SIMULTANEITY WITH DOWNWARD SLOPING DEMAND

Charles F. Manski
Department of Economics
University of Wisconsin-Madison

February 1994

This research is supported by National Science Foundation Grant SES92-23220. I have benefitted from comments received in presentations at the Hebrew University of Jerusalem, the University of Minnesota, and the University of Wisconsin-Madison.

Abstract

This paper considers anew the *simultaneity problem* that arises when observations of transactions are used to study the demand behavior of price-taking consumers. Simultaneity is shown to be a problem of censored outcomes. This fact is used to obtain a basic negative finding on identification in the absence of prior information on the structure of demand or the process of price determination. Then the assumption of downward sloping demand is imposed. The main result of the paper is a proposition showing that this assumption has considerable identifying power. An empirical illustration is provided.

1. Introduction

This paper considers anew the *simultaneity problem* of econometrics. The simultaneity problem arises when observations of transactions are used to study the demand behavior of price-taking consumers or the supply behavior of price-taking (or quantity-taking) firms. It arises when observations of the outcomes of games are used to study the reaction functions of the players. The present analysis applies in all these contexts, but focusses on demand as a leading case.

Section 2 provides background on the econometric analysis of demand. Section 3 observes that simultaneity is actually a problem of censored outcomes. Proposition 0 uses this fact to obtain a basic negative finding on identification in the absence of prior information on the structure of demand or the process of price determination. Section 4 imposes the assumption that demand is downward sloping. Proposition 1 shows that this assumption has considerable identifying power, even when no other prior information is available. Section 5 uses a well-known simultaneity problem of criminology to illustrate Proposition 1. Section 6 briefly discusses the identifying power of exclusion restrictions. Section 7 calls attention to open questions stimulated by this analysis.

2. Econometric Analysis of Demand

The econometric analysis of demand supposes there is a set of isolated markets for a given product, each market separated from the others in time or space. Each market has a value for $[D(\cdot),p,q,x]$. Here x denotes some covariates characterizing a market, q is the quantity of product transacted, and p is the unit price at which these transactions take place. The market demand function $D(\cdot)$ gives the quantity of product that price-taking consumers would purchase if price were set at any level; so $D(t)$ is the quantity demanded if price were set equal to t. The transaction (p,q) is assumed to lie on the demand function, so

(1) $q = D(p)$.

Markets vary in their values of $[D(\cdot),p,q,x]$. This heterogeneity is expressed by making the set of markets a probability space and by treating demand, transactions, and covariates as random variables with some distribution $P[D(\cdot),p,q,x]$. Let $P[D(\cdot)|x]$ denote the distribution of demand functions among markets sharing the same covariates. Econometric demand analysis seeks to learn about $P[D(\cdot)|x]$ when observations of (p,q,x) are obtained by some sampling process, such as random sampling of markets, that reveals $P(p,q,x)$.

The problem of learning $P[D(.)|x]$ may be posed for any specification of the covariates x. Given such a specification, the inferences that may be drawn depend on what is a priori known about the structure of demand and the process of price determination. Econometric analysis has long centered on the case where demand is known to be a linear function of price, with the same slope β in each market and an intercept ϵ that may vary across markets; thus

(2) $D(t) = \beta t + \epsilon$.

Study of this linear model was initiated in the 1920s and crystallized by the early 1950s (see Hood and Koopmans, 1953).

Perhaps the most familiar finding is that the demand slope β is identified given an *exclusion restriction* of the form[1]

(3a) $E(\epsilon|x=x_i) = E(\epsilon|x=x_j)$

(3b) $E(p|x=x_i) \neq E(p|x=x_j)$,

where x_i and x_j are known points on the support of the distribution of covariates. To see that (1), (2), and (3) identify β, use (1) and (2) to write

$$(4) \quad \epsilon = q - \beta p,$$

apply (3a) to find

$$(5) \quad E(q - \beta p | x=x_i) = E(q - \beta p | x=x_j),$$

and apply (3b) to conclude that

$$(6) \quad \beta = \frac{E(q|x=x_i) - E(q|x=x_j)}{E(p|x=x_i) - E(p|x=x_j)} .$$

The sampling process identifies the right side of (6), hence ß.

In fact, (1), (2), and (3) imply that $P[D(\cdot)|x]$ is identified. By (4), knowing ß implies that ϵ is identified in each market whose outcome (p,q) is observed. By (2), knowing ß and ϵ implies knowledge of the demand function $D(\cdot)$. Knowing $[D(\cdot),x]$ in a random sample of markets reveals $P[D(\cdot)|x]$. Observe that these results are obtained without assuming that firms have supply functions or that demand is downward sloping. The only requirement of firms is that they behave in a manner that makes $E(p|x=x_i) \neq E(p|x=x_j)$. There is no need to assume that $\beta \leq 0$ as conditions (1), (2), and (3) suffice to reveal ß.

3. Simultaneity is Selection

In a series of recent articles on the *selection problem* (Manski, 1989, 1990, 1994), I have found that fresh insights emerge if one puts aside conventional econometric modelling and examines the probabilistic structure of the identification problem posed by the censoring of outcomes. The present paper began as an effort to address simultaneity in the same spirit, but I soon learned that my earlier analysis of selection is relevant more than in spirit. In fact, the problem of simultaneity is selection.

Let x be a specified point on the support of the covariate distribution and let t be a specified price. Consider the distribution $P[D(t)|x]$ of demand at price t among markets with covariates x. Write

$$(7) \quad P[D(t)|x] = P[D(t)|x,p=t]P(p=t|x) + P[D(t)|x,p\neq t]P(p\neq t|x).$$

By (1),

$$(8) \quad P[D(t)|x,p=t] = P(q|x,p=t).$$

Random sampling of (p,q,x) reveals $P(p=t|x)$, $P(p\neq t|x)$, and $P(q|x,p=t)$, but does not reveal $P[D(t)|x,p\neq t]$.

This is precisely the selection problem. The selection probability is $P(p=t|x)$. The censoring probability is $P(p\neq t|x)$. The distribution of outcomes conditional on selection is $P[D(t)|x,p=t]$, or $P(q|x,p=t)$ by (8). The distribution of outcomes conditional on censoring is $P[D(t)|x,p\neq t]$. One wishes to learn about $P[D(t)|x]$, the distribution of outcomes that would be observed if price were set equal to t in all markets with covariates x.

IDENTIFICATION IN THE ABSENCE OF PRIOR INFORMATION: In the absence of prior information on the structure of demand or the process of price determination, observations of outcomes in markets where $p \neq t$ reveal nothing about the censored demand distribution $P[D(t)|x,p\neq t]$. So we have this basic finding, which paraphrases Proposition 1 of Manski (1994):[2]

<u>Proposition 0</u>: Let (1) hold. Let Δ be the range of the demand function $D(\cdot)$. Let $f(\cdot)\colon \Delta \to R^1$ be a specified function with $K_0 \equiv \inf_u f(u)$ and $K_1 \equiv \sup_u f(u)$. Then

$$(9)\quad K_0P(p\neq t|x) + E[f(q)|x,p=t]P(p=t|x) \;\leq\; E\{f[D(t)]|x\} \;\leq\; K_1P(p\neq t|x) + E[f(q)|x,p=t]P(p=t|x).$$

In the absence of other information, this bound is sharp. ■

The bound (9) is always a subset of $[K_0,K_1]$, but its width increases as the selection probability $P(p=t|x)$ falls. In practice, the probability that price takes a particular value t is typically small, so the bound is typically wide. When price is continuously distributed, $P(p=t|x) = 0$ and the bound is the uninformative interval $[K_0,K_1]$. Thus, it seems appropriate to think of Proposition 0 as a negative finding showing that inference is infeasible in the absence of prior information about the structure of demand or the process of price determination.

4. Ordered Outcomes

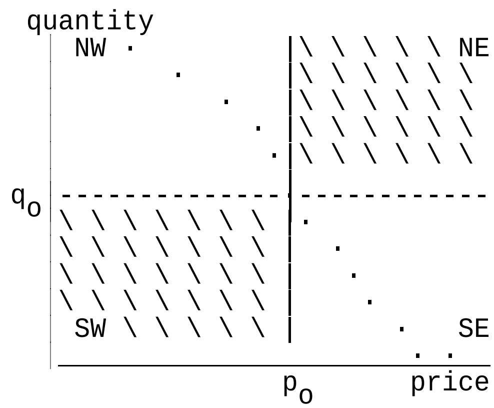
Given prior information about the structure of demand, observations of outcomes in markets where $p \neq t$ can be informative about the censored demand distribution $P[D(t)|x,p\neq t]$. If an economist is willing to assume anything about the structure of demand, it generally is that demand is weakly downward sloping; that is,

$$(10)\quad t' > t \Rightarrow D(t') \leq D(t).$$

Yet the literature on simultaneity has not studied the identifying power of this most common economic assumption. I do so here.

It is intuitive that the assumption of downward sloping demand should have identifying power. Consider Figure 1. If we know only that (1) holds, then observation of a market outcome (p_o,q_o) reveals only that $D(\cdot)$ is some function passing through (p_o,q_o). But if we know that (1) and (10) hold, then

Figure 1: Feasible Demand Functions



observation of (p_o,q_o) reveals that the downward sloping graph of $D(\cdot)$ must lie entirely within the northwest (NW) and southeast (SE) regions of the figure.

I have previously studied the simplest nontrivial case of assumption (10). The literature analyzing treatment effects supposes that a binary variable z determines which of two outcomes is observed; outcome y(1) is observed if z = 1 and outcome y(0) is observed if z = 0. Manski (1994, Section 6.1) examines the identifying power of the *ordered outcomes* assumption

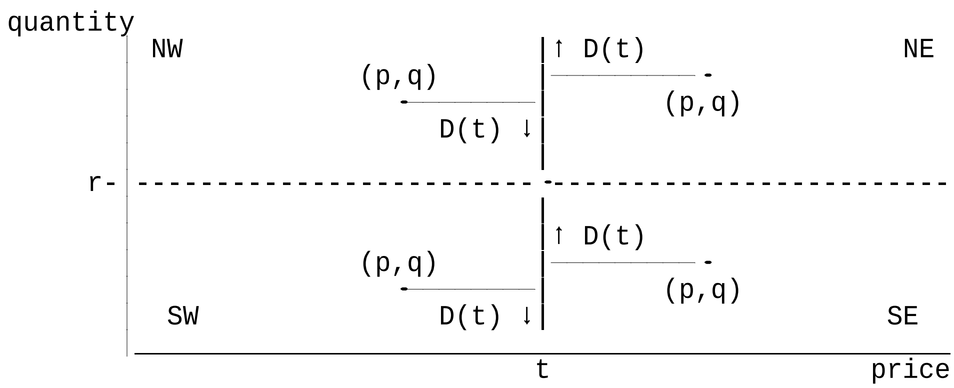
(11) $y(1) \le y(0)$.

Assumption (11) is the special case of (10) with t = 0 and t' = 1. Henceforth, I use the term *ordered outcomes* to refer to assumption (10) and not just to the special case (11).[3]

GRAPHICAL ANALYSIS: A simple way to see the identifying power of assumptions (1) and (10) is to consider the probability $P[D(t) \le r | x]$ that the demand at price t is smaller than a specified constant r. Figure 2 decomposes the possible market outcomes into four regions based on the position of (p,q) relative to (t,r).

Observe that the following holds in each region:

Figure 2: Identification of $P[D(t) \le r | x]$



SW region: $p \le t \cap q \le r \;\Rightarrow\; D(t) \le D(p) = q \le r$

SE region: $p > t \cap q \le r \;\Rightarrow\; D(t) \ge D(p) = q$

NW region: $p < t \cap q > r \;\Rightarrow\; D(t) \le D(p) = q$

NE region: $p \ge t \cap q > r \;\Rightarrow\; D(t) \ge D(p) = q > r.$

Each observation of (p,q) in the SW region implies that $D(t) \leq r$, so we may conclude that $P(p\leq t \cap q\leq r|x) \leq P[D(t)\leq r|x]$. Each observation in the NE region implies that $D(t) > r$, so $P[D(t)\leq r|x] \leq 1 - P(p\geq t \cap q>r|x) = P(p<t \cup q\leq r|x)$. Observations of (p,q) in the SE and NW regions do not reveal whether D(t) is less than or greater than r. So we have this sharp bound on $P[D(t)\leq r|x]$:[4]

(12) $P(p\leq t \cap q\leq r|x) \leq P[D(t)\leq r|x] \leq P(p<t \cup q\leq r|x)$.

The width of the bound depends on the distribution of market outcomes. At one extreme, the support of $P(p,q|x)$ may be concentrated in the SE and NW regions of Figure 2. Then (12) becomes $0 \leq P[D(t)\leq r|x] \leq 1$. At the other extreme, the support of $P(p,q|x)$ may be concentrated in the SW and NE regions. Then (12) becomes $P[D(t)\leq r|x] = P(p\leq t \cap q\leq r|x)$. So assumptions (1) and (10) may reveal nothing about $P[D(t)\leq r|x]$ or may identify this quantity.

THE ALGEBRA OF IDENTIFICATION: Let us now proceed more formally and abstractly. The main finding of this paper is a sharp bound on the mean of any monotone function of the demand at a specified price, as follows.

<u>Proposition 1</u>: Let (1) and (10) hold. Let $f(\cdot): \Delta \rightarrow R^1$ be a specified weakly increasing function. Let $K_0 \equiv \inf_u f(u)$ and $K_1 \equiv \sup_u f(u)$. Then

(13) $$K_0P(p<t|x) + E[f(q)|x,p\geq t]P(p\geq t|x) \leq E\{f[D(t)]|x\} \leq K_1P(p>t|x) + E[f(q)|x,p\leq t]P(p\leq t|x).$$

In the absence of other information, this bound is sharp. ■

PROOF: Write

$$(14)\quad E\{f[D(t)]|x\} = E\{f[D(t)]|x,p<t\}P(p<t|x) + E\{f[D(t)]|x,p=t\}P(p=t|x) + E\{f[D(t)]|x,p>t\}P(p>t|x).$$

By monotonicity of $D(\cdot)$ and $f(\cdot)$,

$$(15)\quad \begin{aligned} p < t &\Rightarrow K_0 \le f[D(t)] \le f(q) \\ &\Rightarrow K_0 \le E\{f[D(t)]|x,p<t\} \le E[f(q)|x,p<t]; \\ p = t &\Rightarrow f[D(t)] = f(q) \\ &\Rightarrow E\{f[D(t)]|x,p=t\} = E[f(q)|x,p=t]; \\ p > t &\Rightarrow f(q) \le f[D(t)] \le K_1 \\ &\Rightarrow E[f(q)|x,p>t] \le E\{f[D(t)]|x,p>t\} \le K_1. \end{aligned}$$

Inserting these lower and upper bounds in (14) yields (13). The bounds (15) are sharp, so the bounds (13) are sharp.

Q.E.D.

Proposition 1 has negative and positive implications. The negative implication concerns inference on $E\{f[D(t)]|x\}$ when f(.) has unbounded range. Inspection of (13) shows that the lower bound on $E\{f[D(t)]|x\}$ is $-\infty$ if $K_0 = -\infty$ and $P(p<t|x) > 0$. Similarly, the upper bound is ∞ if $K_1 = \infty$ and $P(p>t|x) > 0$.

The positive implication concerns inference when the range of f(.) is bounded. If $-\infty < K_0 \le K_1 < \infty$, the bound (13) has finite width

$$\{E[f(q)|x,p<t] - K_0\}P(p<t|x) + \{K_1 - E[f(q)|x,p>t]\}P(p>t|x).$$

The maximum width is $K_1 - K_0$, which occurs if $E[f(q)|x,p<t] = K_1$, $E[f(q)|x,p>t] = K_0$, and $P(p=t|x) = 0$. The minimum width is zero, which occurs if

$E[f(q)|p<t,x] = K_0$ and $E[f(q)|p>t,x] = K_1$. Thus, assumptions (1) and (10) may reveal nothing or may identify $E\{f[D(t)]|x\}$, depending on the distribution of market outcomes.

APPLICATIONS: One immediate application of Proposition 1 is to inference on $E[D(t)|x]$, the mean demand at price t in markets with covariates x. Suppose that demand is non-negative but unbounded from above, so $D(\cdot)$ has range $\Delta = [0,\infty)$. Let $f(\cdot)$ be the identity function, so $K_0 = 0$ and $K_1 = \infty$. If $P(p>t|x) > 0$, (13) reduces to

(16a) $E(q|x,p\geq t)P(p\geq t|x) \leq E[D(t)|x] \leq \infty.$

If $P(p>t|x) = 0$, (13) reduces to

(16b) $E(q|x,p=t)P(p=t|x) \leq E[D(t)|x] \leq E(q|x).$

Another immediate application is to inference on $P[D(t)>r|x]$, the probability that the demand at price t is larger than a specified constant r. Here $f(\cdot) = 1[\cdot > r]$, $K_0 = 0$, and $K_1 = 1$. So (13) reduces to

(17) $P(p\geq t \cap q>r|x) \leq P[D(t)>r|x] \leq P(p>t|x) + P(p\leq t \cap q>r|x).$

Given that $P[D(t)\leq r|x] = 1 - P[D(t)>r|x]$, we may also conclude that

(12)' $1 - P(p>t|x) - P(p\leq t \cap q>r|x) \leq P[D(t)\leq r|x] \leq 1 - P(p\geq t \cap q>r|x).$

This restates the result (12) obtained earlier using Figure 1.

As a final application, I use the bound (12) to derive bounds on quantiles of $P[D(t)|x]$. Let $\alpha \in (0,1)$ and let $q_\alpha[D(t)|x]$ denote the α-quantile of D(t) conditional on x. Corollary 1.1 shows that there is either an informative lower bound on $q_\alpha[D(t)|x]$ or an informative upper bound but not both.

<u>Corollary 1.1</u>: Let $\Delta = [0,\infty)$. If $P(p<t|x) \geq \alpha$, then

(18a) $0 \leq q_\alpha[D(t)|x] \leq \{\inf r \text{ s.t. } P(p\leq t \cap q\leq r|x) \geq \alpha\}$.

If $P(p\leq t|x) < \alpha$, then

(18b) $\{\sup r \text{ s.t. } P(p<t \cup q\leq r|x) < \alpha\} \leq q_\alpha[D(t)|x] < \infty$.

In the absence of other information, these bounds are sharp. ■

PROOF: By definition, $q_\alpha[D(t)|x] \equiv \inf r: P[D(t)\leq r|x] \geq \alpha$. To obtain the lower bound on $q_\alpha[D(t)|x]$, use (12) to find

(19) $P(p<t \cup q\leq r|x) < \alpha \Rightarrow P[D(t)\leq r|x] < \alpha \Rightarrow q_\alpha[D(t)|x] > r$.

Hence the greatest lower bound on $q_\alpha[D(t)|x]$ is $\{\sup r \text{ s.t. } P(p<t \cup q\leq r|x) < \alpha\}$. To obtain the upper bound, use (12) to find

(20) $P(p\leq t \cap q\leq r|x) \geq \alpha \Rightarrow P[D(t)\leq r|x] \geq \alpha \Rightarrow q_\alpha[D(t)|x] \leq r$.

Hence the least upper bound on $q_\alpha[D(t)|x]$ is $\{\inf r \text{ s.t. } P(p\leq t \cap q\leq r|x) \geq \alpha\}$. These bounds are sharp because the bound (12) is sharp.

If $P(p<t|x) \geq \alpha$, then $P(p<t \cup q\leq r|x) \geq \alpha$ for all r; so the greatest lower bound is zero. If $P(p\leq t|x) < \alpha$, then $P(p\leq t \cap q\leq r|x) < \alpha\}$ for all r; so the least upper bound is infinity.

Q.E.D.

CONDITIONING ON AN OBSERVED MARKET OUTCOME: Proposition 1 may be applied given any specification of the covariates x. Conditioning on an observed market outcome warrants special comment.

A common objective of empirical analysis is to infer the structure of demand in a market whose outcome is observed. Suppose one observes that, in a given market, a quantity q_o is transacted at a unit price p_o. Consider the problem of inference on demand conditional on this observed outcome and on some covariates w. Then the distribution of interest is $P[D(\cdot)|w,p=p_o,q=q_o]$.

At the beginning of this section we saw that, given assumptions (1) and (10), observation of (p_o,q_o) reveals $D(\cdot)$ to be a downward sloping function that lies entirely within the northwest and southeast regions of Figure 1. That discussion relied only on observation of (p_o,q_o), and did not bring to bear observations on outcomes in other markets. We may apply Proposition 1 to determine if stronger conclusions can be drawn when such observations are available.

The answer is negative. Applied to $P[D(\cdot)|w,p=p_o,q=q_o]$, the bound (13) becomes

$$(21a)\quad f(q_o) \;\le\; E\{f[D(t)]|w,p=p_o,q=q_o\} \;\le\; K_1 \qquad \text{for} \quad t < p_o,$$

$$(21b)\quad E\{f[D(t)]|w,p=p_o,q=q_o\} \;=\; f(q_o) \qquad \text{for} \quad t = p_o,$$

$$(21c)\quad K_0 \;\le\; E\{f[D(t)]|w,p=p_o,q=q_o\} \;\le\; f(q_o) \qquad \text{for} \quad t > p_o.$$

These are the same conclusions as may be drawn when only (p_o,q_o) is observed.

DEMAND AS A RANDOM PROCESS: Propositions 0 and 1 concern the distribution $P[D(t)|x]$ of demand at a specified price t. Viewed abstractly, the distributions $P[D(t)|x]$, $t \in [0,\infty)$ are the marginals of the random process $P[D(\cdot)|x]$ indexed by t. Knowledge of the marginals of this process suffices to answer many questions of economic interest. The marginals reveal the level of demand at any specified price. They also reveal $E[D(t_1) - D(t_0)|x]$, the expected difference between the demands at any two specified prices t_0 and t_1. Nevertheless, knowledge of the marginals does not give a complete description of the random demand process. To do that, we need to determine what the assumption of downward sloping demand implies about the dependency structure of $P[D(\cdot)|x]$. This question is left for future work.

Focussing as it has on the linear model (2), the econometric literature on simultaneity has been able to avoid explicit treatment of demand as a random process. In the linear model setting, the problem of identifying $P[D(\cdot)|x]$ reduces to that of identifying the scalar slope parameter ß. This extraordinary simplification of the inferential problem ceases to be available when assumption (2) is dropped in favor of (10).

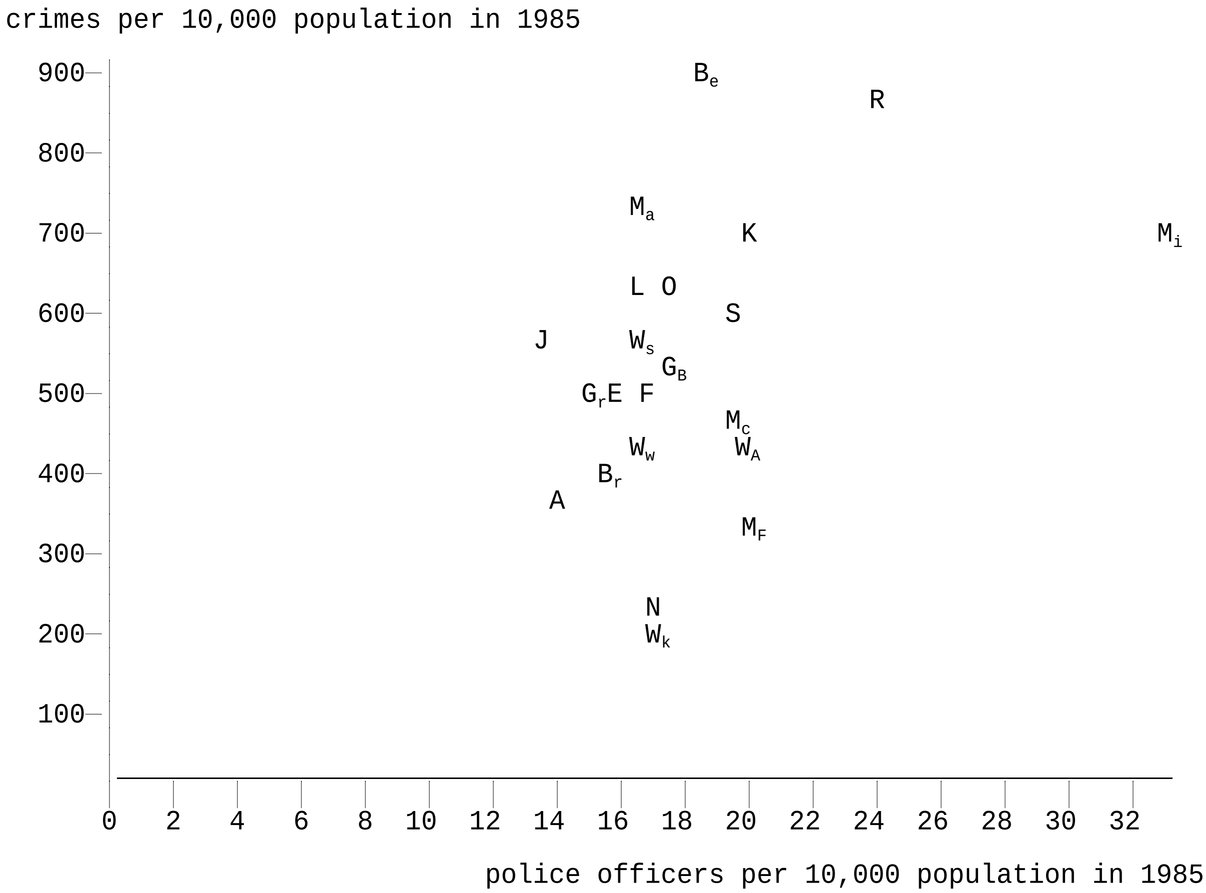
5. Empirical Illustration: The Effect of Policing on Crime

This section uses a well-known simultaneity problem of criminology to provide an empirical illustration of Proposition 1.

The problem arises when observations of crime rates and deterrence policies are used to study the effect of deterrence policies on criminal behavior. The standard setup assumes a set of isolated jurisdictions. Each jurisdiction has a crime function $D(\cdot)$ giving the crime rate that would occur if deterrence (i.e. the price of crime) were set at any level. Some process determines the actual deterrence level p in each jurisdiction. The realized crime rate is then $q = D(p)$. The problem is to learn about the distribution $P[D(.)|x]$ of crime functions among jurisdictions with covariates x.

Empirical analysis has centered on linear crime models of form (2), combined with exclusion restrictions of form (3). See Blumstein, Cohen, and Nagin (1978) for a review and critique of the literature. Our interest is to learn what inferences are possible if it is assumed that crime is a weakly

Figure 3: Crime Rates and Police Densities in Wisconsin Cities



Key
A - Appleton (14.0,379.2)
B_e - Beloit (18.5,906.0)
B_r - Brookfield (15.7,400.3)
E - Eau Claire (16.0,493.9)
F - Fond du Lac (17.2,487.2)
G_B - Green Bay (18.2,540.7)
G_r - Greenfield (15.1,508.4)
J - Janesville (13.4,561.5)
K - Kenosha (19.8,690.4)
L - La Crosse (17.2,647.1)
M_a - Madison (17.2,722.5)
M_c - Manitowoc (19.4,458.2)
M_F - Menomonee Falls (20.3,341.1)
M_i - Milwaukee (32.8,706.6)
N - New Berlin (17.0,222.6)
O - Oshkosh (17.3,631.5)
R - Racine (24.1,857.2)
S - Sheboygan (19.2,601.3)
W_k - Waukesha (17.1,221.2)
W_s - Wausau (16.5,552.9)
W_w - Wauwatosa (17.5,450.8)
W_A - West Allis (20.2,425.6)

Source: U.S. Bureau of the Census, County and City Data Book 1988, Table C.

decreasing function of the deterrence level.

MEASURING CRIME AND DETERRENCE: Empirical studies have measured crime and deterrence in many different ways. To provide an accessible illustration, I use data for American cities with population over 25,000 published in U.S. Census Bureau, County and City Data Book 1988. The crime rate is an FBI estimate of the number of serious crimes committed per 100,000 resident population in the year 1985 (Table C, item 31). The deterrence measure is an FBI estimate of the number of police officers per 10,000 resident population in 1985 (Table C, item 33). I use these crime rates and police densities as reported, except that I rescale the crime rate to have the same population base as the police density.

I focus on the twenty-two cities in my own state of Wisconsin.[5] These observations form a dataset large enough to yield interesting findings, but small enough to permit the reader to follow the calculations easily. Figure 1 presents the raw data and displays the configuration of police densities and crime rates.

ANALYSIS: Assume that the crime functions in these Wisconsin cities are a random sample of size N = 22 drawn from the distribution $P[D(\cdot)|x]$ of crime functions in jurisdictions sharing specified covariates x.[6] Let $P_N(p,q|x)$ denote the empirical distribution of (police density, crime rate) outcomes. Then consistent nonparametric estimates of the bounds proved in Proposition 1 and in Corollary 1.1 may be obtained by replacing the features of $P(p,q|x)$ appearing in (13) and (18) with the corresponding features of $P_N(p,q|x)$. For example, consider $P[D(t)\le r|x]$. The lower bound is $P(p\le t \cap q\le r|x)$ and its estimate is the sample frequency of the event $(p\le t \cap q\le r)$.

Table I presents a set of such estimates. The left part of the table gives estimated bounds on $P[D(t)\le r|x)]$ at various levels of (t,r). The right part gives estimated bounds on the mean crime rate $E[D(t)|x]$ and median crime rate $M[D(t)|x]$ at various levels of t. To keep the discussion centered on the problem of identification, I shall discuss the table as if its entries are the bounds rather than consistent estimates thereof. One may address questions of sampling precision by placing confidence bands around the estimated bounds, much as in the empirical study of the selection problem conducted by Manski et al. (1992).

Table I: Estimated Bounds on Features of $P[D(\cdot)|x]$

	$P[D(t)\le r\|x]$				$E[D(t)\|x]$	$M[D(t)\|x]$
	r = 200	r = 400	r = 600	r = 800		
t = 12	[.00,.00]	[.00,.18]	[.00,.64]	[.00,.91]	[536.6,∞)	[493.9,∞)
t = 16	[.00,.23]	[.05,.36]	[.23,.64]	[.23,.91]	[452.6,∞)	[458.2,∞)
t = 20	[.00,.82]	[.14,.86]	[.55,.91]	[.77,.95]	[105.9,∞)	[0,552.9]
t = 24	[.00,.91]	[.18,.91]	[.64,.91]	[.86,.95]	[71.1,∞)	[0,540.7]

Source: Figure 3, Proposition 1, and Corollary 1.1.

The most striking feature of the bounds on $P[D(t)\le r|x)]$ is how much they vary in width as (t,r) varies. We are able to draw only very weak inferences when (t,r) = (12,800) and (t,r) = (24,200). In these cases, the table entries are $P[D(12)\le 800|x] \in [.00,.91]$ and, likewise, $P[D(24)\le 200|x] \in [.00,.91]$. We can, however, draw rather strong inferences when (t,r) = (12,200) and (t,r) = (24,800). Here $P[D(12)\le 200|x] = 0$ and $P[D(24)\le 800|x] \in [.86,.95]$. That is, abstracting from questions of sampling precision, we find that setting the police density equal to 12 officers per 10,000 population always yields a crime rate larger than 200 crimes per 10,000 population. Setting the police density equal to 24 officers per 10,000 population has at least a .86 chance of yielding a crime rate less than or equal to 800 crimes per 10,000 population.

My earlier analysis of the selection problem emphasized that identifying the mean and median of a distribution pose fundamentally different problems when outcomes are censored (see Manski, 1994). Inspection of the bounds on

$E[D(t)|x]$ and $M[D(t)|x]$ at $t = 20$ illustrates this difference well. The bound on the mean crime rate is $[105.9,\infty)$ and the bound on the median crime rate is $[0,552.9]$. Thus setting the police density equal to 20 would yield a mean crime rate of at least 105.9 and a median crime rate no larger than 552.9.

6. Exclusion Restrictions

Our analysis of demand has imposed no assumptions on the process of price determination. If an economist is willing to assume anything about price determination, it probably is that some covariate affects price but not demand. In the linear model setting, equation (3) formalizes the usual exclusion restriction. In the present more abstract setting, we might express an exclusion restriction by assuming that

$$(22a)\quad P[D(\cdot)|x=x_i] \;=\; P[D(\cdot)|x=x_j]$$
$$(22b)\quad P(p|x=x_i) \;\neq\; P(p|x=x_j).$$

Or we might impose the weaker assumption that

$$(23a)\quad E\{f[D(t)]|x=x_i\} \;=\; E\{f[D(t)]|x=x_j\}$$
$$(23b)\quad P(p|x=x_i) \;\neq\; P(p|x=x_j).$$

where $f(\cdot)\colon \Delta \to R^1$ is a specified function.

What is the identifying power of assumption (22) or (23)? There is a ready answer to the latter question. Proposition 6 of Manski (1994) shows that if (23) holds, the bound in Proposition 0 may be replaced by its intersection across the two covariate values x_i and x_j. That is,

$$(24)\quad \sup_{k=i,j} \{K_0P(p\neq t|x=x_k) + E[f(q)|x=x_k,p=t]P(p=t|x=x_k)\} \leq E\{f[D(t)]|x=x_i\}$$
$$\leq \inf_{k=i,j} \{K_1P(p\neq t|x=x_k) + E[f(q)|x=x_k,p=t]P(p=t|x=x_k)\}$$

is the sharp bound on $E\{f[D(t)]|x=x_i\}$. The same reasoning implies that if (23) holds and $f(\cdot)$ is a monotone function, the bound in Proposition 1 may be replaced by its intersection, namely

$$(25)\quad \sup_{k=i,j} \{K_0P(p<t|x=x_k) + E[f(q)|x=x_k,p\geq t]P(p\geq t|x=x_k)\} \leq E\{f[D(t)]|x=x_i\}$$
$$\leq \inf_{k=i,j} \{K_1P(p>t|x=x_k) + E[f(q)|x=x_k,p\leq t]P(p\leq t|x=x_k)\}.$$

The identifying power of the stronger exclusion restriction (22) is an open, and seemingly difficult, question. It is also an important question. Economists willing to assert exclusion restrictions often are willing to do so in the stronger form of (21), not just in the weaker form of (22).

7. Conclusion

Analysis of simultaneity was so central to the early development of econometrics that it was long common for econometricians to think of identification and simultaneity as synonymous. Many econometrics texts, even recent ones, discuss identification only in the context of simultaneity. Particularly revealing is the title chosen by Fisher (1966) for his monograph on the simultaneity problem. He titled the book *The Identification Problem in Econometrics* and justified this choice by writing in the preface (page vii):

> Because the simultaneous equation context is by far the most important one in which the identification problem is encountered, the treatment is restricted to that context.

From today's perspective, Fisher's judgment of the preeminence of simultaneity among all identification problems seems strained. Nevertheless, simultaneity certainly remains an important problem of econometrics.

Econometric thinking about simultaneity has long focussed on the linear model (2). The present analysis opens a new way of thinking by showing that simultaneity is a problem of censored outcomes. Proposition 1 demonstrates how fruitful this way of thinking can be.

This paper opens a long agenda for future research. In Section 4, I raised the question of inference on the dependency structure of demand. In Section 6, I observed that we do not know the identifying power of exclusion restrictions of the form (22). Here, I would add this: we should seek to subsume the assumptions of downward sloping demand and linear demand within a more general study of the identifying power of restrictions on the shape of the demand function.

To assume downward sloping demand assumption is to impose a specific bound on the first derivative of the demand function, or a specific Lipschitz condition. To assume linear demand is to impose a specific value for the second derivative of the demand function, or a specific concavity/convexity assumption. Learning the identifying power of other shape restrictions would widen the options available to empirical researchers analyzing simultaneity problems.

Notes

1. The standard setup does not assert (3) directly, but rather invokes assumptions that imply (3). These assumptions are

(a) firms are price takers with linear supply functions $S(t) = bt + e$;

(b) market outcomes satisfy the equilibrium condition $q = D(p) = S(p)$

(c) $E(\epsilon,e|x) = (x'\alpha,x'a)$ for some parameters (α,a).

Condition (3a) is expressed by assuming that some component of α equals zero. Condition (3b) holds if the corresponding component of a is non-zero.

2. The proof is simple. Write

$$E\{f[D(t)]|x\} = E\{f[D(t)]|x,p\neq t\}P(p\neq t|x) + E(q|x,p=t)P(p=t|x).$$

The sampling process reveals all the quantities on the right side except for $E\{f[D(t)]|x,p\neq t\}$. In the absence of prior information, we know only that $E\{f[D(t)]|x,p\neq t\} \in [K_0,K_1]$.

3. Much of the literature on treatment effects invokes the *shifted outcomes* or *constant-treatment-effect* assumption; that is, $y(1) = y(0) + \beta$, where β is a parameter. This assumption is the simplest nontrivial case of the linear model (2), obtained by evaluating (2) at $t = 0$ and at $t = 1$.

4. A complementary result holds for upward sloping functions. Assume $D(\cdot)$ to be upward sloping rather than downward sloping. Then observations in the SE and NW regions are informative, but those in the SW and NE regions are not.

5. County and City Data Book 1988 lists twenty three Wisconsin cities as having population over 25,000 but statistics are not reported for one of these, the city of Superior.

6. Specification of the covariates x should reflect one's beliefs about the variation in crime functions across cities. For example, suppose one believes that the distribution of crime functions across Wisconsin cities is the same as the distribution of crime functions across all midwestern cities. Then one can let [x = midwestern cities] and use the Wisconsin data to infer the distribution of crime functions across all midwestern cities.

References

Blumstein, A., J. Cohen, and D. Nagin (editors) (1978), Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions on Crime Rates, Washington, D.C.: National Academy Press.

Fisher, F. (1966), The Identification Problem in Econometrics, New York: McGraw-Hill.

Hood, W. and T. Koopmans (editors) (1953), Studies in Econometric Method, New York: Wiley.

Manski, C. (1989), "Anatomy of the Selection Problem," Journal of Human Resources, 24, 343-360.

Manski, C. (1990), "Nonparametric Bounds on Treatment Effects," American Economic Review Papers and Proceedings, 80: 319-323.

Manski, C. (1994), "The Selection Problem," in C. Sims (editor) Advances in Econometrics: Sixth World Congress, Cambridge, England: Cambridge University Press.

Manski, C., G. Sandefur, S. McLanahan, and D. Powers (1992), "Alternative Estimates of the Effect of Family Structure During Adolescence on High School Graduation," Journal of the American Statistical Association, 87: 25-37.

U.S. Bureau of the Census, County and City Data Book 1988, Washington, DC: U.S. Government Printing Office.