(w)\|^2]$ $\to 0$ as $K \to \infty$.*

Denote a non-negative integer by δ and $|h|_\delta = max_{|\mu|\leq\delta}\, sup_{w\epsilon\mathcal{W}}|\partial^\mu h(w)|$.

Assumption 7 *$a(h)$ is a scalar, $\|a(h)\| \leq |h|_\delta$ for some $\delta \geq 0$, and there exists β_K such that as $K \to \infty$, $a(p^{K\prime}\beta_K)$ is bounded away from zero while $\mathbf{E}[\{p^K(w)'\beta_K\}^2] \to 0$.*

Assumption 8 *$\sqrt{n}K^{-s/d} \to 0$ and $\sqrt{n}L^{-s_1/d_1} \to 0$; for power series $(K^9L + K^8L^2 + K^6L^3 + K^2L^6)/n \to 0$, and for splines $(K^5L + K^4L^2 + K^3L^3 + KL^4)/n \to 0$.*

Assumption 9 *One of these two conditions must hold: (a) x is univariate and if a spline is used it is at least a quadratic one, and $\sqrt{n}K^{1-s} \to 0$; (b) x is multi-variate, $p^K(w)$ is a power series, $h_o(w)$ is differentiable of all orders, there is a constant C with the absolute value of the jth derivative bounded above by $C(C)^j$, and $\sqrt{n}K^{-\epsilon} \to 0$ for some $\epsilon > 0$.*

[7]Note first, that given $\hat{h} = \sum_j \hat{\beta}_j * p_{jK}(x)$, $\hat{\theta} = a(\hat{h}) = A\hat{\beta}$, $A = (a(p_{1K})), \ldots, a(p_{KK})$ is simply a linear combination of the second-step least squares coefficients.

Denote $\zeta_\delta(K) = max_{|\mu|\leq\delta}\, sup_w \|\partial^\mu p^K(w)\|$.

Theorem 3

If Assumptions 1-3, 5, either 6 or 7, & 8 and 9 hold, then $\hat{\theta} = \theta_o + 0_p(\zeta(K)/\sqrt{n})$ *and*

$$\sqrt{n}\hat{V}^{1/2}(\hat{\theta} - \theta_o) \xrightarrow{d} N(0, I) \tag{2.4.1.3}$$

Further, if Assumption 6 is satisfied

$$\sqrt{n}(\hat{\theta} - \theta_o) \xrightarrow{d} N(0, \overline{V}), \qquad \hat{V} \xrightarrow{p} \overline{V} \tag{2.4.1.4}$$

Proof: Appendix

2.5 Semi-parametric Estimation

We now consider how we may partially parametrize our general formulation to consider semi-parametric estimation. Semi-parametric estimation is useful to consider since an analyst occasionally has *a priori* knowledge about the functional form of at least part of the model; such information can be incorporated into estimation and thereby ease the computational burden by reducing the number of variables in non-parametric estimation. Partially parametric models have been used in applications before (*e.g*, Hausman and Newey (1995)). The semi-parametric model we consider is the familiar partially linear model (*e.g*, Robinson (1988)), in which we assume that the known functional part of $g_o(x)$ is linear and additive:

$$y = g_0(x) + \epsilon \tag{2.5.1}$$

$$= x_1'\beta_0 + g_{20}(x_2) + \epsilon \tag{2.5.2}$$

The results from Section 3 extend quite simply to the semi-parametric model here by restricting the approximating functions in $p^K(w)$ to contain the the componenets of x_1 and functions (power series or splines) of elements in x_2. Similarly, a semi-parametric formulation of the selection equation could be incorporated into the analysis, specifying $\pi_0(z) = z_1\gamma_0 + \pi_{20}(z)$, and corresponding to this, using components of z_1 and functions of components of z_2 in the approximating series in $r^L(z)$.

Since the parameters in equation (5.2) are simply functionals of $h(w)$, $\sqrt{n}$-consistent estimates

of β may be obtained under the same conditions as those listed in Section 4. There, the primary condition needed for $\sqrt{n}$ consistency was the existence of a function $\nu(w)$ such that $a(h_0) = \mathrm{E}[\tau(w)\nu(w)h(w)]$, for any linear functional $a(h_0)$. An explicit expression for $\nu()$ can be derived here.

Define $u(w) = x_{10} - E[x_{10}|x_{20}, d]$, the residual term from projecting x_{10} on the space of (x_{20}, d). Assume x_{10} is not almost surely predictable by (x_{20}, d) so that $E[\tau(w)u(w)u(w)']$ is bounded away from zero. Then

$$\beta = E[E(u(w)u(w)')^{-1}u'h(w)] \tag{2.5.3}$$

and $\nu(w) = E[u(w)u(w)']^{-1}u(w)$. Thus, $\sqrt{n}$-consitency and asymptotic normality of $\hat{\beta}$ follows under the assumptions and results in Theorem 3.

2.6 Empirical Application

In order to demonstrate the practical purpose of the estimator introduced in this paper, we present an empirical application in which sample selection is an obvious issue. We study the effects of the provision of abortion services available to pregnant mothers on the distribution of birthweights of their newborns. The outcome is selected since not every pregnancy results in a birth.

There is a large body of work that studies the effects of available abortion services and state abortion laws on birth outcomes (*e.g*, Joyce 1987, Grossman and Jacobowitz 1981, Grossman and Joyce 1990). An important finding of these studies is that exactly those women who are least likely to have healthy babies are the same who are least likely to carry the fetus to term. Therefore, restricted abortion access is likely to decrease the average health status of the set of babies eventually born. Birthweight is one of the best indicators of a newborn's health, and is thus the outcome most often studied. Further, from a policy perspective, birthweight is a useful measure, since low birthweight (less than 2500 grams) babies experience higher incidence of infant mortality, and account for 57% of costly neo-natal intensive care (Schwartz 1989).

Selection arises in this study precisely because birthweight is observed only for the set of pregnanices carried to term; and, the woman who chooses to have an abortion cuts a remarkably different profile from a woman who does not (see Joyce and Grossman [8]). The model is specified

as:

$$\begin{aligned} \text{Birthweight}_i &= 1(d=1) \circ g(\mathbf{fc}_i, \mathbf{mc}_i, \mathbf{hcf}_i) + \epsilon_i \\ d_i &= 1\{\pi(\mathbf{mc}_i, \mathbf{hcf}_i, \mathrm{Z}_i) + \eta_i > 0\} \end{aligned} \tag{2.6.1}$$

where i denotes a particular birth, **fc** is a vector of fetal characteristics such as its sex, race and birth-order, **mc** denotes maternal characteristics some of which affect only birthweight (drinking/smoking behaviour, employment, race, age), some of which affect only the probability of delivery (education, marital status, religious affiliation), and some such as income which affect both. **hcf** denotes a vector of health care facilities including the number of abortion providers and community health centers, primary health care centres, the number of ob/gyns and local health departments offering subsidized care. "Restriction" is a dummy which takes the value 1 for states which restrict Medicaid funding of abortion.

An alternative specification considers semi-parametric estimation of equation (6.1). Non-linearities may be important for some co-variates such as income, age, the number of large hospitals (denoted $\mathbf{mc}_1$), and less so for co-variates such as the number of other siblings (denoted $\mathbf{mc}_2$). Further, several of the co-variates are dummies which we simply pull out of the function g and π:

$$\begin{aligned} \text{Birthweight}_i &= 1(d=1)\{g(\mathbf{mc}_{1i}, hc_{1i}) + \sum_{j=1}^{J} fc_{ji}\beta_j + \sum_{k=1}^{K} mc_{2ki}\gamma_k + \sum_{l=1}^{L} hc_{2i}\psi_l + \epsilon_i\} \\ d_i &= 1\{\pi(\mathbf{mc}_{1i}, \mathbf{hcf}_{2i} + \sum_{j=1}^{J} fc_{ji}\beta_j + \sum_{k=1}^{K} mc_{2ki}\gamma_k + \sum_{l=1}^{L} hc_{2i}\psi_l \\ &\quad + \text{Restriction}_i\,\delta) + \eta_i > 0\} \end{aligned} \tag{2.6.2}$$

The list of variables used in each of the specifications in (6.1) and (6.2) are presented in the tables along with the results. The data used in the study come from the National Longitudinal Survey of Youth (NLSY) which is not a representative sample for this study as it over-samples young, poor and minority women who are most likely to require Medicaid funding and whose babies are most likely to be of low birthweight. The sample consists of 6365 pregnant women, aged between 15 and 32 in the year of their pregnancy. Between 1980 and 1989, 88.1% of these women carried the fetus to term.

We report a variety of results to facilitate the comparison between (i) least squares estimation

without any correction for selection, (*ii*) the Heckman 2-step estimator with jointly normal (ϵ, η), (*iii*) semi-parametric estimators with series estimation of the selection rule and least squares estimation of the outcome equation $(y = x\beta + \lambda(e_d))$; and (*iv*) non-parametric estimation of both steps. For non-parametric estimation of the selection rule, two alternate specifications are employed. One uses power series which trims out predicted probabilities outside $[0, 1]$, and the other uses functions of the inverse of the Mills' Ratio for the basis functions. The number of functions are chosen by cross-validation of the data.

2.6.1 Results

The first two panels of table 2 report least squares estimates from linear models. In the first panel, there is no correction for sample selection, and in the second panel the correction term comes from a linear probability model. Simply correcting for sample selection magnifies the estimate on the "Abortion Providers" variable about twofold, from 3.511 to 6.820. We infer that selection is an important issue as the t-statistic on the selection correction term is significant at the 95% level. The third panel reports estimates from a Heckman two-step estimator. Using probit in the first step, as opposed to a linear probability model appears to have no differential effect, and thus the estimated coefficients on "Abortion Providers" in both the second and third panels are close, and the selection correction term using Heckman's is again significant at the 95% level.

The fourth and fifth panels of table 2 report results from semi-parametric estimation of the model. The estimated regressions are not as general as that in equation (6.2), since the outcome equation is linear in the covariates and non-parametric only in the selection correction. By cross-validation, second order terms in "Abortion Providers", and interactions of this with income, age and a dummy for low income are used in semi-parametric estimation of the selection rule. In the outcome equation, two powers of the selection correction are found to be appropriate from cross validation of the data.

The effect of considering semi-parametric estimation is to magnify the estimated coefficient on "Abortion Providers" relative to a model where there is no correction for selection, and shrink it somewhat relative to a model in which the selection correction term enters linearly. This result is obtained either when power series or functions of the Mills' ratio are used for the selection correction term. However, note that the selection correction terms using functions of the Mills ratio (in panel 5) are not significant. Furthermore, note that while the coefficient on "Abortion Providers" is

Table 1 : The Effects of Access to Medical, Prenatal and Abortion Services on Birthweight : Parametric (Heckman, linear probability models), Semi-parametric selection correction and Non-parametric selection correction

	Least squares: No correction for selection	Sample Selection Correction		Semi-parametric Selection Correction		Non-parametric Estimation in both steps
		Linear in both steps	Heckman two-step	Power series in $\lambda(\pi(w'\gamma))$	Functions of Mills Ratio: $\varphi[\pi(w'\gamma)]/\Phi(\pi(w'\gamma))$	
% Abortion Prv	3.511 (2.061)	6.820 (3.062)	6.793 (3.874)	5.58 (2.501)	5.586 (2.504)	31.75 (0.902)
Age	0.167 (1.857)	-0.021 (-0.158)	-0.051 (-0.55)	0.066 (0.482)	0.0657 (0.478)	-0.646 (-0.630)
Drink	0.330 (0.531)	0.246 (0.382)	0.406 (0.667)	0.0931 (0.146)	0.084 (0.133)	0.1100 (0.172)
Smoke	-6.39 (-9.934)	-6.23 (-9.225)	-5.99 (09.47)	-6.183 (-9.261)	-6.182 (-9.628)	-6.093 (-9.018)
AFQT	1.68 (3.015)	2.412 (3.665)	2.534 (4.411)	2.081 (3.003)	2.093 (3.029)	2.030 (2.846)
Other Clinics	-0.15 (-1.103)	0.024 (-1.181)	-0.261 (-1.73)	-0.143 (-0.695)	-0.150 (-0.726)	-0.1962 (-0.484)
Subsidized care	-0.391 (-1.097)	-0.288 (-0.539)	-0.765 (-2.00)	-0.075 (-0.136)	-0.0821 (-0.148)	0.0424 (0.075)
Income	0.352 (2.469)	0.274 (1.768)	0.342 (2.343)	1.843 (3.186)	0.289 (1.886)	1.840 (3.186)
SCT (linear)	--	19.98 (2.118)	-14.98(16.23)	9.282 (1.707)	116.76 (0.216)	9.207 (1.742)
SCT (square)	--	--	--	3.840 (0.0333)	-36.723 (1.283)	3.840 (0.060)
SCT (cubic)	--	--	--	--	--	-36.723 (-0.128)
SCT (fourth power)	--	--	--	--	--	-28.234 (-0.134)
$Income^2$	--	--	--	--	--	-0.112 (-1.990)
$Income^3$	--	--	--	--	--	0.0024 (1.942)
$\%AbortionProv^2$	--	--	--	--	--	-101.49 (-0.707)
$\%Abortion Prov^3$	--	--	--	--	--	167.93 (0.746)
% Ab. Pr * Income	--	--	--	--	--	-1.123 (-1.931)

Notes:
SCT: Selection Correction Term. In the linear case this is simply E(d | covariates in the selection rule).

affected, none of the other coefficients (that were statistically significant) in the first three panels appears to change noticeably. One interpretation is that the tight constraints in a linear selection rule are invalid and weakening those restrictions allows a more accurately estimated coefficient on "Abortion Providers".

Finally, we report results from non-parametric estimation of both steps. We treat the dummy variables as essentially parametric since they can simply be pulled out of the unknown function. Power series in both "Abortion Providers" and "Income" are used, along with their interactions. By cross-validation of the data I choose five functions of the co-variates and four in the selection correction term. The average derivative estimate of the effects of abortion provision on birthweights, evaluated at the means is very close to both the semi-parametric specifications at 5.76. The standard errors are significantly inflated in the fully non-parametric specification, presumably yielding to the curse of dimensionality. However, the non-parametric estimates are probably the most accurate (even if imprecise) and thus, the conclusion of this exercise is that by restricting access to abortion, states effectively increase the pool of newborns born with low birthweight.

2.7 Conclusion

The research effort on sample selection since Heckman (1974,1976) has been one of the most active areas of microeconometrics research, motivated partly by the wide variety of applications that selection arises in. Acknowledging some of the parametric limitations of Heckman's estimator, a new literature has begun to emerge with the intent of weakening the parametric assumptions to consider semi-parametric estimation of selection models. The existing set of estimators focus on non-parametric estimation of either the selection rule or the outcome equation, but not both together.

Our paper recognizes a limitation of this new literature, which only partly addresses the disadvantages embodied in using parametric assumptions in modelling sample selection. Specifically, we develop an estimator that represents the most general formulation to date for sample selection, requring no parametric assumptions on either the regression functions or the joint distribution of the pair of stochastic terms. The single restriction required for estimation is an additive error term in both stages, a restriction that is weak and we expect to hold in a wide variety of applications.

The proposed estimator is a simple two-step estimator that employs functions of the non-parametrically estimated propensity score from the first step in the second. We derive theorems for the mean-square and uniform convergence rates for the estimator, demonstrating its optimality in terms of achieving Stone's (1982) bound. We then show that interesting functionals of our estimator are $\sqrt{n}$-consistent, and asymptotically normal.

To test our estimator, we present an application which studies the effect of abortion services available to pregnant women on the distribution of the birthweights of their newborns (a selected outcome). We show that the proposed estimator is easily computable and yields non-ignorable differences from parametric models.

2.8 References

Ahn, Hyung-Taik and James L. Powell (1993). " Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism," Journal of Econometrics; v58 n1-2 July 1993, pp. 3-29.

Cavanagh, Chris and Robert Sherman (1993). "Rank Estimators for Monotone Index Models," BellCore Working Paper 84 May 1992.

Cosslett, Steve (1983). "Distribution-Free Estimator of a Regression Model with Sample selectivity," manuscript, Center for Econometrics and Decision Sciences, University of Florida, June 1985.

Currie, Janet; Nixon, Lucia A., and Cole (1996). "Restrictions on Medicaid Funding of Abortion," The Journal of Human Resources, Volume 31, Number 1, Winter 1996.

Eissa, Nada (1995). "Taxation and Labor Supply of Married Women: The Tax Reform Act of 1986 As a Natural Experiment," National Bureau of Economic Research Working Paper: 5023

Gronau, R (1973). "The Effects of Children on the Housewife's value of time," Joural of Political Economy, 81, s168-s199.

Grossman, Michael and Steven Jacobowitz (1981). "Variations in infant mortality rates among counties of the United States: The Roles of public policieis and programs," Demography 18(4), 695-713.

Grossman, Michael and Theodore Joyce (1990). "Unobservables, Pregnancy Resolutions, an Birth Weight Production Functions in New York City," Journal of Political Economy, 98(5), 983-1007.

Han, Aaron (1985). "Non-parametric Analysis of a Generalized Regression Model: The Maximum Rank Correlation Estimator," Journal of Econometrics; v35 n2/3 July 1987, pp. 303-16.

Hausman, J.A (1985). "The Econometrics of Nonlinear Budget Sets," Econometrica; v53 n6 November 1985, pp. 1255-82.

Hausman, J.A and Newey W.K (1995). "Nonparametric estimation of exact consumer surplus and deadweight loss," Econometrica, 63, 1445-1476.

Heckman, J.J (1974). "Shadow Prices, Market Wages, and Labor Supply," Econometrica, 42, 679-693.

Heckman, J.J (1976). "The Common Structure of Statistical Models of Trucation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models, " Annals of Economics and Social Measurement, 5, 475-492.

Ichimura, H (1987). "Estimation of single index models," MIT PhD.

Joyce, Theodore (1987). "The Impact of Induced Abortion on black and white birth outcomes in the United States," Demography, 24(2), 113-136.

Kane, Thomas J. and Staiger, Douglas (1996). "Teen Motherhood and Abortion Access," Quarterly Journal of Economics; v111 n2 May 1996, pp. 467-506.

Kyriazidou, E. (1997) . " Estimation of a Panel Data Sample Selection Model", Econometrica; v65 n6 November 1997, pp. 1335-64.

Maddala G.S (1994). "A Survey of the Literature on Selectivity Bias as It Pertains to Health Care Markets,"in *Econometric methods and applications.* Elgar; distributed in the U.S. by Ashgate, Brookfield, Vt., 1994, pp. 330-45.

Manski, Charles. 1975. "Maximum Score Estimation of the Stochastic Utility Model of Choice," Journal of Econometrics; v3 n3 Aug. 1975, pp. 205-28.

Meyer, Bruce (1989). "Unemployment Insurance and Unemployment Spells," Econometrica; v58 n4 July 1990, pp. 757-82.

Mroz, Thomas (1987). "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions," Econometrica; v55 n4 July 1987, pp. 765-99.

Newey, W.K (1984). "A Method of Moments Interpretation of Sequential Estimators," Economics Letters 14, 201-206.

Newey, W.K. (1988). "Two step series estimation of Sample Selection Models," MIT Department of Economics Working Paper.

Newey, W.K, Powell J.L, Vella, F (1997). "Nonparametric Estimation of Triangular Simultaneous Equations Models," MIT, Department of Economics.

Schwartz, Rachel (1989). "What Price Prematurity?," Family Planning Perspectives, 21(4), 170-174.

Stone C.J. (1982). "Optimal global rates of convergence for nonparametric regression." Annals of Statistics 10, 1040-1053.

Stone, C.J. (1985). "Additive regression and other nonparametric models." Annals of Statistics 13, 689-705.

2.9 Appendix A

Proofs of Theorem 1, Theorem 2.1 and Theorem 2.2 :

Follows directly from the proofs for Lemmas 4.1 and $A1$ in Newey, Powell and Vella (1997).

Proof of Theorem 3 :

The proof for this theorem is an application of the proof for Theorem 5.1 of Newey, Powell and Vella (1997).

Chapter 3

Organization of Health Care and Composition of Deliveries

3.1 Introduction

In recent years, much attention has been focussed on models of the demand-inducing behaviour of physicians. The underlying hypothesis of such models is as follows: physicians derive utility from income and leisure, and disutility from inducing demand for unnecessary services. The disutility may arise from ethical considerations, or from reputation effects which penalize physicians by lowering demand for their services in the future (Evans 1974, Dranove 1988, Pauly and McGuire 1991). A natural implication of the model is that when income is tailored to specific procedures, physicans will exploit their agency relation with patients to perform more remunerative procedures if the marginal benefit of a specific procedure outweighs the associated marginal costs. This basic implication of the induced-demand model has been explored in several empirical studies (*e.g*, Cromwell and Mitchell, 1986; Grytten *et al*, 1991; Gruber and Owings, 1996) which demonstrate that demand-inducing behaviour by physicians is widespread.

In this study, I focus on a specific medical service, that of the delivery of babies, to examine the county-wide composition between cesarean-section (c-section) and natural deliveries and its link to the growth in managed care.[1] There is widespread belief, and some empirical evidence, that a

[1] In this paper, "managed care" will primarily refer to Health Maintenance Organizations (HMOs). A more inclusive definition could include HMOs as well as the physicians, hospitals and other providers that HMOs contract with.

high fraction of c-section deliveries are the result of induced-demand by obstetricians/gynecologists (Stafford, 1980 ; Shiono *et al,* 1987; Dranove and Wehler, 1994; Gruber and Owings, 1996). In 1991, the average comprehensive fee for a cesarean was $7879 compared to the average cost of a natural childbirth at $4456, which was 40% lower. Of this, the physician's fee was $2,200 for cesareans versus $1492 for a natural, representing a 30% differential (Health Insurance Association of America, 1991). Furthermore, the average time required to perform a cesarean is shorter than that for a natural, allowing a physician higher utility by skewing deliveries in favour of c-sections until the marginal gain of higher income and a reduced workload is offset by the marginal cost of inducement.[2] Over the 1965 – 1985 period, cesarean deliveries rose from 5% of all deliveries to 25%, a five-fold increase that was arguably not all due to rapid changes in medical technology. Gruber and Owings (1996) note that in the same time-period, fertility levels in the U.S. declined by 13.5%, representing a negative income shock to ob/gyns, and conclude that this exogenous source of income-pressure led to the substitution towards the more highly reimbursed c-sections.

This paper argues that the inducement-effect of declining fertility levels on c-section deliveries may be offset by a corresponding reverse effect on c-section rates due to increased managed care activity, with the net result of decreasing the observed rate of c-section deliveries. I take as my starting point a simple time-series correlation which suggests that this might in fact be the case (Figure 1): after unabated growth in the number of cesareans from the mid 1960s until the late 1980s, cesareans have for the first time fallen gradually between 1990-1994 (Centre for Disease Control, 1994); this is particularly striking when we take into account that fertility levels have continued to fall in the 1990-1994 period, contradicting the basic premise of the induced-demand hypothesis. Concurrently, HMOs grew from serving 8% of the health insurance industry in 1986, to 18.9% in 1988, and 25.3% in 1991 (HIAA, 1991). I argue that the surge in managed care activity has altered both, the potential for inducing demand for c-sections by non-managed care physicians, as well as the financial incentives and practice environment facing managed care physicians, resulting in lower observed c-section rates.

Apart from the ethical issues that arise, induced demand for services which produce the same

[2]Another factor that may be of importance in deciding between a cesarean versus a vaginal delivery is that cesareans can be scheduled. In fact, raw means in my data shows that amongst cesarean deliveries, 13% more babies were born on Tuesday than on any other day of the week; doctor's day off has traditionally been Wednesday. Alternatively, ob/gyns may prefer not to perform surgeries on Monday immediately after the weekend. This suggests a non-pecuniary behavioural motive for c-sections.

outcome as less costly procedures is an important issue in the ongoing national debate on rising health care costs. If specific procedures can be identified as those that arise largely due to inducement, policies that structure financial incentives for using more cost-effective substitutes could be designed to reduce national medical expenditures. For example, if greater managed care acitivity in a market (measured as the market share of HMOs in a health care service area) results in altering the composition of deliveries from the more highly reimbursed cesarean section towards natural childbirths, policies that favour the growth of managed care will result in lowering health care costs. In recent years, policies to re-structure the reimbursement of deliveries and re-organize the practice environment for obstetrics have been widely discussed (Myers and Gleicher, 1988). The evaluation of these policies rests on tangible evidence supporting such reform; this paper provides such evidence.

From a policy perspective, targeting deliveries in an effort to control health care costs is likely to be more successful than targeting other practices which are costly and have high incidence (for example, coronary bypass or hysterectomies). Induced demand can take place at two levels: inducement for the treatment, and conditional on requring a treatment, induced demand for a specific service. For a policy to effectively address inducement, it must be possible to distinguish inducement from necessity in the former, and there must exist close substitutes with very different reimbursement schedules for the latter. In the case of deliveries, the issue of whether the treatment (either a natural or a c-section) itself has been induced, does not arise, since ob/gyn and patient information of the symptoms requiring treatment is symmetric. Further, deliveries comprise of two well-defined services that are close substitutes and produce the same outcome. Insofar as there exists a reimbursement differential for the two substitute services, one service must favour inducement more than the other, and this service can be identified simply by comparing reimbursements.[3]

There are at least three principal mechanisms by which HMOs are known to affect the nature of health care services in a market [Newhouse (1984), Luft et al (1986), Feldman and Dowd (1986), Welch (1994), Baker and Corts (1996)]. First, the underlying principle of HMOs is *preventive* care, which leads to early detection and prognosis, and thus, potentially avoids costly surgical treatment. In 1991 HMO coverage of pre-natal and well-baby care exceeded coverage by traditional insurance,

[3]Contrast this with, for example, a coronary bypass where asymmetric information regarding the necessity of treatment is severe, making it difficult to ascertain if demand has been induced. Further, *conditional* on requiring treatment, there is no issue of inducement as no substitute service exists.

with 98% of HMOs offering well-baby care compared to 39% of conventional insurance (Source Book, 1991); there is corresponding evidence that pre-natal interventions result in better peri-natal outcomes, decreasing the need for costly neo-natal intensive care (Joyce et al 1988, Wise et al 1988). Increased HMO activity could therefore contribute in non-trivial ways to the reduction of costly neo-natal care. This could arise because HMO physicians encourage the use of preventive services, and also from HMO enrollees who perhaps self-select into HMOs because of unobserved tastes for preventive care. Thus, by altering the demand for the typical bundle of medical services, HMOs may affect the health care system by reducing the frequency of particular treatments, as well as the availability of specific procedures in the market.[4]

Second, HMOs play the dual role of insuring as well as purchasing (*i.e.*, specifying the hospitals/physicans where health care may be obtained under the terms of the contract), and thus have strong incentives in price-shopping for low-cost physicians and providers who practice conservative medicine. HMOs not only charge lower deductibles than traditional insurers, but typically set co-payment rates to zero, thus bearing much of the marginal cost of each additional service provided. By contrast, traditional insurers pay little attention to providers and medical services, instead, charging higher premiums to offset higher bills. Thus, increased enrollment of the population in HMOs might lead to changes in provider behaviour, fostering the spread of conservative practices and lower costs amongst phsyicians and hospitals that want to compete for HMO customers. For example, an ACOG report (1996) reveals that HMOs have become an important source of customers for ob/gyns, with 80.1% of ob/gyns participating in salaried managed care positions in 1994, up from 67.2% in 1991, even though fee-for-service ob/gyns earned up to 25% higher incomes in 1991.

Another mechanism by which HMOs may affect the nature of health care in a market is by altering the financial incentives associated with physican services. One of the defining characteristics of managed care is that reimbursement for services is often capitated or, physicians are paid fixed salaries; both mechanisms are intended to reduce the incentives associated with performing medically unnecessary procedures. Further, HMOs may design incentive schemes to reward phys-

[4]For another classic preventive medical service, take mammographies. In 1991, 96% of HMOs covered mammographies, while only 57% of traditional insurance companies did (HIAA, 1991). It is reported that mammographies reduce the risk of breast cancer by upto a third suggesting that growth in HMOs can affect the net spending on cancer therapy (National Cancer Institute, 1993). For instance, Baker and Brown (1997) find that increased HMO activity results in the adoption of more mammography facilities as well as higher usage.

icans for adopting practice styles that match their preferences for conservative practices. Clearly, HMOs are likely to be more successful in severing the link between induced demand and income for those services where (1) there is symmetric knowledge about symptoms favouring treatment and (2) where services producing the same outcome can be reimbursed equally. Deliveries are one such service.

Finally, another channel by which the growth in HMOs are thought to influence market-wide activity in health care is by influencing non-managed care physicians to adopt the services offered by managed care (see Baker and Brown, 1997). This line of argument is drawn from models of physican learning which suggest that physicians tend to mimic the practices and behaviour of other physicians in the market (*e.g*, Phelps, 1992).[5] Since HMOs favour physicians whose practice patterns are aligned with their conservative, low-cost preferences, the growth in HMOs in an health care market can alter the ratio of managed care to non-managed care physicians, and encourage the diffusion of those practices preferred by managed care organizations. Much of the existing empirical evidence on the effects of increased HMO activity documents such "spillover" phenomena. For example, traditional insurers are known to respond to growing managed care activity by imposing stricter guidelines on hospital use, increasing utilization review [Dowd (1988), Frank and Welch (1985)], resulting in decreased inpatient stays [McLaughlin (1987, 1988)].

Focussing on the effect of managed care on the composition of deliveries is useful, because the cost savings from a compositional change will be substantial. Obstetrics/gynecologic operations are the third most frequent procedure performed in the United States, with cesarean sections being performed at the rate of 388.4 per 100,000 people in 1991 (Source Book of Health Insurance, 1991). One study suggests that in 1991, over half the cesareans performed were unnecessary, resulting in 1.1 million extra inpatient days at a cost of over 1 billion (Public Health Citizen's Research, 1997). Consequently, U.S. government health officials have set a target of reducing the c-section rate to 15% by the year 2000. If HMOs do have a market-altering effect of reducing the number of unnecessary cesarean sections, they may induce large cost savings and help curb the rate of growth in health care costs.

In the next section I discuss in some detail the mechanisms by which HMO affiliation may influence physicians' choices in the mode of delivery, and how changes in HMO market-shares can

[5]One reason this might arise is that medical malpractice laws often use "standard" local practices as a benchmark against which to judge negligence suits.

result in altering the c-section rate, using a model of induced-demand. I then turn to an econometric specification of the model to be estimated, testing the main implication of the theoretical model: as HMO market shares rise, controlling for fertility levels, are c-section rates decreasing in HMO market share? An important implication of the theoretical model is that there will be differential effects on c-section rates depending on where along $[0, 1]$ HMO penetration rates fall. Specifically, the model predicts that c-sections are likely to be (weakly) rising in HMO penetration rates up to a critical penetration, after which the cesarean rate will fall with increasing HMO penetration.

A distinguishing feature of the current work is in minimizing specification error in estimation by using very general *non-parametric* methods that allow for an unspecified non-linear relation between c-section rates and HMO market shares. Using series estimation, I find that a penetration rate of 20% appears to be pivotal, beyond which every ten percent increase in HMO market-shares results in an average decrease of 1.8 percent in cesarean delivery rates; a sizable effect that translates into approximately $11 million in net savings at 1991-1992 reimbursement levels and c-section rates. The effect is most pronounced for HMO penetration between 35% and 45%, and diminishes as penetration rates approach 50%. The results establish that any linear modelling techniques (even those that correct for the endogeneity) highly understate the HMO effect, and completely ignore important non-linearities. By contrast, the methodology offered in this paper is robust to a wide variety of specification errors, simple to implement and simple to draw inference from.

This paper uses a rich data set containing information on all births in the United States in 1991 and 1992, collected by the National Centre for Health Statistics; this information is aggregated up to the county-level and used along with county-level information on HMO market-shares, demographic controls, county health care facilities, county business patterns and state health legislation acts. An extension of the study determines whether the changes in c-section rates come with associated changes in the quality of deliveries, say, because HMOs effectively induce demand for naturals. I use a variety of measures of quality, ranging from neo-natal infant mortality rates, to the number of documented complications at birth and Apgar Scores. I find no effect of HMO activity on quality. From the perspective of health care reform, therefore, increased managed care activity is an improvement in the sense of Pareto, decreasing costs without any obvious adverse effects.

3.2 HMOs and C-Sections: Model and Links

3.2.1 A Model of Induced Demand, Market-Shares and C-section Rates

Begin with the canonical representation of physician utility:

$$\begin{aligned} U &= U(Y, L, I) \\ U_Y, U_L > 0, U_I < 0 \qquad & U_{YY}, U_{LL}, U_{II} < 0, \end{aligned} \tag{2.6.1}$$

where Y represents physician income, L represents leisure and I is the level of inducement. $U_I < 0$ arises from our assumption that, as professionals bound by a code of ethics, physicians derive disutility from exploiting their agency relation to induce demand. Assume additively separable preferences. The model below builds on the early work of Newhouse (1970) and Evans (1974), extending it in several dimensions.

Let m denote Managed Care and, t denote "traditional", by which we will refer to ob/gyns paid by conventional indemnity insurance. The following model of induced demand studies t and m ob/gyns separately, identifying the relative magnitudes of their response to a change in the stock of births in their market. Such a change is exogenous, and represents a negative income shock, as noted by Gruber and Owings (1996). Further, the model determines how changing market-shares of managed care versus non-managed care affects the c-section delivery rate. Begin with non-managed care physicans:

Traditional:

Let

$$\begin{aligned} U_t &= U(Y^t, L^t, I^t) \\ Y^t &= Y_n N_t + Y_c C_t \end{aligned} \tag{2.6.2}$$

where N is natural childbirths, C is cesareans, Y_c is ob/gyn net revenue for a cesarean, Y_n represents ob/gyn net revenue for a natural, with $Y_c - Y_n > 0$ as is well known.

$$L^t = \overline{L} - \tau_n N_t - \tau_c C_t \tag{2.6.3}$$

where τ_n is the time per unit spent on a natural childbirth, τ_c is the time per unit spent on a cesarean, and $\overline{L}$ is the stock of leisure time. The average time for a natural childbirth is substantially greater than that for a cesarean; hence, $\tau_n - \tau_c > 0$. Let $\overline{B}$ represent all births, and α represent the market-share of traditional insurance. Then,

$$\begin{aligned} B_t &= \overline{B}\alpha, \\ C_t &= \phi(i)B_t \\ N_t &= (1-\phi(i))B_t \end{aligned} \tag{2.6.4}$$

where i will represent 'inducement per birth', or, the probability that a physician will induce demand for a c-section for a given birth; it follows that I, the total level of inducement is, Bi, and $\phi(i)$ is the c-section delivery rate with $\phi'(\cdot) > 0$. We further assume that $\phi''(i) = 0$. Since some fraction of births are appropriately diagnosed as requiring c-sections, assume $\phi(0) > 0$. The ob/gyn solves:

$$\max_i \; U(Y^t, L^t, I^t), \tag{2.6.5}$$

subject to the constraints in equations (2) – (4). The first-order condition is: (suppressing the t superscript)

$$U_Y(\cdot)[B\phi'(i)\Delta Y] + U_L(\cdot)[B\phi'(i)(\Delta\tau))] + U_I(\cdot)B = 0 \tag{2.6.6}$$

where ΔY represents $(Y_c - Y_n)$, and $\Delta\tau$ represents $(\tau_n - \tau_c)$. Comparing this f.o.c. to the one derived next for managed care ob/gyns will highlight clearly the relative inducement effects. The comparative statics follow from differentiating the f.o.c:

$$\frac{\partial i}{\partial B_t} = -\frac{\phi'(i)[\Delta Y U_{YY}(\cdot)(\Delta Y\phi(i) + Y_n) + \Delta\tau U_{LL}(\cdot)(\Delta\tau\phi(i) - \tau_n)] + U_{II}(\cdot)\, i]}{B_t[\phi'(i)^2(\Delta Y)^2 U_{YY}(\cdot) + \phi'(i)^2(\Delta\tau)^2 U_{LL}(\cdot) + U_{II}(\cdot)]} < 0 \tag{2.6.7}$$

Signing the expression is simple since every term apart from U_{II}, U_{YY} and U_{LL} is positive. Simply, (7) states that an exogenous decline in a source of income leads to more inducement. While the above analysis focuses on a decline in the stock of births, other exogenous shocks such as a mandated reduction in the c-section fee, yield the same comparative statics results. Consider next the managed care ob/gyn:

Managed Care:

Let

$$U = U^m(Y^m, L^m, I^m)$$
$$Y^m = \overline{Y}(N_m + C_m) \tag{2.6.8}$$

where $\overline{Y}$ is the capitated payment for a delivery *irrespective* of the specific mode of delivery chosen. Let

$$\begin{aligned} L_m &= \overline{L} - \tau_n N - \tau_c C \\ B_m &= \overline{B}(1-\alpha), \\ C_m &= \phi(i)B_m, \\ N_m &= (1-\phi(i))B_m \end{aligned} \tag{2.6.9}$$

Maximizing utility subject to the income, leisure and births constraints yields the first-order condition:

$$U_L(\cdot)[B\phi'(i)\Delta\tau] + U_I(\cdot)B = 0 \tag{2.6.10}$$

Let i^{*m} be the implicit solution to (10) and to facilitate comparison, assume equal market shares. Substituting i^{*m} into (6), we see the immediate implication of capitated payments: there is no income-related inducement effect since $\Delta Y = 0$. At the optimum the managed care ob/gyn trades off the disutility of inducement with the utility of increased leisure from unnecessarily prescribing a cesarean; by contrast, for the non-managed care ob/gyn, at a given level of inducement the marginal gains include a pecuniary as well as a time component. (This is clearly visible in equation (6): $B\phi'(i)$ represents the additional c-section births, each of which reimburses the non-managed care physician ΔY; a term absent in the managed care ob/gyn's optimization).[6] Thus, if $U_{YY}, U_{II}, U_{LL} < 0$

[6]The leisure constraint makes little substantive difference, only affecting the scale of the results. Hence, we retain it to prevent corner solutions: with no leisure-related constraint, the managed care ob/gyn would pick $i^{*m} = 0$.

globally, at the optimums:

$$i^{m*} < i^{t*} \Rightarrow \phi(i^{m*}) < \phi(i^{t*}) \tag{2.6.11}$$

or, the cesarean rate is strictly lower for managed care.

Evaluating the change in inducement for an exogenous change in births,

$$\frac{\partial i}{\partial B_m} = -\frac{\Delta\tau U_{LL}(\cdot)(\Delta\tau\phi(i) - \tau_n)] + U_{II}(\cdot)\, i}{B_m[\phi'(i)^2(\Delta\tau)^2 U_{LL}(\cdot) + U_{II}(\cdot)]} < 0 \tag{2.6.12}$$

Recalling that α denotes market-share for traditional insurance, equations (7) and (12) imply that:

$$\begin{aligned} &\frac{\partial i^m}{\partial \alpha} > 0, \quad \frac{\partial i^t}{\partial \alpha} < 0, \\ &\Rightarrow \frac{\partial (i^t - i^m)}{\partial \alpha} < 0 \end{aligned} \tag{2.6.13}$$

The next comparative statics of interest is the question of what happens to the c-section rate as managed care activity begins to grow. Let $\overline{C}$ be the total number of births, and $\bar{c} = \overline{C}/\overline{B}$ denote the c-section rate:

$$\bar{c} = \phi(i^t)\ \alpha + \phi(i^m)\,(1 - \alpha) \tag{2.6.14}$$

Differentiating,

$$\frac{\partial \bar{c}}{\partial \alpha} = \frac{\partial i^m}{\partial \alpha} + \alpha\left(\frac{\partial i^t}{\partial \alpha} - \frac{\partial i^m}{\partial \alpha}\right) + (\phi(i^t) - \phi(i^m)) \tag{2.6.15}$$

which cannot be signed *a priori*. The first term is positive, the second term ($\alpha\,\frac{\partial [i^t - i^m]}{\partial \alpha}$) is negative as shown in equation (13), and the third term is likely to be positive for some range of values of α around 0.5, but is of indeterminate sign in general. Thus, determining the sign and magnitude of the comparative statics in (15) is an empirical question that is the focus of the analysis in this paper.

I hypothesize that at high values of α, the non-managed care ob/gyn faces a large future income stream via the large stock of births covered by traditional insurance, and thus has weak incentives to induce demand. Consequently, for small local changes in α, the c-section rate might well be

falling in α (the middle term is being weighted heavily). As HMO market shares begin to rise, non-managed care physicians respond to their declining income stream by shifting births to c-sections, and the inducement rate i^t becomes larger than i^m almost surely. As a result, for local changes in α:

(i) the inducement by HMO ob/gyns falls; and,

(ii) fewer births are covered by non-HMOs so that even if this fraction of births are more likely to be delivered by c-section, there are fewer *total* c-sections, with the net result of decreasing the c-section rate. Thus, c-section rates are non-linear, say quadratic, in traditional's market-share.

Note that the response of c-section rates to fertility rates is unambiguously negative:

$$\frac{\partial \overline{c}}{\partial \overline{B}} = (1-\alpha)\frac{\partial i^m}{\partial \overline{B}} + \alpha \frac{\partial i^t}{\partial \overline{B}} < 0 \tag{2.6.16}$$

Equation (16) is the basis for previous studies to examine the response of c-section rates to declining fertility levels. Without controlling for contemporaneous changes in HMO market-shares, however, there is no *a priori* reason to believe that the cumulative effect of falling fertility rates and increasing managed care market-shares will be to increase c-section rates: the 1990-1994 decline in fertility rates and the corresponding fall in c-section rates is testimony to this. The hypotheses are two:

Hypothesis 1 : up to a critical penetration rate, α^*, as managed care activity grows (*i.e,* α falls) the c-section rate will fall:

$$\frac{\partial \overline{c}}{\partial \alpha} < 0, \qquad \forall \alpha > \alpha^* \tag{2.6.17}$$

Hypothesis 2 : Beyond the critical penetration rate, increased penetration of HMOs will result in decreasing the c-section rate. In particular, this response of c-section rates to managed care growth will swamp out the opposing effect on c-section rates of declining fertility levels:

$$\frac{\partial \overline{c}}{\partial \alpha} > 0 \qquad \forall \alpha < \alpha^* \tag{2.6.18}$$

In the empirical section I evaluate the hypotheses by examining the variation in county c-section rates due to variations in HMO penetration, holding county fertility rates and other demographics fixed.

3.2.2 Cesarean Sections

I now illustrate the specific links between observed c-section rates and HMO activity in the context of the model previously developed. These links can be divided into supply shifters (related to increased number of HMO ob/gyns and their low incentives in performing c-sections), and demand shifters such as increased use of pre-natal diagnostic services and adverse selection in non-HMOs. Some understanding of the leading causes of c-sections, the typical risk set, and variations in c-sections by diagnosis is useful in establishing the links between managed care activity and the incidence of cesareans. Prior to discussing the links, I briefly discuss the trends in cesareans, and the main diagnoses which favour a c-section over a natural childbirth.

Trends

Figure 2 maps the total and the primary c-section rates from 1970 to 1994. Of interest in this study are the rates between 1989 and 1994, a period over which cesarean usage gradually dropped for the first time since its widespread introduction in the early 1960s. The data reveal that between 1989 and 1992 the national c-section rate fell 5 percentage points from 22.8% to 21.9% (MVSR, 1994). The graphs do not mask a lot of variation over age groups or states. In 1991-1992, which is the time-frame considered in this study, c-section rates fell almost universally across all counties and states in the data set. The mean increase in those counties where the rate rose was 0.015. Regarding variation by age, in 1992, c-section rates were higher for mothers in the 40-45 age group (31.5%) than for teenaged mothers (15.0%), but both were lower than their respective rates in 1989.

It is widely believed that the diffusion in c-sections has been spurred by contemporaneous advances in medical technology. The two most prevalent obstetric procedures are *electronic fetal monitoring* (EFM) and *ultrasound,* which promote better deliveries by detecting abnormalities and other pre-existing conditions, but might thereby also increase the incidence of cesareans. In 1992, the use of EFM increased for the fifth year in a row, and was used by 80% of pregnant mothers, while ultrasound usage increased to 61% (Vital Statistics Report, 1994). There is some debate,

and no consensus, regarding the associated benefits of the rapid rise in c-sections from 1965 - 1988 (Finkler and Wirtschafter 1993, Mariskind 1979). Certainly, birth outcomes improved in terms of neo-natal mortality rates, birth injury rates and birthweight in tandem with the increased use of super. or technology (William and Hawes 1979), but over this same time period, other technological advances such as EFM, ultrasound and intensive care decreased the probability of poor outcomes (Vital Statistics Report, 1994).[7]

The other trend which appears to have tracked rising c-section rates very closely is that of malpractice premiums (HIAA, 1991). In 1991 the average malpractice premium for ob/gyns was approximately 9 times as high as that for any other group in the medical profession (AMA, 1991). This rise parallels the rapid rise in EFM and other obstetric procedures, suggesting that some of the increase in c-sections may have been to avoid malpractice suits. Some evidence of such 'defensive' medicine is offered in Tussing and Wojtowycz (1991).

Symptoms

Four symptoms account for over 80% of cesareans performed in the U.S.[8]. A leading cause of c-sections is repeat provider behaviour, *i.e.*, the observed phenomenon that if a previous birth was a cesarean, the likelihood of every subsequent birth being delivered by cesarean-section is extremely high. In 1994, 82.5% of women with previous cesareans had their next birth by cesarean delivery (NCHS, 1994). The most commonly cited medical explanation for this behaviour is that women with previous cesareans are at high risk of rupturing the original incision if the subsequent birth is natural (Guide to Pregnancy and Childbirth, 1996).

The second most frequent diagnosis is a *breech* or abnormal presentation, where the fetus is positioned in a way that largely precludes natural delivery; for example, transversally across the womb. A third cause is related to fetal health (respiratorial, congenital, low birthweight) and goes under the rubric of *fetal distress* and, a fourth is *dystocia* or, non-progressive labor, which is closely related to maternal health and age. Both fetal distress and dystocia are decidedly subjective diagnoses relative to the first two and may thus depend on non-medical factors such as time-involved

[7]In fact, while cesareans are known to save lives when they are deemed necessary, they increase risks to both mother and child when induced. Since increased HMO activity is associated with decreased inducement (as developed in the theoretical model), an extension of this study examines how the recent decrease in c-section rates has affected birth outcomes. There appears to be no relation between the two variables for three different measures of birth outcomes.

[8]This rate has varied very slightly; in my data sets they account for approximately 85% which compare with those found by others (Tussing and Wojtowycz 1991, McCloskey et al 1994)

and remuneration. I argue next that each of these is a dimension along which a particular birth is less likely to be delivered through a cesarean rather than a natural, if the birth is covered by managed care.

Links of C-sections to HMO Activity

First, increases in managed care activity might decrease the potential candidates for cesarean deliveries. Conditional on age and reproductive history, cesareans are much less likely to be necessary when fetal status and health are frequently monitored using electronic fetal monitoring (EFM), ultrasound, and sonographies. In particular, EFM enables the detection of breech babies, whose positioning can be corrected if detected in a timely manner, removing the need to deliver the baby by a c-section. Given this fact, increased managed care activity may affect the risk set for cesareans in two ways:

One, there is empirical evidence that pre-natal interventions result in better peri-natal outcomes and reduced expenditure on neo-natal intensive care (Joyce et al, 1994). Thus, in keeping with their general principle of encouraging preventive services to avoid potentially costly surgery, HMOs might encourage the use of pre-natal services with high frequency. In fact, in 1991 100% of HMOs offered diagnostic procedures, compared to only 67% of traditional insurance program; and, 98% offered well-baby care compared to 39% of conventional insurance.[9] (Source Book of Health Insurance, 1991). Second, women with unobserved preferences for preventive health care may self-select into HMOs and themselves increase the demand for EFMs and other pre-natal services, partly because the marginal cost of every additional service is very low for HMO enrollees.[10] Thus, increased managed care activity might alter the pattern of deliveries by reducing the number of potential candidates for a cesarean.

A second mechanism by which managed care may affect the market-wide cesarean rate is linked to HMOs' selection criteria for physicians and providers, and further reinforced by their capitated payments. If managed care organizations choose practitioners whose practice styles match their preferences, then physicians affiliated with managed care are more likely on average to practice a

[9]I am unable to determine if a 100% offered pre-natal diagnostic services; however, preventive and diagnostic facilities such as mammographies are the hallmark of HMOs, so I assume that coverage of pre-natal services in HMOs is almost universal.

[10]There is some evidence that while HMOs attract a disproportionate number of healthy individuals, their enrollees also demand services at higher rates than enrollees of traditional insurance (Hellinger 1987)

relatively conservative brand of medicine, to prefer less invasive procedures, and recommend more outpatient than inpatient treatments (conditional on expecting the same outcome under both). Thus, insofar as cesareans require lengthier post-delivery hospital stays resulting in higher costs, HMO cesarean-rates are likely to be lower than non-HMO cesarean rates ($\phi(i^m) < \phi(i^t)$, as predicted by the model).[11] These preferences will be further reinforced by introducing a capitated reimbursement scheme that does not reward HMO physicians for performing cesareans. Thus, an increase in HMO penetration may translate into fewer c-sections due to HMO unobserved preferences for naturals, as well as a weakened link between mode of delivery and income.

The diagnoses along which the above reduction in c-section rates is likely to take place are repeat cesareans, *dystocia,* and fetal distress. First, with advances in medical technology, the medical literature notes that the logic underlying repeat cesareans is apparently outdated or non-existent (Weiss, 1997; ACOG, 1996). Professional associations such as the American College of Obstetricians and Gynecologists have released official statements recommending "... that the concept of routine repeat cesarean be replaced by a specific indication for surgery, and that most women can be counseled and encouraged to labor and have a vaginal birth after a cesarean (VBAC)."[12] Thus, repeat cesareans appear to be a margin for HMO physicians to exercise their preference for less invasive procedures, as they appear to be neither necessary nor more remunerative than vaginal births. Likewise, with more careful diagnosis, an HMO ob/gyn may be less likely to classify a specific symptom as *dystocia* or fetal-distress. (Tussing and Wojtowycz, 1994). Further, conditional on such a diagnosis, an HMO ob/gyn may be less likely to recommend a c-section delivery. Thus, unobserved HMO preferences for conservative practices may result in lower c-section rates where HMO activity is highest.[13]

For various reasons, each of the above mechanisms which result in lower c-section rates may be reinforced by changes in the practice patterns of non-managed care physicians. First, non-managed care physicians may begin to mimic HMO ob/gyn behaviour in order to attract HMO contracts and thereby expand their customer base. ACOG (1996) reports indicate that among ob/gyns,

[11]This may be true even if managed care reimbursement for cesareans is higher than that for naturals, as it is with traditional insurance. However, without a capitated financial scheme, HMO preferences alone are unlikely to result in lower *market-wide* c-section rates since increased managed care penetration essentially sorts ob/gyns into two sets, leaving the average c-section rate unchanged.

[12]The success rate of VBACs is 86% where "unsuccessful" is the set of planned VBACs that were subsequently reverted to c-sections.

[13]An ICEA report notes that increased diagnosis, along with active discouragement of repeat-cesaraeans can halve c-section rate. In other parts of the world these efforts have reduced c-section rates to less than 12%.

there was a dramatic rise between 1989 and 1992 in the number of ob/gyns that gave up full-time fee-for-service practices in exchange for salaried contracts with HMOs. Second, the diffusion of a dominant practice is an observed phenomenon across various medical categories, as suggested by physician learning models. This diffusion may be especially rapid for deliveries, since ob/gyns face the highest malpractice premiums in the medical profession making it particularly attractive to conform with the 'norm'. Such learning may also spread amongst patients, perhaps by increasing wariness against the less-commonly used practice (Dranove, 1988).

3.2.3 Related Literature

This paper relates to the literature on two dimensions. Methodologically, this research is most closely related to the work by Chernew (1995), Baker (1995b), Baker and Corts (1996) and Baker and Brown (1997), who pioneered the use of detailed measures of HMO activity in examining narrow measures of health care (*e.g,* mammography pricing and usage, insurance premiums, and medicare fees) to avoid problems of inference that might arise in using aggregate measures of medical care (such as regional hospital costs), since these measures cannot capture unobservable local effects, and aggregate over areas that are unlikely to approximate health care markets.

Two other strands of the literature are relevant for this study. One is the body of work that explores the extent of induced demand, many of which focus on cesareans (Shiono *et al* (1988), Grytten *et al* (1990), Dranove and Wehner (1994), Yip (1994), Gruber and Owings (1996)), but do not consider the HMO v/s FFS issue. Further, many of these studies face a basic identification problem as they use variations in fees to examine quantity responses. A second body of related work is the medical and economics literature on the determinants of cesarean-sections at the individual level, which establishes that coverage by an HMO is less likely to result in delivery by c-section (Dale and Tussing (1994), McCloskey et al (1991), Kizer (1988), Stafford (1987)). Such studies limit the type of inference that the current paper seeks to establish. First, each of these suffers from a drawback by not taking into account the obvious endogeneous determination of insurance choice.[14] Second, the focus of this paper is in determining how managed care activity alters the c-section rate, separating its effect from that of induced demand. None of the above works addresses

[14]Since HMOs promote preventive health care and limit coverage of expensive therapy they may be more appealing to healthier individuals for whom the risk of needing costly medical treatment is low. Existing doctor-patient relationships are also much less likely to be relevant for healthy individuals who may therefore switch readily to HMOs from traditional insurance. Because health is a strong predictor of the probability of a cesarean-section for a woman, insurance choice is therefore an endogenous determinant of a c-section.

this issue.

More generally, this paper contributes to the health economics literature in two ways. First, it links the literature on induced-demand models with that of the growing work on the effects of HMO activity by identifying a medical service that is simultaneously an obvious choice for induced demand on the part of physicians, as well as a natural candidate for a medical service in which HMO v/s non-HMO phsycians will behave differently, in part because of the different practice environments facing each. Second, it adds to the vast body of research that examines how managed care might influence the pricing, provision and quality of health care services (Newhouse (1984), Luft *et al* (1986), Feldman *et al* (1986), Noether (1988), McLaughlin (1988), Welch (1994), Baker and Corts (1996), Kessler and McClellan (1996), Baker and Brown (1997)), and analyzes the policy implications of the findings.

To reiterate, the results found in this paper suggest that increased HMO activity skews deliveries away from the highly-reimbursed cesareans towards natural childbirths, implying first-order (net) cost savings, and come with no tandem decline in the 'quality' of deliveries. Therefore, policies that are structured to encourage the growth of managed care activity are likely to lower spiralling health care costs from this medical service. In the following section, I present a detailed analysis of the econometric modelling and estimation tools used in deriving these results.

3.3 Data and Econometric Specification

The analysis is done in two parts. The first part is the focus of this paper and analyzes the effects of HMO activity on cesarean section rates. In this part, I estimate regressions of county c-section rates on HMO market shares, demographic controls and health care facilities. Below I discuss the data and the regression modelling for this part. The second part is an extension that determines the effects of growing HMO activity on various quality-measures of birth. The dependent variable in these regressions are alternatively neo-natal infant morality rates (by county), documented complications at birth, and Apgar scores (which are summary measures of a newborn's health at 1 and 5 minutes after the birth).

3.3.1 Data

This study is facilitated by the enactment of a 1989 federal law which required every birthing centre to record the mode of delivery on the birth-certificate. The data come from the NCHS which documents the town or city of birth, detailed parental and hospital characteristics, and detailed birth information (type, complications, birth-weight) of each of the 4 million-plus babies born in a hospital setting in 1991 and 1992.[15] The data from the NCHS were merged with firm data from County Business Patterns (1991,1992), data on state health regulatory laws from the Health Care Financing Administration (HCFA) and data on health care facilities and demographic characteristics from the Area Resource Files (ARF).

HMO Market share data

The biggest drawback of the available data is that payor-source cannot be identified by birth. Instead, I have data on county HMO market shares, measured roughly as the fraction of a county's population that are HMO enrollees. This study uses two alternate measures of HMO penetration/market shares: one is computed directly from the ARF (1991) and is the fraction of a county's *insured* population covered by HMOs. These data were collected only for 1991 and 1994, limiting their applicability and accuracy for the study.[16] An alternative measure is that used in Baker and Corts (1996). Counties are used to approximate a market for the relevant health care service, although Health Care Service Areas (HCSAs) (which are larger areas, usually groups of counties) may better approximate the relevant market (Makuc, Hoglund et al (1991)). However, HCSAs are not very well-defined and further, I assume that while HCSAs may be superior measures for those services that are less commonly available, such as M.R.I scans, they may be less relevant for deliveries.

Only those organizations officially listed as Health Maintenance Organizations are used in the HMO market share variable. Over time, a variety of organizational structures that combine features of both traditional insurance and traditional HMOs have emerged, blurring the original distinction between managed care and traditional insurance; these include Independent Practice Associations

[15] Of the babies born in 1991-1992, 99% were born in either a hospital or an accredited birthing centre.

[16] The 1994 data were used to approximate the market shares for 1992; this sort of approximation is unproblematic for the early 1980s when market shares did not vary much over years; however, there was reasonable growth between 1992 and 1994, rendering estimates based on this variable potentially biased and inconsistent.

(IPAs), Point-Of-Service (POS) Plans, and Preferred Provider organizations (PPOs). Where c-sections are concerned, IPAs differ from traditional insurance only in degree. POS Plans and PPOs allow greater flexibility in choosing physicians; it is entirely possible that the physicians affiliated with PPOs and POS face reimbursement schedules much like traditional insurance. Thus, to avoid difficulty in interpretation, only those firms that are organized like traditional HMOs are used in constructing the HMO variable.

In the two years studied in this paper, there is significant variation in managed care activity. In 1992, HMO penetration rates ranged from 0 per cent in Calcasieu, Louisiana to a high of 60.8% in Monroe, New York, with a mean national penetration rate of 15 7%. The mean c-section rate in counties with HMO market shares over 15.7% was 19.5%, and the mean c-section rate in the complementary set was 22.2%. Some striking patterns of HMO activity and c-section rates emerge from studying the data: nationwide, southern states (Alabama, Arkanasas, Florida, Georgia, Louisiana and Kentucky) have the lowest penetration rates and the highest c-section rates (*e.g,* Lousiana is 28.5%, Kentucky is 23.3% and Georgia is 22.49%, which are all above the 1991 mean national c-section rate of 21.9%.) Of course, causality cannot be inferred immediately from these sample statistics since the southern states also have disproportionately higher numbers of women with a larger risk for cesareans (higher poverty levels, larger percent of black women).

Other Co-variates

The strategy here is to use the HMO penetration rate in a county as indicative of the likelihood that an individual mother is covered by an HMO. In keeping with the aggregation level of the HMO market share variable, individual level data from the NCHS are aggregated up to the county level and merged with other data from the ARF (1991) and CBP (1991, 1992); these include aggregate demographic characteristics (*e.g,* the dependent variables: total cesarean section rate, primary c-section rates, the neo-natal infant mortality rate and mean Apgar scores; and other covariates: *e.g,* percent of women in the 40-45 age group, 35-40 age group, 30-35 age group, percent educated, percent of women covered by AFDC, percent married), aggregate health care features (*e.g,* number of hospital beds per 1000, number of obstretician/gynecologists per 1000 population) and business characteristics such as the average number of workers in a firm. Counties with fewer than 100,000 population cannot be identified in the NCHS data, and were thus dropped from the analysis. This reduced the number of counties from 3,080 to 508 for each year. Detailed summary statistics are

presented in Table 1.

3.3.2 Econometric Modelling and Issues

Tests of equations (17)-(18) are carried out under a variety of specifications in order to ensure that the results are robust to varying assumptions about the functional form and the stochastic distribution. Two main methodological issues arise in the empirical analysis. One is the issue of endogeneity, a feature pervasive in such studies, but scarcely addressed in the related literature. The other is specification error due to incorrectly imposed linearity and invalid assumptions about the stochastic distribution. A distinguishing feature of this work is in relaxing these restrictions to consider *non-parametric* or *semi-parametric* estimation to minimize specification error.

Functional Form and Non-parametric estimation

The prototypical analysis of equation (18) would begin with least squares estimation of the following model:

$$\begin{aligned} c &= \alpha_0 + \beta\, hmo + \sum_{k=1}^{K} d_k' \gamma_k + \sum_{l=1}^{L} hcf_l' \delta_l + \epsilon \\ &= \alpha_0 + \beta\, hmo + \mathbf{d}'\gamma + \mathbf{hcf}'\delta + \epsilon \end{aligned} \tag{2.6.1}$$

where the unit of observation is a county, c is the c-section rate, $\mathbf{d}$ denotes a vector of demographic characteristics such as age break-down of the female population and wealth, $\mathbf{hcf}$ is a vector of health care facilities and ϵ denotes unobservable disturbance terms such as tastes for preventive care, or community health consciousness that shifts observed c-section rates; ϵ could contain a county or state fixed effect. Motivated by the comparative statics derived for the theoretical model, I avoid testing a specification such as (19) that is linear in market-share, instead allowing for a general, flexible method that permits non-linearity of an unspecified form. The added generality should better capture the inherent non-linearities; least squares estimation with (1) quadratic, (2) cubic and (3) fourth order terms yielded respectively (1) a significant coefficient on the second-order term, (2) significant coefficients on the linear and third order terms and (3) significant coefficients on all but the linear term, suggesting that the appropriate regression design had to be carefully modelled.

It is well-known that small mis-specifications of functional forms can lead to large biases in

estimated parameters, and thus result in flawed inference [Bickel and Doksum (1976), Han (1985)]. With specification errors, confidence intervals are centred away from the truth, and are invalid even asymptotically. If we have *a priori* knowledge about linearity, ordinary least squares provides a straightforward method to estimate an economic relation, and is best; in the current application, however, it is apparent that linearity is invalid in the sense that $(\hat{\beta}, \hat{\gamma}, \hat{\delta}) \overset{p}{\not\mapsto} (\beta, \gamma, \delta)$. Instead, non-parametric (series) estimation provides a way to uncover the underlying relation we are interested in with no functional form restrictions or distributional assumptions, thus minimizing specification error. Furthermore, in terms of inference, non-parametric methods and linear models have equal power, since non-parametric analogues of (β, γ, δ) are simple to compute and interpret.

I begin by reformulating the model in (19) with no parametric assumptions, and one weak restriction that involves additivity of the error term. For unknown $f(\cdot)$ and an unspecified c.d.f of ϵ, let

$$c_{it} = f(\, hmo_{it},\ \mathbf{d}_{it},\ \mathbf{hcf}_{it}\,) + \epsilon_{it} \tag{2.6.2}$$

where i denotes county and t denotes the year. The goal is to recover the local effect of *hmo* on c, a numeric corresponding to β in equation (19). In the empirical analysis that follows, I report estimates from linear models, non-linear (parametric) specifications and non-parametric specifications to demonstrate the inconsistencies and flawed inference that results from using tightly constrained models unless the constraints are valid.

The paper focuses on series estimation of the function $f(\cdot)$, which is shown to be useful for the additive structure that is estimated (an exhaustive review of series estimators, corresponding convergence rates and asymptotics can be found in Hausman and Newey (1995) and Newey (1996, 1997a, 1997b)). Series estimation begins with the assumption that for some $L > 0$, an $(L \times 1)$ sequence of functions can approximate the underlying function $f(\cdot)$ by a linear combination of the functions, with the property that the approximation to f is arbitrarily close for large L. This underlying principle of series estimation is a simple restatement of the well-known Weierstrass theorem. For example, power series estimation of (20) with a $(K \times 1)$ vector of functions, $p^K(hmo, \mathbf{d}, \mathbf{hcf}) = (p_{1K}(\cdot), \ldots, p_{KK}(\cdot))'$ yields:

$$f(hmo_{it}, \mathbf{d}_{it}, \mathbf{hcf}_{it}) \ \approx \ p^K(hmo_{it}, \mathbf{d}_{it}, \mathbf{hcf}_{it})'\pi$$

$$= \sum_{l_1=0}^{L_1} (hmo_{it})^{l_1} \alpha_{l_1} + \sum_{l_2=1}^{L_2} (\mathbf{d}_{it})^{l_2} \alpha_{l_2} + \sum_{l_3=1}^{L_3} (\mathbf{hcf}_{it})^{l_3} \alpha_{l_3} \qquad (2.6.3)$$

$$\pi = (\alpha'_{l_1}\ \alpha'_{l_2}\ \alpha'_{l_3})$$

giving us the non-parametric estimate of interest:

$$\hat{E}(c_{it}|hmo_{it}, \mathbf{d}_{it}, \mathbf{hcf}_{it}) = \hat{f}(hmo_{it}, \mathbf{d}_{it}, \mathbf{hcf}_{it}) \quad = \quad p^K(hmo_{it}, \mathbf{d}_{it}, \mathbf{hcf}_{it})'\hat{\pi} \qquad (2.6.4)$$

$$\hat{\pi} = \left\{ \sum_{i=1}^{N} \sum_{t=1}^{T} p^K(\cdot) p^K(\cdot)' \right\}^{-1} \cdot \left\{ \sum_{i=1}^{N} \sum_{t=1}^{T} p^K(\cdot)' c_{it} \right\} \qquad (2.6.5)$$

i.e, , $\hat{\pi}$ is the vector of least squares coefficients from regressing c_{it} on $p^K(\cdot)$, treating the approximating sequence as the regression function. That series estimatation is so closely related to (ordinary) least squares, facilitates the asymptotics, which are shown to be simple extensions of the OLS case (Andrews (1991), Newey (1996)).

Often the regression estimate $\hat{f}$ is of little intrinsic interest to the modeler, and the purpose of estimation lies in some numerical characteristic of f. For example, we might be interested in the (local) concavity of convexity of the estimated conditional mean, such as β from equation (19). Such functionals are simple to compute, facilitating the comparison of estimates from parametrized and non-parametrics models (a detailed discussion of linear functionals of series estimators is presented in Newey (1996)). This paper is specifically concerned with the functional that measures the local effect of HMO market shares on c-section rates; estimates of this functional are presented and discussed in Section 5.

To summarize, equation (20) represents the type of general formulation that will be estimated in this paper. A practical drawback of series estimators is the accompanying multi-collinearity induced by using several functions of the vector of covariates, a problem known as the curse of dimensionality. Semi-parametric models in which some components are parametrically specified provide one way of circumventing this problem: For example:

$$c = f(hmo) + \sum_{k=1}^{K} d_k \beta_k + \sum_{l=1}^{L} hcf_l \gamma_l + \epsilon \qquad (2.6.6)$$

The empirical section will report estimates from semi-parametric models. As written, however, the model is incorrectly specified due to endogenously determined HMO locational choices. The next step is to extend the linear instrumental variables estimator to equation (20). This exercise is non-trivial, since valid instruments for *hmo* cannot be used in a "plug-in" sense for unknown $f(\cdot)$,[17] and the theory for a non-parametric analog to the standard IV estimator is undeveloped.

The solution to this problem comes from Vella (1991), who analyzed the problem for the parametric case, Newey, Powell and Vella (1996) for the non-parametric and semi-parametric case, and Das (1997) for an extension of Newey et al to the panel model; it is discussed briefly in the following section.

Endogeneous HMO Locational Choices

Endogeneity is endemic in this study. To illustrate this problem clearly, note that we have the following simultaneous equation system. First, rewrite equation (20) as

$$\begin{aligned} c_{it} &= f(hmo_{it}, \mathbf{d}_{it}, \mathbf{hcf}_{it}) + \alpha_{1i} + \epsilon_{1it} \\ &= f(hmo_{it}, \mathbf{d}_{it}, \mathbf{hcf}_{it}) + \nu_{1it} \end{aligned} \tag{2.6.7}$$

where α_1 is the (time-invariant) part of the error term that may be correlated with the covariates, and ϵ_1 is white noise. Next, taking $\mathbf{d}$ and $\mathbf{hcf}$ as exogenously determined, assume that HMO market shares are determined as follows:

$$\begin{aligned} hmo_{it} &= g(\mathbf{d}_{it}, \mathbf{hcf}_{it}, \mathbf{Z}_{it}) + \alpha_{2i} + \epsilon_{2it} \\ &= g(\mathbf{d}_{it}, \mathbf{hcf}_{it}, \mathbf{Z}_{it}) + \nu_{2it} \end{aligned} \tag{2.6.8}$$

where g is unknown, $\mathbf{Z}$ is a vector of local market-conditions that includes regulatory factors, the number of incumbent insurers, inefficiency or slack in the existing health care system, and incumbent market power. (α_2, ϵ_2) are vectors of unobservables, uncorrelated with demographics, health care features or the set of instrumental variables.

The source of endogeneity lies in the fact that the same unobservables that affect c also affect

[17] This regression is "forbidden": using a non-linear function of the predicted value of an endogenous variable (see Hausman 1984).

hmo, such that $E(\nu_1|\nu_2) \neq 0$; this is plausible and at least three scenarios underlying the endogeneity can be described. First, unobservable tastes for preventive health care or health consciousness might encourage HMOs to the county while simultaneously (negatively) affecting the cesarean rate. Second, a casual survey of my data reveals that HMOs are more likely to locate in up-county, urban, wealthy areas where they may be better ensured of potential enrollees (some empirical evidence is also given in Porell and Wallack, 1990). Simultaneously, c-sections are least likely to be needed for healthier women, who in turn are more likely to live in up-county, wealthy residences.

Another explanation is that insofar as HMOs are the efficient and low-cost competitors to traditional insurers, they may view slack incumbent behaviour (which may be correlated with high c-section rates) as a signal that there is potential to correct inefficient behaviour. The first two scenarios imply a negative bias from simultaneity, while the last induces a positive bias; the net bias cannot be signed.

As described, the source of endogeneity appears to lie in the time-invariant component α_1. Unobservable tastes for health, underlying health-status or community health-consciousness are unlikely to change except over long periods of time, so that for the short panel in this study (1991 – 1992), I assume that

$$E(\alpha_1|\alpha_2) \neq 0, \quad E(\epsilon_1|\epsilon_2) = 0, \quad \text{implying} \quad E(\nu_1|\nu_2) \neq 0 \tag{2.6.9}$$

This is an important assumption to make since α is time-invariant; then, just as in the linear case, differencing the model between t and $t-1$ in equation (24) would remove the source of bias and f could be consistently estimated using non-parametric methods for the panel model:[18]

$$c_{it} - c_{i,t-1} = \left(f(hmo_{it}, \mathbf{d}_{it}, \mathbf{hcf}_{it}) - f(hmo_{i,t-1}, \mathbf{d}_{i,t-1}, \mathbf{hcf}_{i,t-1}) \right) + (\epsilon_{1it} - \epsilon_{1i,t-1}) \tag{2.6.10}$$

Alternatively, we could proceed by exploiting the triangular nature of the system.

To proceed without differencing, assume that $E(\nu_1|\nu_2, \mathbf{d}, \mathbf{hcf}, \mathbf{Z}) = E(\nu_1|\nu_2)$, a condition weaker

[18]Numerous issues arise in differencing the model with unknown $f(\cdot)$. First, $f(hmo_{it}, \mathbf{d}_{it}, \mathbf{hcf}_{it}) - f(hmo_{i,t-1}, \mathbf{d}_{i,t-1}, \mathbf{hcf}_{i,t-1}) \neq f(hmo_{it} - hmo_{i,t-1}, \ldots)$ but rather, is equal to $k(hmo_{it}, hmo_{i,t-1}, \ldots)$ for unknown k; second, a less recognized problem is that even though f is potentially non-linear, differencing the model differences out all time-invariant variables, so that an estimate of f cannot be recovered without some additional work. Techniques to recover f are known (see Porter (1996) and Das (1997) for details).

than assuming the independence of $(\mathbf{d}, \mathbf{hcf}, \mathbf{Z})$ from ν_1. Then,

$$\begin{aligned} E(c_{it}|hmo_{it}, \mathbf{d}_{it}, \mathbf{hcf}_{it}, \mathbf{Z}_{it}) &= f(hmo_{it}, \mathbf{d}_{it}, \mathbf{hcf}_{it}) + E(\nu_{1it}|\mathbf{d}_{it}, \mathbf{hcf}_{it}, \mathbf{Z}_{it}, \nu_{2it}) \\ &= f(hmo_{it}, \mathbf{d}_{it}, \mathbf{hcf}_{it}) + E(\nu_{1it}|\nu_{2it}) \\ &= f(hmo_{it}, \mathbf{d}_{it}, \mathbf{hcf}_{it}) + \lambda(\nu_{2it}) \end{aligned} \tag{2.6.11}$$

where $\lambda(\cdot)$ is an unknown function of ν_2. Although ν_2 is unavailable, a consistent estimate of ν_2 can be obtained from non-parametrically estimating equation (24): $hmo - \hat{g}(\mathbf{d}, \mathbf{hcf}, \mathbf{Z}) = \hat{\nu_2}$. The additive regression in (27) as well as the differenced panel model in (26) are the focus of estimation in the empirical section. Both correct for endogeneity and permit a consistent estimate of $f(\cdot)$ from which a functional (such as β) can be extracted.

The triangular simultaneous equation system described in equations (19)-(20), and in particular the additive error terms in each stage, is an integral part of the analysis. If $\mathbf{Z}$ included c-section rates (c), the reduced form for c-sections could not have an additive-in-errors specification, making it impossible to estimate the model for unknown $f(\cdot)$ and $g(\cdot)$. Fortunately, it is plausible that c is not amongst the many local-market conditions that affects HMO locational choice: deliveries are but one of several medical services that HMOs cover, so that c-section rates alone may neither be indicative of slack behaviour by incumbent insurers, nor important enough to affect HMO locational decisions except at the margin. Nothing in the literature appears to suggest an alternative view (*e.g*, Goldberg and Greenberg (1981), Welch (1984), Robinson (1991)).

Instrumental Variables

Potentially two sets of instruments are available for the analysis. One set, $\mathbf{Z}_1$, is that used previously in studies of HMO activity (*e.g*, Baker and Corts 1996, Baker and Brown 1997) and is related to firm size and type in the market. By a 1978 federal law, all firms with 25 or more employees must offer an HMO if it wants to be offered on their health plans. Thus, the number of establishments that employ more than 25 workers in a county provides a rough guide to the size of the pool of potential enrollees. Note that smaller employers are also more likely to self-insure (Gruber and Poterba, 1995) so that the average employer size is indicative of the potential pool of enrollees on a second dimension as well. Likewise, the number of white-collar workers is assumed to be a valid instrument, since on average white-collar workers are more likely than blue-collar counterparts to

purchase insurance (American Medical Statistics, 1996).

The second set of instruments, $\mathbf{Z}_2$, is the set of state laws regulating the entry, expansion and operation of managed care. This source of exogenous variation is surprisingly under-utilized in the existing literature. State regulatory laws range from requiring certificates of authority for locating and capital requirements, to dual choice options (discussed above) advertisement restrictions, and enabling laws that favour HMO formation in varying degrees (HCFA, 1996). There is tremendous variation across states, with the drawback that the unit of observation for the dependent variable is the county, leaving much within-state variation unexplained.

In the empirical section, I report results with varying sets of instruments. One set uses only state laws, and the other combines the state laws with the establishment variables. For the establishment variables to be valid instruments, they must be uncorrelated with the unobservable component α_1. It is entirely possible that large firms are concentrated in larger and wealthier counties, where the population is healthier and have stronger preferences for preventive care; likewise, larger concentrations of white-collared workers may be found in counties with favourable health characteristics. If true, these possibilities render the establishment variables invalid as instrumental variables.

The state laws, on the other hand, are arguably purely exogenous, strongly correlated with HMO locational choices (and thus, county HMO market-shares) and uncorrelated with health care preferences or other unobservable population health characteristics. A remote possibility is that incumbent non-managed care insurers may be firmly entrenched, have the leverage to influence state laws in their favour and against the formation of HMOs. Then, if incumbents are most likely to influence state regulatory behaviour where they are most slack, we might pick up a spurious negative correlation between HMO market-shares and c-section rates. I rule this possibility out by observing that there is no obvious within-law pattern across states: states that favour HMO activity on some dimensions discourage it on others, and no state falls in either extreme, uniformly discouraging or uniformly supporting the creation of HMOs.

3.4 Results

3.4.1 Effects on cesarean rates

I identify important non-linearities in the effect of HMO market-shares on c-section rates. First, as predicted by the theoretical model, cesarean rates are not uniformly decling in HMO market shares (see Figure 3). There appears to be two "pivotal" points between which HMO activity results in decreased c-section rates, corresponding to market shares of 20.6% and 55% (the mean penetration rate in my data set is 15%.) Second, (1) the non-linearities are under-emphasized by non-parametric estimation that does not take into account the endogeneous determination of HMO activity; and (2) the (average) effect of HMO activity is substantially understated by any linear modelling technique and overstated by unadjusted (for endogeneity) non-parametric estimation.

Table 2*a* reports least squares, instrumental variables (two-stage least squares), semi-parametric (SP) and semi-parametric instrumental variables (SPIV) estimation of the following model:

$$c = f(hmo) + \sum_{k=1}^{K} d_k \beta_k + \sum_{l=1}^{L} hcf_l \gamma_l + \lambda(\hat{\nu_2}) \tag{2.6.1}$$

where $f(\cdot) = 1$ for the parametric models, and remain unspecified for the semi-parametric models. Likewise, $\lambda(\cdot)$ is unspecified for the semi-parametric models and equal to $\psi \hat{\nu_2}$ in the parametric models, where ψ is an additional parameter to be estimated. Table 3*a* and 3*b* repeat the analysis in tables 2*a* and 2*b* with a larger set of instruments. There is little substantive difference in the semi-parametric estimates across the two sets; as a conservative measure, the results discussed below focus on the smaller set of instruments (only the 6 state laws) presented in Tables $2(a, b, c)$ since the establishment variables may not be valid as instruments for the reasons presented in the previous section.

As discussed earlier, at low levels of activity HMO ob/gyns behave in a manner consistent with the induced demand hypothesis, substituting natural childbirths with the leisure-intensive c-sections (recall, there is no financial discrimination between the two procedures for HMOs). Moreover, simply because non-managed care ob/gyns are managing a larger number of total births, their effect on c-section rates will dominate a possibly opposite effect due to HMO ob/gyns. The net effect is that c-section rates are increasing in HMO activity in this range. In Figure 3, this is represented as the positive slope of the curve in market shares. This effect diminishes as market

shares approach 20%.

When HMO penetration levels approach 20%, we see the expected negative effect of increasing HMO market shares on c-section rates. The graph may be capturing several effects. First, we know from the theoretical model that the optimal inducement rate is declining in market shares; the negative sloped portion of the curve represents the cumulative effect of the declining inducement rates due to managed care ob/gyns and the increaseing inducement rates of non-managed care ob/gyns; that it weighs out in favour of the HMO ob/gyns is simply a result of the higher weighting due to the larger fraction of births now managed by HMOs. Between the two pivotal points it suggests that a ten percent increase in HMO market shares results in 1.8 percent decrease on average in c-section rates within the pivotal bounds.

Table 2c reports average changes in c-section rates over many smaller bands of HMO penetration rates. For a 10% change in HMO penetration, these vary from -1.4% between 25% and 30%, to -3.5% when HMO activity is between 35% and 40%, and appear to taper as HMO activity exceeds 50%. In between, the effects vary substantially, and a simple average is uninformative about the real effects of HMO penetration, since there are ranges over which HMO effects on c-section rates are trivial and close to zero (such as in the range close to the peak, between 20% and 25%). Inference based on any linear modelling technique that yields only a simple average effect would be seriously biased. These results clearly go to show that, as the theoretical model predicts, there are differential effects of HMO activity depending on where along the unit interval penetration falls. For illustrative purposes an average effect is sometimes useful to consider; this corresponds to a 1.8% decline in c-section rates for every 10% increase in HMO penetration.

By contrast, inference from least squares suggests a constant effect of HMO activity on c-section rates at any penetration rate. The corresponding effect is about four times as small as that from the SPIV results (a 0.4 percent decrease in c-section rates), while two-stage least squares would suggest an effect slightly closer to the semi-parametric IV estimate, *i.e*, a 1 percent fall in c-section rates for a ten per cent increase in HMO market shares. Thus, non-linearities matter and result in a substantial difference in inferrence. Accounting for the non-linearities is not sufficient, however. A semi-parametric estimate that does not also correct for endogeneity suggests that a 10% increase in HMO market shares will result in a 1.2% decline in the range over which c-section rates are falling in the penetration rate.

Table 2*a* reports estimates on the variable of interest ("hmo") from a variety of specifications. Coefficients on the remaining co-variates are presented in Table 2*d*. The first panel of Table 2*a* shows that OLS estimates are very small, although significantly different from zero. A simple correction for the endogeneity using 2-stage least squares magnifies this effect. Note that the coefficient on the residual-adjusted term *res* is significant, suggesting that endogeneity is certainly an issue. The semi-parametric IV results arise from minimal restrictions on the model and are presumably most accurate. Choosing the optimal number of functions for $f(\cdot)$ and for $\lambda(\cdot)$ comes from *cross validation* of the data. This criterion is conceptually similar to that of the Adjusted R-squared, trading off higher explanatory power with higher standard errors (a clear discussion of the CV criterion with an application is presented in Hausman and Newey (1995)). It is computed by summing the squared residuals for each the ith observation using the parameter estimates from each of the not-i observations. Choosing the number of functions to minimize the CV criterion minimizes the asymptotic bias in series estimation. By the CV criterion, I choose four power series in *hmo* and two in $\hat{v_2}$ (Table 2b).[19] A chi-squared test of the over-identifying restrictions rejects the null at the 5% level.

In Tables 3*a*-3*b*, the analysis is repeated with a larger set of instruments that includes the establishment variables. There is some, but no substantive change in the average effects of HMO activity on c-section rates. Specifically, the estimates reported in Table 3*a* suggest that for a 10% change in HMO penetration rates, there is an average effect of reducing c-section rates by 1.48%. The non-linearities are smoothed out somewhat and there is less variation across different bands of HMO penetration. The chi-squared test for over-identifying restrictions rejects at the 5% level. The results here essentially rely on the validity of the instruments, which must be assumed; for the reasons discussed earlier, the establishment variables may be invalid and the results based on the full set of instruments may therefore be biased. Thus, I focus on estimates reported in Table 2.

As a specification check, I estimate a more general model by using power series in per capita income, county fertility rates, per cent married, number of hospital beds per 1000 pop., ob/gyn

[19]The standard errors are computed to reflect the two-stage nature of estimation. This class of estimators is that in which consistency of the first step affects consistency of the second, a criterion which requires a correction for the appropriate standard errors in the second stage, and results in inflating them. The case above is a special case of two-step estimators for which Whitney Newey (1984) has established an explicit expression for the asymptotic variance. The class of estimators Newey studies is defined by the property that the first step first-order conditions are uncorrelated with the second step first-order conditions, a criterion satisfied here.

density and per cent covered by AFDC, and interactions of these.[20] Figure 5 plots this curve along with the semi-parametric IV curve. There is some difference in the *average* effect of HMO penetration rates on c-section rates over the range that c-sections are falling, but nothing substantive. The difference in the two curves appear to be non-trivial, but the fully non-parametric curve is estimated imprecisely with larger standard errors, so such inference may be unwarranted. The curve peaks earlier (corresponding to a penetration rate of 18%), but the estimation appears to succumb to the curse of dimensionality, so that the (White heteroskedasticity-corrected) standard errors are large and the confidence intervals are correspondingly large, limiting precise inference. In particular, for a 10% increase in HMO market shares, the average decline in c-section rates (over the portion of the curve where the estimated c-section rate is falling) is 1.6%. In summary, the gain in moving from the semi-parametric IV (SPIV) to the non-parametric IV (NPIV) model yields little in terms of precision, so focussing on the SPIV may be of more interest.

3.4.2 Panel Estimation

The next set of results report results from linear and semi-parametric panel estimation of the model. This exercise is undertaken for two reasons. First, the panel structure allows us to difference the model and remove the unobserved time-unvarying heterogeneity factor α that may be correlated with the HMO market share variable. As discussed earlier, much of the unobserved components that are correlated with HMO market shares are likely to be time-invariant such as tastes for preventive care, cultural changes (such as the acceptance of cesareans as a common substitute for naturals) and underlying health demographics. If this is a valid assumption, differencing the model permits estimation without the use of instrumental variables.

A more interesting reason is that differencing the model provides a method to disentangle the short-run mechanisms by which HMO activity affects cesareans from those of the long-run. Specifically, we might expect that physicians respond immediately to changes in their financial environment or to practice guidelines and reward mechanisms (such as those adopted by HMOs), so that analyzing a cross-section is informative about the short-run responses of the c-section rate to HMO penetration rates. By contrast, physician or patient learning, the evolution of the "norm" in deliveries, and the gradual adaptation of malpractice standards are more likely to be captured

[20]The dummy variables can simply be pulled out of the unknown function and evaluated at their two values, and are therefore treated as parametric.

over time. Therefore, variation in the *difference* of cross-sections over time will be more informative about changes in c-section rates that respond gradually to changes in practice environment.

The ideal data set would have a long time-series for such an analysis. I am constrained to having only two years' data. I estimate models of the form:

$$c_{it} - c_{i,t-1} = f(hmo_{it}) - f(hmo_{i,t-1}) + \sum_{k=1}^{K} d_k \beta_k + \sum_{l=1}^{L} hcf_l \gamma_l + (\epsilon_{1it} - \epsilon_{1i,t-1}) \qquad (2.6.1)$$

For the linear model, $f(hmo_{it}) - f(hmo_{i,t-1}) = \delta(hmo_{it} - hmo_{i,t-1})$. For the semi-parametric model, differences of the power series at time t and $t-1$ are used to approximate $f_t - f_{t-1}$. Table 4 presents the results.

Panel estimation reveals some new information about changes in c-section rates. The first two columns of Table 4 are carried out under the assumption that the endogeneity stems from the time-invariant component α; thus instrumental variables is not an issue once the model has been differenced. The coefficient on HMO market share ("hmo") is still fairly small, but about twice as large as the OLS estimate on a non-differenced model (Table 2*d*). Inference on this parameter is valid only if the assumption about the source of endogeneity is valid. As a check, the second two columns report estimates from 2-stage least squares estimation using Average Number of workers, and County establishment size as instruments. The coefficient on the HMO market share variable is substantially magnified relative to OLS, and highly significant. It suggests that learning by physicians and patients, the diffusion of medical practices and the evolution of cultural norms, while gradual, have important roles in explaining the trend of cesarean section rates.

Semi-parametric panel estimation preserves some of the non-linearities uncovered earlier, but distinctly smoothes out most of it: Figure 6 plots both the semi-parametric IV curve as well as the curve derived from panel estimation.[21] Because no constant is estimated in panel estimation, the SPIV is differenced from its mean to compare the two. Two main facts emerge: (1) the hump-shaped effect disappears and (2) while the non-linearities are less important, c-section rates are now uniformly falling in HMO activity.

For slow-moving changes, these are unsurprising findings. If trends, technological changes and

[21] Specifically, taking the parameters obtained from the panel estimation in Table 4, I reconstruct the estimated c-section rate at *levels* of HMO penetration rate.

cultural norms have made cesareans widely acceptable, then c-section rates will have an upward drift over time that will dampen any reverse effect due to short-run changes in financial incentives and ob/gyn practice environment. Gruber and Owings (1996) discuss asymmetric responses to growth and decline in fertility rates as one manifestation of such a trend in c-section rates (although they find no evidence for it).

3.4.3 Quality Effects

The second set of results examines the effects of HMO penetration rates on the three quality measures. Unlike the section above, I disregard non-parametric estimation in examining the birth-quality effects of growing HMO activity. First, there is no theoretical reason to believe that infant mortality or complications at birth might be affected differentially at lower versus higher levels of HMO market activity. One possibility is that both measures may be positively correlated with c-section rates, so that if the c-section rate is non-linearly changing with HMO market shares, so might these quality measures. I expect this is a small second-order effect. Second, as a check and in keeping with the general estimation of the preceding section, non-parametric estimation revealed no discernible non-linear pattern for either of the three measures. As a conservative measure, the reported IV results only use the state laws as instruments for HMO market share.

1. Neo-neo natal infant mortality rate

I begin with the neo-neo-natal (within 6 hours of birth) infant mortality rate. This measure of infant mortality, as opposed to the standard 7 or 28 day rate is preferrable in linking the infant's death to the mode of delivery rather than some extraneous non-hospital factor. I exclude those deaths that are less likely to be a result of the birthing process (such as death from congenital heart problems). A simple correlation of the neo-neo infant mortality rate and HMO market shares is -0.4899, suggesting that a real link might exist.

The first panel of Table 5, which reports least squares estimates, suggests that HMOs may have some salubrious consequences on the neo-neo natal infant mortality rate. Before dismissing this result due to the endogeneity involved, it is worth noting that the result is not entirely implausible. First, if the decrease in c-section rates due to increased HMO activity is because those births which would otherwise have been c-sectioned (and medically unnecessary) are now delivered naturally, this is exactly the result we would expect – C-sections save lives when they are medically necessary, but

pose risks to mother and fetus when induced (Guide to Pregnancy and Childbirth, 1996). A second reason this result may hold is that by increasing pre-natal interventions and diagnostic procedures, HMOs could be decreasing the number of babies that are born with harmful pre-existing conditions.

On the other hand, endogenous determination of HMO market shares might induce this negative result if HMOs are locating in wealthy areas where unobservables negatively affect the infant mortality rate. In fact, a two stage-least squares correction for the inherent endogeneity does remove the obtained result. The coefficient on HMO Marketshare drops from a significant -0.03 to 0.009 which is not distinguishable from zero.

2. Complications at Birth/Abnormalities of the Newborn

A second metric of the 'quality' of birth is a reported complication at birth, or a documented abnormality of the newborn. I divide the listed complications into those that are likely to be affected by the mode of delivery (such as anesthetic complications, fetal injury, excessive bleeding and respiratorial distress) from those that are unlikely to be (fetal alcohol syndrome, feverish babies (*febrile* conditions), anemic babies, and precipitious labor). In the regressions, the former set is used as the measure of the rate of complications.

Controlling for the c-section rate is particularly important for this regression. First, c-sections result in about twice the blood loss as vaginal deliveries (Risks of Cesarean Sections, 1997). Second, anesthesia is almost always administered for cesareans, and less so for natural childbirths. Thus, the complications arising from anesthesia are much more likely to be reported for cesarean births. Third, fetal injury from surgical procedures is more likely with cesareans simply by virtue of c-sections being a surgery (Risks of Cesarean Sections, 1997). Thus, if HMOs are reducing the number of cesareans performed, we might observe a smaller number of such reported complications simply because of the effect of HMOs on the c-section rate. However, we are interested in determining whether or not HMOs have tangible effects on quality independently of their effects on the c-section rate. This is achieved by using the c-section rate as a control variable.

The second panel of Table 5 reports least squares and two-stage least squares estimates for this regression. The OLS numbers pick up a positive impact of HMO activity on the birth-injury rate. If HMOs are prescribing naturals when cesareans are appropriate (perhaps because financial

rewards are tied to naturals), this is the pattern we would expect. Again, this may be spurious because of endogenous HMO market shares. Also, there may not exist sufficient variation in the number of reported complications across counties. In the data, an average of 2 births per 1000 has some documented complications as I measure it, with a standard deviation of 0.004. Only two variables are statistically significant (even though we reject the hypothesis that all coefficients are 0 by an F-test). Correcting for the endogeneity removes the positive effect, reducing the coefficient to 0.0008 which is not statistically different from zero.

3. Apgar Scores

The Apgar Score was developed as a summary measure of a newborn's health measured at 1 minute and 5 minutes after birth. The scores measure the heart rate, respiratory conditions, muscle tone, reflex irritability and color, and are used to predict a newborn's probability of survival. (Monthly Vital Statistics Report, 1994). If any effects of HMO activity on the Apgar scores are noted, they are more likely to reflect the pre-natal care that HMOs emphasize, rather than the method of delivery. Thus, this quality-measure may be less relevant for the current study, but interesting in its own right.

The scores range from 0 to 10, with 7 indicative of good health (MVSR, 1994). I recode the data as a dummy that indicates whether or not the score is larger than 7, and then measure the percentage of births in a county that are of high scores. In my data set, twenty-nine states did not collect information on Apgar scores for both 1991 and 1992. The other twenty-one account for just over 50% of the births.

I find no discernible effect of HMO activity on either the 5 minute or the 1 minute score. Table 5 reports the 5 minute score results. This is true for both the least squares and IV results. Some reading of the literature suggests that this is not a surprising finding. First, in 1994 only 1.4% of all babies had low 5-minute scores (MVSR, 1994). Further, after declining sharply from 1984 to 1990 (a period consistent with growing HMO activity), there has apparently been little change in average Apgar scores (MVSR, 1994).

In summary, the empirical analysis has found two results:

First, the growth in HMO activity has real effects on the cesarean section rate. This finding is

robust, but over-emphasized if (1) the endogeneity of HMO market shares is not addressed and (2) if linear modelling techniques are used. Endogeneity matters. Much of the medical literature that examines the effect of HMO-affiliation on various outcomes completely ignores the endogenous determination of insurance choice. Non-linearities are also important. The empirical results are consistent with the theoretical model developed, which clearly suggest that the inducement effects are much likely to be relevant when HMO market activity is small, and become less important as HMOs capture a larger number of total births. Linear modelling techniques are unable to capture the inherent non-linearities.

Second, there appear to be no obvious adverse effects of HMOs on the quality of births. Again, endogeneity matters, while non-linearities are probably inconsequential for this set of results. In particular, going by neo-neo natal infant mortality rates, documented complications at birth and Apgar scores, increased HMO activity has no discernible effects on quality. While this is an interesting result in and of itself, it is particularly important when viewed jointly with the first set of results. Together they imply that, from the point of view of spiralling health care costs, increased HMO activity is a good.

3.5 Conclusion

The growth rate of health care costs over the late 1980s and early 1990s has been staggering (Newhouse 1992; Cutler, 1994). Much research has emerged to address the sources of these costs and thereby appropriately design policies that can re-structure the provision, usage and financial environment of health care services. By altering the financial incentives associated with inducing demand and over-diagnosing (also known as the "DRG creep"), and by encouraging a preventive and more conservative approach to health care, managed care has begun to receive some attention in the national debate on health care reform.

This paper explores the effects of managed care on the composition of deliveries between the highly reimbursed cesarean-sections and natural deliveries, arguing that real mechanisms exist by which increased HMO activity will result in decreased c-section rates. In particular, that this effect will swamp out any reverse effects on c-section rates due to contemporaneous decreases in fertility rates which might encourage obstetricians/gynecologists to *induce* demand for cesareans.

I develop a model that combines features of demand-inducement by ob/gyns, and changing

market shares of HMOs v/s traditional indemnity insurance. An important prediction of this model is that after a critical HMO penetration rate, increasing HMO penetration will result in decreasing the c-section rate. This relation will be non-linear of an unspecified form. Taking the model to the data and analyzing it with very general *non-parametric* and *semi-parametric* estimation tools, I find both that c-section rates are declining in HMO market shares between two pivotal points, and that this relation is highly non-linear: for a 1% increase in HMO penetration rates, it ranges from a 0.04% decrease in c-section rates when HMO penetration is between 20 and 25%, to 0.32% for penetration in $35 - 40\%$. An average estimate is that every 10% increase in HMO penetration results in a 1.8% decline in c-section rates. The corresponding net savings at 1992 c-section rates and reimbursement schedules is $11 million.

For completeness, I also address whether such beneficial effects on health care costs due to increased HMO activity come with tandem declines in "quality" of birth. Three measures of the quality of birth are used: the neo-neo natal infant mortality rate (within 6 hours of birth), reported birth-complications, and Apgar scores. I find no effect of HMO activity on any of these measures. I take the two strands of empirical evidence found in this paper to suggest that increased HMO activity is, from the perspective of health-care reformers, Pareto-improving. This paper therefore provides the evidence needed in the design and evaluation of policies to re-structure the reimbursement of deliveries and the organization and practice environment for obstetricians. It suggests that policies which favour the growth of managed care will have first-order cost savings from this medical service.

Appendix A

References

American Medical Women's Association Guide to Pregnancy and Childbirth, Americal Medical Women, 1996.

Baker, Laurence; Brown, Martin, 1997 "The Effects of Managed Care on Health Care Providers: Evidence from Mammography, Stanford working paper, August 1997.

Baker, Laurence C.; Corts, Kenneth, 1996. "HMO Penetration and the Cost of Heal th Care: Market Discipline or Market Segmentation? " American Economic Review; v86 n2 May 1996, pp. 389-94.

Chernew, Michael, 1995. "The impact of non-IPA HMOs on the number of hospital and hospital capacity, " Inquiry, 32:2, 143-154.

Cromwell, Jerry and Janet Mitchell. 1986. "Physician-induced demand for surgery.". Journal of Health Economics, 5, 293-313.

Cutler, David, 1995. "The Cost and Financing of Health Care," American Economic Review; v85 n2 May 1995, pp. 32-37.

Cutler, David, 1994. "A Guide to Health Care Reform," Journal of Economic Perspectives, v8 n3 Summer 1994, pp. 13-29.

Das, Mitali, 1997. " 'Complete' non-parametric estimation of Panel Data Models, and Non-parametric Instrumental Variables' ", MIT Department of Economics working paper.

Dranove, David, 1988. "Demand inducement and the physician/patient relationsip.". Economic Inquiry, 281-298.

Evans, Robert. 1974. "Supplier-induced demand: some empirical evidence and impliations." in *The Economics of Health and Medical Care.* , Mark Perlman (Ed.), London.

Feldman, Roger, Hung-Ching Chan et al, 1990. "Effects of HMOs on the creation o f competitive markets for hospital services". Journal of Health Economics, 9 207-223.

Feldman, Roger and Bryan Dowd 1986. " Is there a competitive market for hospita l services?" Journal of Health Economics 5, 277-292

Feldman, Roger and Bryan Dowd 1986. " Is there a competitive market for hospita l services?" Journal of Health Economics 5, 277-292

Feldman, Roger, et al 1986. "The competitive impact of health maintentance orga nizations on hcspital finances: an exploratory study". Journal of Health Politics, Po licy and Law, 10:4 (Winter), 675-698.

Frank, Richard and Pete Welch 1985. " The competitive effects of HMOs: a review of the evidence." Inquiry, 22 (Summer) 148-161.

Garber, Alan and Macurdy, Thomas (1992). " Payment Sou.ce and Episodes of Institutionalization," in *Topics in the economics of aging.* A National Bureau of Economic Research Project. University of Chicago Press 1992, pp. 249-71.

Goldberg, Lawrence and Warren Greenberg 1981." The determinants of HMO enrollment and growth." HSR, 16: 421-440.

Gruber, Jonathan and Maria Owings. Spring 1996. "Physician Financial Incentives and Cesarean Section Delivery", Rand Journal of Economics; v27 n1 Spring 1996, pp. 99-123

Grytten, Jostein et al 1990. " Supplier-inducement: Its Effect on dental services in Norway." Journal of Health economics, 9, 483-491.

Health Maintenance Organizations, 1996. Aspen Publishers, Reprint from *A Report to the Governor on State Regulatin of Health Maintenance Organizations,* Health Care Financing Administration.

Joyce, Theodore et al, 1994. " The Impact of Prenatal Exposure to Cocaine on Ne wborn Costs and Length of Stay,"Health Services Research; v30 n2 June 1995, pp. 341- 58.

Kessler, Daniel and McClellan Mark, 1996. " Do Doctors Practice Defensive Medicine?," Quarterly Journal of Economics; v111 n2 May 1996, pp. 353- 90.

Kohrman et al, 1990. "Medicaid Enrollees in HMOs: A Comparative Analysis of Perinatal Outcomes for Mothers and Newborns in a Large Chicago HMO," Center for Health Administration Studies, University of Chicago.

Maddala G.S (1994). "A Survey of the Literature on Selectivity Bias as It Pertains to Health Care Markets,"in *Econometric methods and applications.* Elgar; distributed in the U.S. by Ashgate, Brookfield, Vt., 1994, pp. 330-45.

Luft, Harold, S.C Maerki and J.B. Trauner, 1986. "The competitive effects of health maintenance organizations: another look at the evidence from Hawaii, Rochester and Minneapolis/St. Paul," Journal of Health Politics, Policy and Law, 10, Winter 625-658.

Martin, Margaret, and Elisabeth Bing. Pregnancy and Childbirth; The Basic Illustrated Guide, 1997.

McClellan, Mark, 1997. "Hospital Reimbursement Incentives: An Empirical Analys is," Journal of Economics and Management Strategy; v6 n1 Spring, pp. 91-128.

McGuire, Thomas and Pauly, Mark V, 1991. "Physician response to fee changes with multiple payers.', Journal of Health Economics, 10, 385-410.

McLaughlin, Catherine, 1988. "The effects of HMOs on overall hospital expenses: anything left after correcting for simultaneity and selectivity?". HSR: 23:3, August, 421-441.

Newey, W.K, Powell J.L, Vella, F (1997). "Nonparametric Estimation of Triangular Simultaneous Equations Models," MIT, Department of Economics.

Newhouse, Joseph. 1970. "A Model of Physician Pricing," Souther Economic Journal, 37, 174-183.

Newhouse, Joseph, 1984. " Cream skimming, asymmetric information and a competitive insurance market." Journal of Health Economics 3, 97-100.

Phelps Charles, 1992. "Diffusion of information in medical care." Journal of Economic Perspectives, 6:3, Summer 23-42.

Porter, Jack (1996). "Nonparametric Regression Estimation for a Flexible Panel Data Model", MIT mimeo.

Profile of Ob-Gyn Practice 1991-1994, 1996. The American College of Obstetricians and Gynecologists.

Practice Expenses for Ob-Gyns, 1994, 1996. The American College of Obstetricians and Gynecologists.

Robinson, P.M. (1988). "Root-n-consistent Semiparametric Regression", Econometrica, v56 n4 July 1988, pp. 931-54.

Stoker, T. (1986). " Consistent estimation of scaled coefficients," Econometrica 54, 1461-1481.

Shiono, Patricia et al, 1987. " Reasons for the Rising Cesarean Delivery Rates 1978-1984". Obstetrics and Gynecology, 696-700.

Source Book of Health Insurance, 1991. Health Insurance Association of America.

Stafford, Randall 1990. " Cesarean Section Use and Source of Payment: An Analysis of California Hospital Discharge Abstracts." American Journal of Public Health, 313-315.

Tussing, Dale and Wojotowycz, Martha, 1992. " The Cesarean Decision in New York State." Medical Care, 529-540.

Tussing, Dale and Wojotowycz, Martha, 1994. " Health Maintenance Organizations, Indepenent Practice Association and Cesarean Section Rates., HSR, 29:1, 75-93.

Vital Statistics Report, June 1996. Centre for Disease Control and Prevention. Advance Report of Final Natality Statistics, 1994.

Welch, Pete, 1994. "HMO market share and its effect on local Medicare costs," HMOs and the Elderly, Health Administration Press.

Yip, Winne. 1994 " Physician Responses to Medicare Fee Reductions: Changes in the Volume and intensity of supply of coronary artery bypass graft (CABG) surgery in the medicare and private sector". MIT PhD Thesis.

Fig.1 C-Section rates, fertility rates, and HMO enrollment (1980-1994)

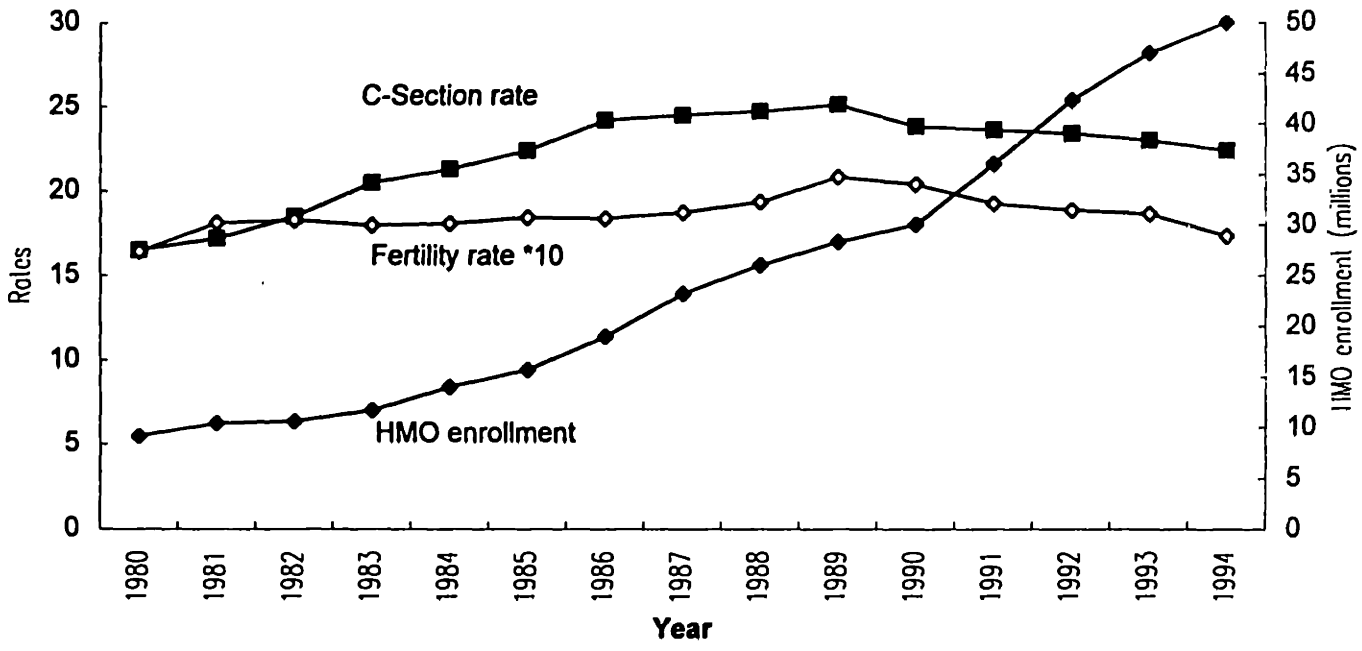

Fig 2 : Total and Primary C-section Rates

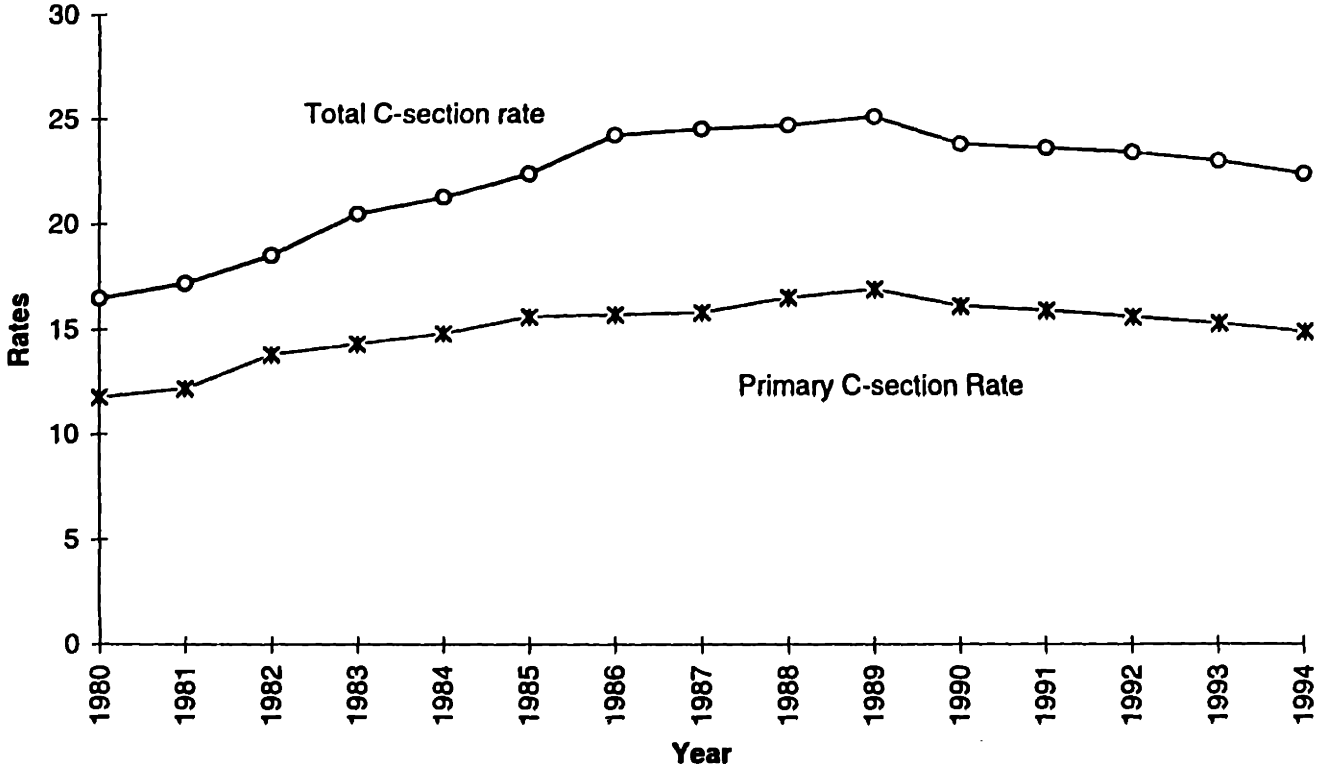

Table 1: Summy Statistics of Variables in Analysis

Variable	Mean		Std Dev		Min		Max	
Year	1991	1992	1991	1992	1991	1992	1991	1992
Total C-section rate	0.220	0.217	0.048	0.054	0	0	0.415	0.5
Primary C-section Rate	0.140	0.133	0.030	0.035	0	0	0.256	0.333
HMO Market Share	0.136	0.164	0.102	0.116	0	0	0.539	0.607
Breech Presentation (%)	0.039	0.040	0.014	0.021	0	0	0.234	0.333
Diagnosed Fetal Distress (%)	0.040	0.040	0.025	0.030	0	0	0.299	0.443
Lowbirthweight (%)	0.065	0.065	0.024	0.023	0	0.021	0.157	0.166
Hosp.beds per 1000 pop	4.934	4.767	2.701	2.577	0	0	17.85	17.17
Ob/gyns per 1000 pop.	4.328	4.363	14.29	14.24	0	0	127.05	124.0
Fertility rate (Births/1000 pop)	0.016	0.016	0.006	0.006	0.0008	0.0005	0.043	0.042
Per capita Income (1000s)	19.01	20.08	4.01	4.24	9.23	9.80	40.49	49.19
% White collar	0.83	0.83	0.138	0.136	0.264	0.265	0.998	0.998
Average number of workers	12.56	16.22	7.23	3.06	2.91	4.03	88.05	90.03
% (Mothers) Age < 20	0.13	0.13	0.05	0.05	0.031	0.028	0.337	0.316
% (Mothers) Age > 35	0.05	0.06	0.019	0.021	0.005	0.016	0.16	0.222
High School graduates (%)	0.383	0.382	0.088	0.067	0.1	0.19	0.6	0.568
College graduates (%)	0.210	0.219	0.058	0.042	0	0.08	0.404	0.412

Notes:

Number of observations = 438 in 1991, 448 in 1992
Data sources
NCHS: Total c-section rate, primary c-section rate, Breech presentation, Diagnosed fetal distress, Low birthweight
ARF : Per capita income, high school graduates, college graduates, age break down, fertility rate
CBP : Percent white collar, average establishment size

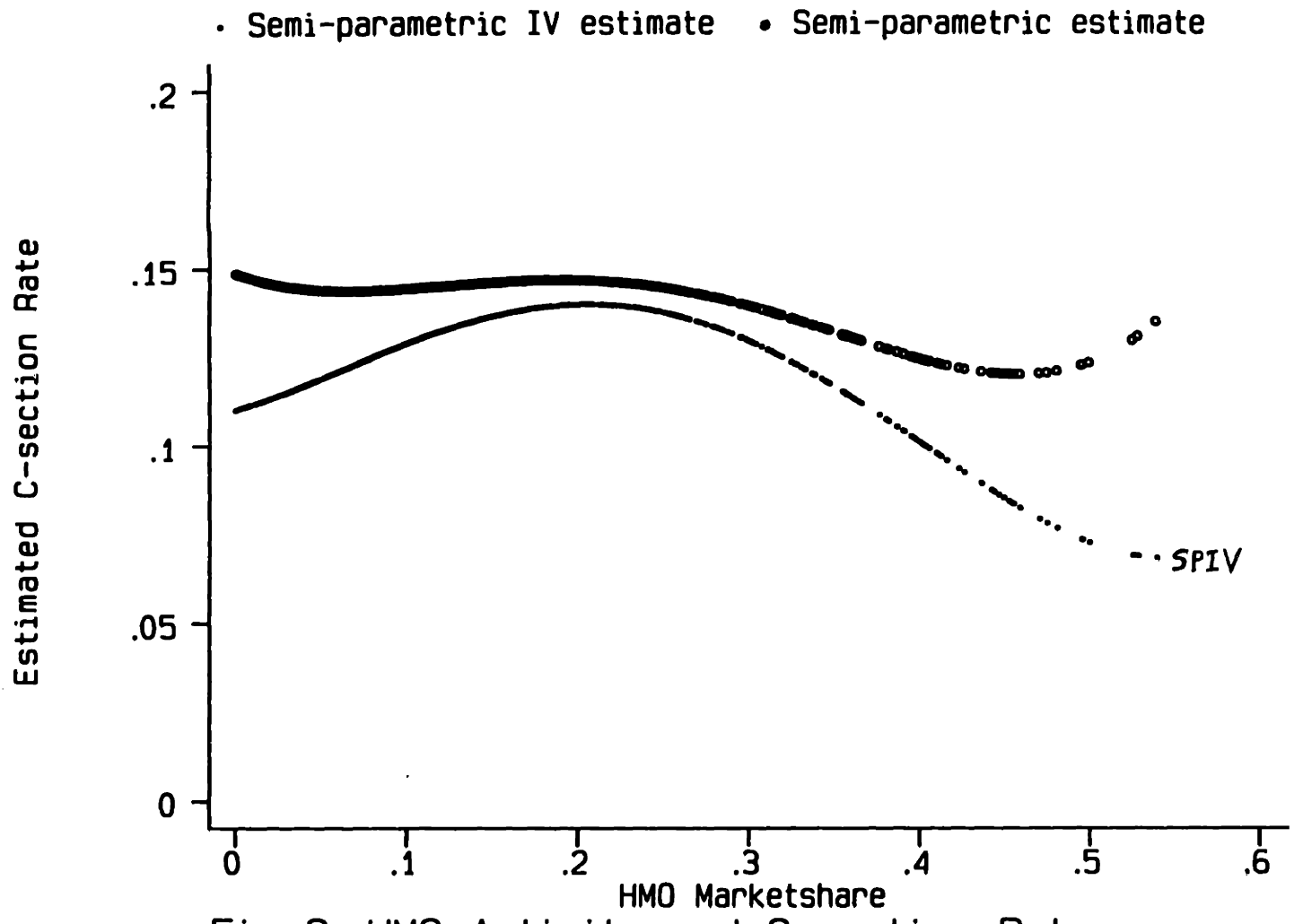

Fig 3: HMO Activity and C-section Rates

Table 2a : Least Squares, Two-stage least squares, Semi-parametric and Semi-parametric IV (NPIV) effects of HMO Market-shares on Total C-section Rates (Instrument Set 1)

		OLS	T-stat	2-stage least sqs	T-stat	Semi-parametric	T-stat	SPIV (Res-adjusted)	T-stat
Instrument Set 1	hmo	-0.0404	-2.077	-0.1020	-1.946	-0.2412	-1.536	0.1378	1.717
	hmo^2					2.6393	1.3612	1.0110	1.739
	hmo^3					-9.1971	-2.314	-6.1618	1.567
	hmo^4					9.2989	3.858	6.5721	2.204
	res			-0.1206	-1.754			-0.0992	-2.345
	res^2							0.7126	2.280

Dependent Variable: Total C-section Rate, n = 805
hmo: HMO marketshare, res = residual from reduced form for hmo
Instrument Set 1: 6 State Legislation Laws
Other co-variates: Fertility rates, % White, % Women 40-45, 35-40, < 15 years, % high school, % College, 3 regional dummies, hospital beds per 1000 pop, Ob/gyns per 1000 , % married, per capita income, % AFDC, % urban
Note: Standard Errors are White-consistent and take into account the two-step nature of estimation.

Table 2b: Cross-Validation statistics:

Powers of hmo	Powers of Res	Cross validation value
2	2	91.879
3	2	91.821
3	3	91.796
4	2	91.734
4	3	91.773

Res is the residual obtained from a non-parametric regression of Hmo marketshare on all exogenous variables.

Table 2c: Average Derivative Estimates of 1 % change in HMO Penetration on % change in C-section Rates.

HMO Penetration Band	Semi-parametric IV	Semi-parametric		
20.6 % - 25 %	- 0.0472	- 0.145	Corresponding estimate from linear IV = **-0.1** at all penetration levels	Corresponding estimate from OLS = **-0.04** at all penetration levels
30% - 35%	- 0.2495	-0.139		
35% - 40%	- 0.3139	-0.1281		
45 - 50%	- 0.2549	-0.1208		
> 50 %	- 0.06	-0.1314		

Table 2d: Least squares, 2sls and Semi-parametric IV effects on C-section rates (Co-variates aside from hmo and res)

	Least squares	2-stage least squares	SPIV(res-adjusted)
per cap income	0.0011 (2.023)	0.0015 (2.562)	0.001 (2.136)
fertility rate	-0.155 (-0.392)	-0.1405 (-0.356)	-0.1460 (-0.367)
% breech babies	0.6774 (7.453)	0.6819 (7.506)	0.6671 (7.251)
% low birthweight	-0.2002 (-1.809)	-0.1765 (-1.585)	-0.205 (-1.852)
% college educated	-0.218 (-0.567)	-0.0172 (-0.448)	-0.0085 (-0.222)
No. of hospital beds/1000 pop.	0.0025 (2.921)	0.0024 (2.835)	0.0026 (3.099)
No. of ob/gyns per 1000 pop.	-0.0002 (-0.443)	-0.0000 (-0.453)	-0.0000 (-0.252)
% AFDC	0.2234 (4.615)	0.2111 (2.224)	0.204 (2.088)

n=886
F(19,866) = 6.40
Adjusted R-squared = 0.1039

Notes:
T-statistics in parentheses.
Controls also include three regional dummies, year dummies, % white,
Four age groups for women (> 45, 40-45, 35-40, 25-35), % urban

The test of over-identifying restrictions is an asymptotic chi-squared test on the regression of the residual from the above regression on the set of instruments. The chi-squared statistic for this regression was 3.09, which was rejected at the 95% confidence level.

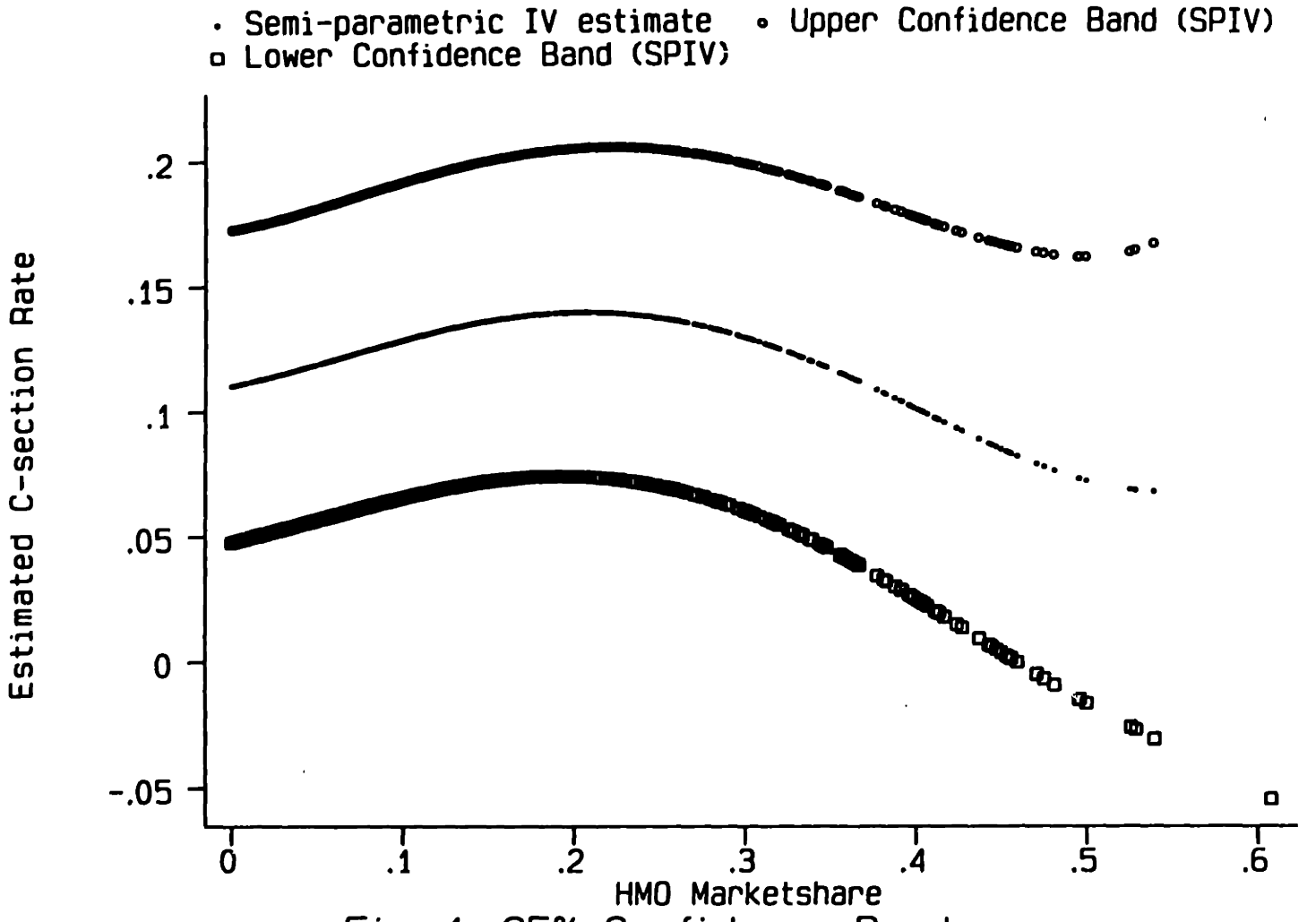

Fig 4: 95% Confidence Bands

Table 3a : Least Squares, Two-stage least squares, Semi-parametric and Semi-parametric IV effects of HMO Market-shares on Total C-section Rates (Instrument Set 2)

		OLS	T-stat	2-stage least sqs	T-stat	Semi-parametric	T-stat	SPIV (Res-adjusted)	T-stat
Instrument Set 2	hmo	-0.040	-2.077	-0.1440	-1.464	-0.2412	-1.536	0.0911	0.4064
	hmo^2					2.639	1.316	1.7113	1.222
	hmo^3					-9.1971	-2.314	-8.1695	4.012
	hmo^4					9.2989	3.858	8.2919	3.889
	res			0.1068	2.025			-0.1242	-1.708
	res^2							0.7043	2.200

Dependent Variable: Total C-section rate n = 805
Instrument Set 2: 6 State Legislation Laws, % White collared workers, Average establishment size.
Other co-variates: Fertility rates, % White, % Women 40-45, 35-40, < 15 years, % high school, % College, 3 regional dummies, hospital beds per 1000 pop, Ob/gyns per 1000 pop % married, per capita income, % AFDC
hmo: HMO marketshare, res = residual from reduced form for hmo

Table 3b: Cross-Validation:

Powers of hmo	Powers of Res	Cross validation value
2	2	91.981
3	2	91.855
3	3	91.756
4	2	91.7498
4	3	91.753

Res is the residual obtained from a non-parametric regression of Hmo marketshare on all exogenous variables.

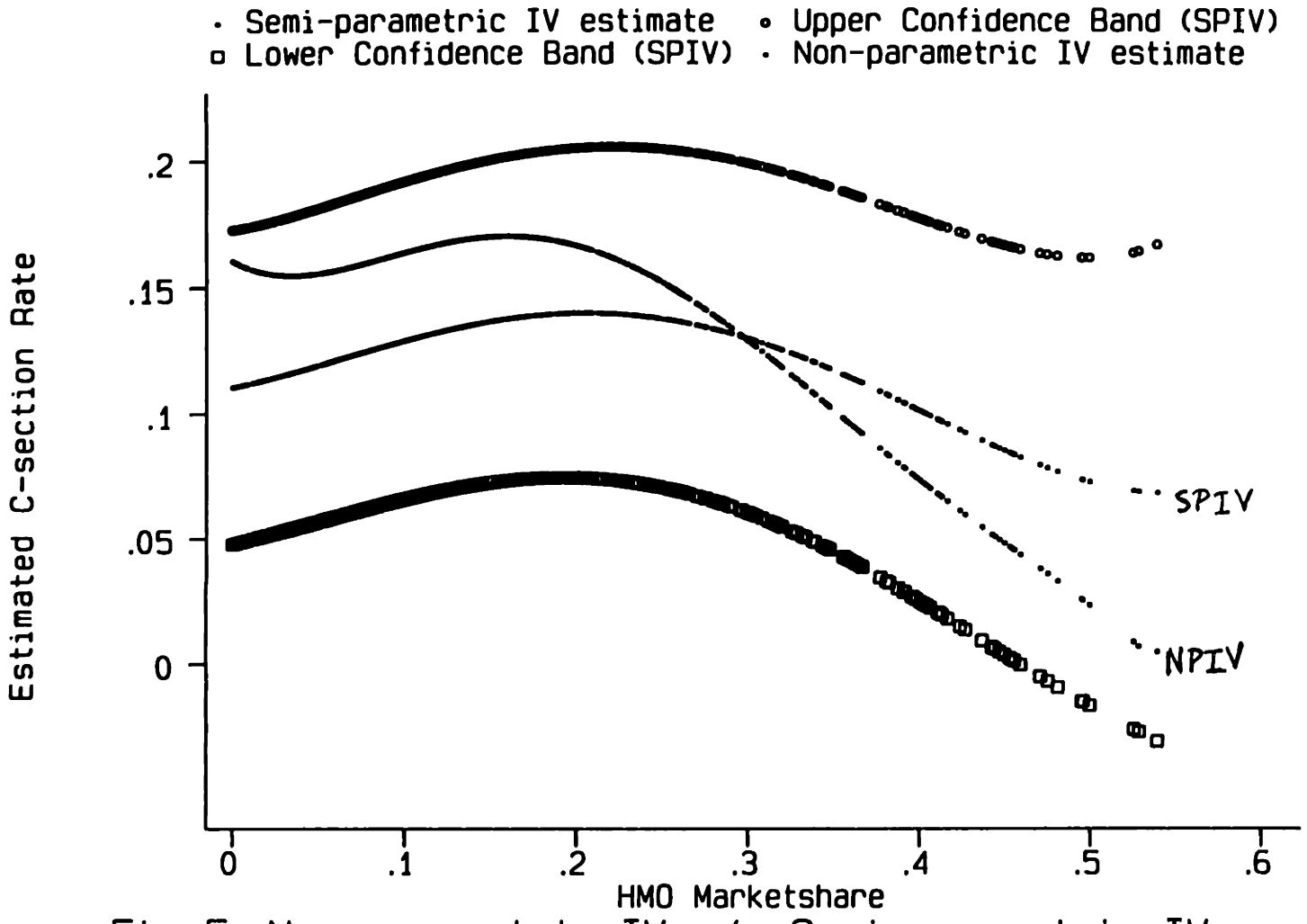

Fig 5: Non-parametric IV v/s Semi-parametric IV

Table 4:

Panel OLS , IV and SPIV results Examining Effects of Change in HMO Marketshare on Change in Total C-section rate

	Least Squares		2-stage least squares		Semi-parametric Residual adjusted	
Variable	Coefficient	S Error	Coefficient	S.Error	Coeff	S.Error
Δ HMO Market Share	-0.075**	0.0309	-0.2576*	0.1430	-0.419**	0.2087
Δ (HMO Market Share)^2					2.318 *	1.201
Δ (HMO Market Share)^3					-5.309	5.509
Δ (HMO Market Share)^4					3.789	5.286
Δ% Breech Presentation	0.547**	0.1023	0.535**	0.102	0.560**	5.390
Δ% Lowbirthweight	-0.2456	0.1519	-0.3209*	0.1615	-0.212	0.1508
ΔHosp.beds per 1000 pop	0.0024**	0.0012	0.0010	0.0016	0.0020**	0.0011
ΔPer capita Income	0.008	0.009	0.0023	0.0016	0.0021**	0.0008
Δ% women on AFDC	-0.0593	0.175	-0.0046	0.1798	0.1640	0.1711
Δ% White (Mothers)	0.0161	0.0363	0.029	0.037	0.099 **	0.0349
Δ% Married (Mothers)	-0.0613	0.0499	-0.1078*	0.0613	-0.0597	0.0533
Δ% High School Educ (Mothers)	0.0780**	0.037	0.0395	0.0470	0.047	0.0383
Δ% Fertility rate	-0.235	0.465	-0.347	0.423	-0.0657	0.1617
Δ% College Educated (Mothers)	-0.055	0.0612	0.0235	0.0847	-0.0209	0.064
Prob > Critical F	0.0029		0.0020		0.0019	
Adjusted R-squared:	0.0382		0.0402		0.0402	
	N = 411		N = 410		N = 400	

Notes:

The Omitted Regional Dummy is West.

* : significant at the 10% level **: significant at 5% confidence level

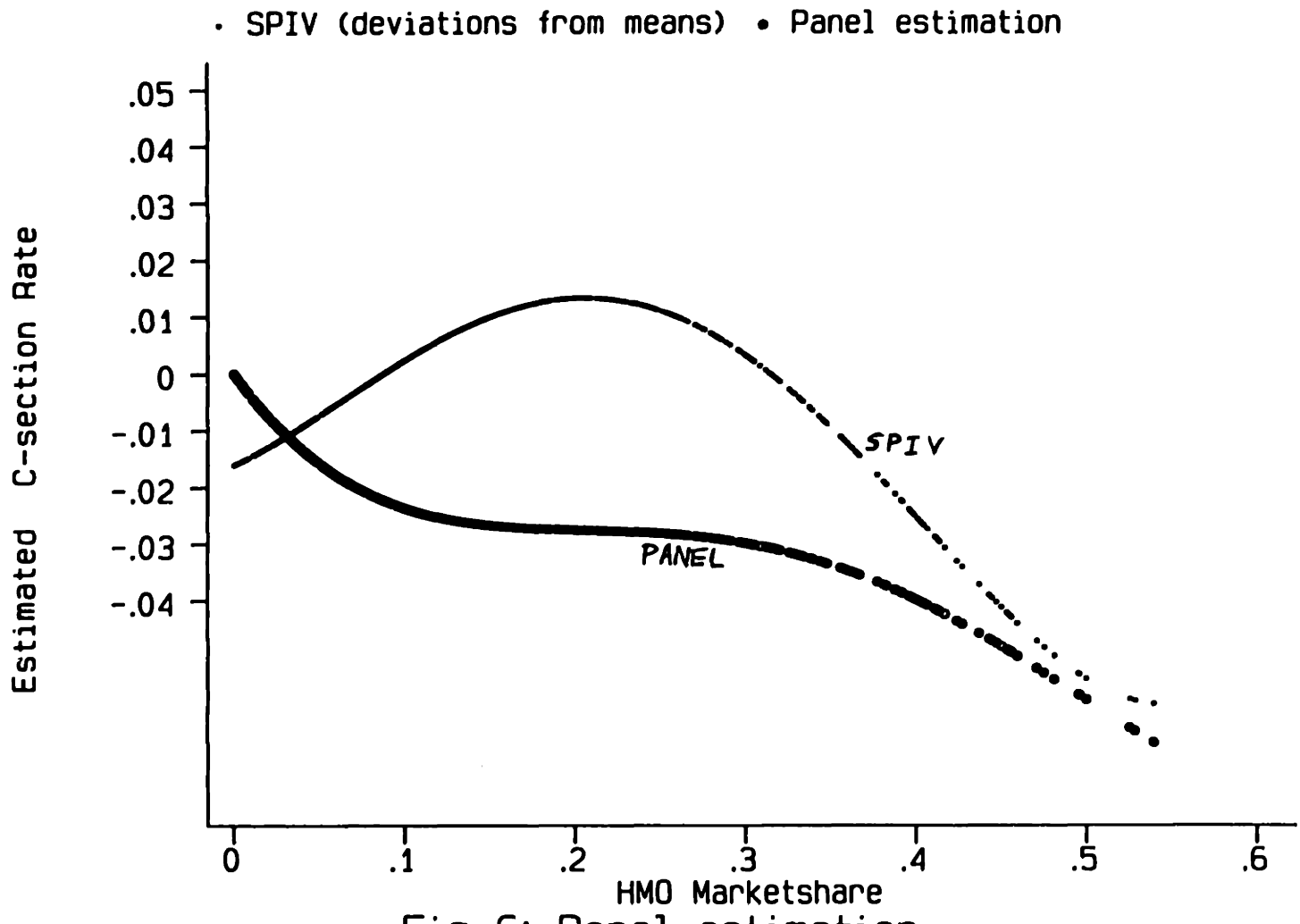

Fig 6: Panel estimation

Table 5 : Quality Effects of Increased HMO Marketshares. Effects on Neo-neo natal Infant mortality rates, Complications at birth, and Apgar Scores

Dep Var :	Neo- neo-natal infant mortality rate		Documented complications/ Birth injury Rate		Apgar Scores	
	least sqs	2 stage ls	least sqs	2-stage ls	least sq	2 stage ls
hmo marketshare	-0.0317 (-5.849)	-0.0108 (-0.500)	0.0059 (2.231)	0.0008 (0.089)	0.0376 (0.434)	-0.2884 (-0.795)
fertility rate	0.0507 (0.447)	0.0435 (0.383)	-0.063 (-1.164)	-0.0614 (-1.120)	-1.728 (-0.920)	-1.616 (-0.858)
c-section rate	-0.0094 (-0.904)	-0.011 (-1.088)	-0.0165 (-3.599)	-0.0163 (-3.557)	-0.1378 (-0.947)	-0.1123 (-0.765)
residual adjusted correction	-----	-0.0434 (-2.019)	-----	0.0051 (0.0519)	-----	0.3260 (0.925)
% Breech babies	0.1130 (3.120)	0.1178 (3.253)	0.0303 (2.412)	0.0298 (2.362)	-0.4035 (-0.693)	-0.429 (-0.736)
% low birthweight	0.191 (5.17)	0.186 (5.512)	0.0358 (2.382)	0.035 (2.326)	0.679 (1.207)	0.6084 (1.094)
per capita income	-0.0027 (-1.830)	-0.0002 (-1.729)	-0.0001 (-1.588)	-0.0001 (-1.600)	0.0036 (1.336)	0.0037 (1.363)
N =	633		814		399	
Adjusted R^2 =	0.6648	0.6626	0.058	0.0358	0.0100	0.0104
	F(19,613) = 63.99	F(18,614)=66.97	F(19,794)=2.54	F(18.795)=2.72	F(18,380)=1.23	F(19,379)=1.21

T-statistics in parentheses.
Other covariates : No. Of hospital beds per 1000, regional dummies, no. of ob/gyns per 1000, % married, % white, % urban, % college educated, % low birthweight, , % distress, year dummy, % women on AFDC