2

Instrumental Variables

2.1. Distributional Assumptions and Credible Inference

Distributional assumptions may enable one to shrink identification regions obtained using empirical evidence alone. When facing the problem of missing outcome data, researchers have generally imposed distributional assumptions that point-identify the outcome distribution $P(y)$. When a single random sampling process generates the available data, it has been particularly common to assert that observed and missing outcomes have the same distribution; that is,

$$P(y) = P(y|z=0) = P(y|z=1). \tag{2.1}$$

The distribution $P(y|z=1)$ is revealed by the sampling process, so $P(y)$ is point-identified. Someone asserting (2.1) cannot be proved wrong; after all, the empirical evidence reveals nothing about $P(y|z=0)$.

An assumption may be non-refutable and yet not credible. Researchers who assert (2.1) almost inevitably find this assumption difficult to justify. Analysts who assert other point-identifying assumptions regularly encounter the same difficulty. This should not be surprising. The empirical evidence reveals nothing at all about the distribution of missing data. An assumption must be quite strong to pick out one among all possible distributions.

There is a fundamental tension between the credibility and strength of conclusions, which I have called the *Law of Decreasing Credibility*. Inference using the empirical evidence alone sacrifices strength of conclusions in order to maximize credibility. Inference invoking point-identifying distributional assumptions sacrifices credibility in order to

achieve strong conclusions. Between these poles, there is a vast middle ground of possible modes of inference asserting assumptions that may shrink the identification region $H[P(y)]$ but not reduce it to a point.

This chapter examines the identifying power of various distributional assumptions that make use of instrumental variables. Some such assumptions imply point identification, whereas others have less identifying power and, perhaps, greater credibility. For simplicity, the analysis below presumes that a single random sampling process generates the available data, that realizations of y are either completely observed or entirely missing, and that all realizations of the instrumental variable are observed. Distributional assumptions using instrumental variables may also be applied when data are available from multiple sampling processes, when interval measures of outcomes are observed, and when some realizations of the instrumental variable are missing.

2.2. Some Assumptions Using Instrumental Variables

As in Chapter 1, suppose that a sampling process draws persons at random from population J and that the outcome y is observable if $z = 1$. Moreover, suppose now that each person j is characterized by a covariate v_j in a space V. Let v: J → V be the random variable mapping persons into covariates and let $P(y, z, v)$ denote the joint distribution of (y, z, v). Suppose that all realizations of v are observable. Observability of v provides an instrument or tool that may help to identify the outcome distribution $P(y)$. Thus v is said to be an *instrumental variable*.

The sampling process asymptotically reveals the distributions $P(z)$, $P(y, v|z = 1)$, and $P(v|z = 0)$. It is uninformative about the conditional distributions $[P(y|v = v, z = 0), v \in V]$. The presence of an instrumental variable does not, per se, help to identify $P(y)$. However, observability of v may be useful when combined with distributional assumptions. This chapter examines the identifying power of six such assumptions.

Sections 2.3 and 2.4 study identification of $P(y)$ under assumptions that assert forms of statistical independence among the random variables (y, z, v). Section 2.3 assumes that observed and missing outcomes have the same distribution conditional on v; that is, outcomes are *missing-at-random* conditional on v.

Outcomes Missing-at-Random (Assumption MAR):

$$P(y|v) = P(y|v, z=0) = P(y|v, z=1). \tag{2.2}$$

Section 2.4 assumes that y is statistically independent of v; that is:

Statistical Independence of Outcomes and Instruments (Assumption SI):

$$P(y|v) = P(y). \tag{2.3}$$

Section 2.5 studies identification of the expectation $E[g(y)]$ of a real-valued function $g(y)$ under assumptions that are weaker than Assumptions MAR and SI. First the forms of statistical independence asserted in (2.2) and (2.3) are weakened to the mean-independence assumptions

Means Missing-at-Random (Assumption MMAR):

$$E[g(y)|v] = E[g(y)|v, z=0] = E[g(y)|v, z=1] \tag{2.4}$$

and

Mean Independence of Outcomes and Instruments (Assumption MI):

$$E[g(y)|v] = E[g(y)], \tag{2.5}$$

respectively. Then, Assumptions MMAR and MI are weakened to the monotonicity assumptions

Mean Missing Monotonically (Assumption MMM):

$$E[g(y)|v, z=1] \geq E[g(y)|v] \geq E[g(y)|v, z=0] \tag{2.6}$$

and

Mean Monotonicity of Outcomes and Instruments (Assumption MM): Let V be an ordered set.

$$E[g(y)|v=\mathrm{v}] \geq E[g(y)|v=\mathrm{v}'], \forall (\mathrm{v}, \mathrm{v}') \in \mathrm{V} \times \mathrm{V} \text{ such that } \mathrm{v} \geq \mathrm{v}'. \tag{2.7}$$

Taken together, these six distributional assumptions provide a variety of ways in which a researcher can use instrumental variables to help identify outcome distributions when some outcome data are missing. Researchers contemplating the use of instrumental variables should, of course, pay due attention to the credibility of these and other assumptions. Empirical researchers often ask whether some observable covariate is or is not a "valid instrument" in an application of interest. The expression "valid instrument"

is imprecise because it focuses attention on the covariate used in the role of v. Credibility depends not on the covariate per se but on the assumption that the distribution $P(y, z, v)$ is assumed to satisfy.

To simplify the presentation, the analysis below supposes that the covariate space V is finite and that $P(v=\mathrm{v}, z=1) > 0$ for all $\mathrm{v} \in \mathrm{V}$. These regularity conditions are maintained without further reference.

2.3. Outcomes Missing-at-Random

Assumption MAR is a non-refutable hypothesis that point-identifies $P(y)$. Proposition 2.1 shows how.[1]

Proposition 2.1: Let assumption MAR hold. Then $P(y)$ is point-identified with

$$P(y) = \textstyle\sum_{\mathrm{v} \in \mathrm{V}} P(y|v=\mathrm{v}, z=1)P(v=\mathrm{v}). \tag{2.8}$$

Assumption MAR is non-refutable. □

Proof: The Law of Total Probability gives

$$P(y) = \textstyle\sum_{\mathrm{v} \in \mathrm{V}} P(y|v=\mathrm{v})P(v=\mathrm{v}). \tag{2.9}$$

Assumption MAR states that

$$P(y|v) = P(y|v, z=1). \tag{2.10}$$

Applying (2.10) to (2.9) yields (2.8). The right side of (2.8) is point-identified by the sampling process, so $P(y)$ is point-identified. Assumption MAR is non-refutable because the empirical evidence reveals nothing about $P(y|v, z=0)$.

Q. E. D.

A researcher applying assumption MAR must specify the instrumental variable v for which the assumption holds. Assumption (2.1) is the special case in which v has a degenerate distribution. As in that case, the credibility of assumption MAR is regularly a matter of controversy.[2]

2.4. Statistical Independence

Assumption SI has the same identifying power as does observation of data from multiple sampling processes. The space V of values for the instrumental variable plays the same role here as did the set M of sampling processes in Section 1.4. Proposition 2.2 gives the basic result, and two corollaries flesh it out.

Proposition 2.2: (a) Let assumption SI hold. Then the identification region for $P(y)$ is

$$H_{SI}[P(y)] = \bigcap_{v \in V} \{P(y|v=v, z=1)P(z=1|v=v) + \gamma_v \cdot P(z=0|v=v),\ \gamma_v \in \Gamma_Y\}. \tag{2.11}$$

(b) Let the set $H_{SI}[P(y)]$ be empty. Then assumption SI does not hold. □

Proof: (a) Application of equation (1.2) to each conditional distribution $P(y|v=v)$, $v \in V$ gives the identification region for this distribution using the empirical evidence alone; that is,

$$H[P(y|v=v)] = [P(y|v=v, z=1)P(z=1|v=v) + \gamma_v \cdot P(z=0|v=v),\ \gamma_v \in \Gamma_Y]. \tag{2.12}$$

Moreover, the identification region for the set of distributions $[P(y|v=v), v \in V]$ is the Cartesian product $\times_{v \in V} H[P(y|v=v)]$.

Assumption SI states that the distributions $P(y|v=v)$, $v \in V$ coincide, all being equal to $P(y)$. Hence $P(y)$ must lie in $\cap_{v \in V} H[P(y|v=v)]$. Any distribution in this intersection is feasible, so $H_{SI}[P(y)]$ is the identification region.

(b) If assumption SI holds, the set $H_{SI}[P(y)]$ is necessarily non-empty. Hence, the assumption cannot hold if $H_{SI}[P(y)]$ is empty.

Q. E. D.

Part (a) of the proposition shows that the identifying power of assumption SI can range from point identification of $P(y)$ to no power at all, depending on the nature of the instrumental variable. Point identification occurs if there exists a $v \in V$ such that $P(z=1|v=v)$; then one of the sets whose intersection is taken in (2.11) is a singleton. When Y is countable, Corollary 2.2.1 below gives a simple necessary and sufficient condition for point identification.

Assumption SI has no identifying power if (a) z is statistically independent of v and (b) y is statistically independent of v conditional on the event $\{z = 1\}$; that is, if $P(z|v) = P(z)$ and $P(y|v, z = 1) = P(y|z = 1)$. Then $H[P(y|v=v)]$, $v \in V$ are all the same as the identification region obtained using the empirical evidence alone. This shows that identification cannot be achieved by construction of a trivial instrumental variable that uses a randomization device to assign a covariate value to each member of the population. A covariate v generated by a randomization device is necessarily statistically independent of the pair (y, z). Such a covariate satisfies assumption SI but has no identifying power.

Part (b) of the proposition shows that assumption SI is refutable. If $H_{SI}[P(y)]$ is empty, the assumption logically cannot hold. Of course, non-emptiness of $H_{SI}[P(y)]$ does not imply that the assumption is correct.

Observe that the identification region $H_{SI}[P(y)]$ has the same structure as the region $H_M[P(y)]$ obtained by combining data from multiple sampling processes (see Proposition 1.3), with V here playing the role of M there. Hence, there are instrumental-variable analogs to Corollaries 1.3.1 and 1.3.2. These are given in Corollaries 2.2.1 and 2.2.2 below. The proofs are analogous to those of the earlier corollaries and so are omitted.

Corollary 2.2.1: Let assumption SI hold. Let $\eta \in \Gamma_Y$. For $B \subset Y$, define

$$\pi_V(B) \equiv \max_{v \in V} P(y \in B|v=v, z=1)P(z=1|v=v). \tag{2.13}$$

(a) Then $\eta \in H_{SI}[P(y)]$ if and only if $\eta(B) \geq \pi_V(B)$, $\forall B \subset Y$.

(b) Let Y be countable. Then $\eta \in H_{SI}[P(y)]$ if and only if $\eta(y) \geq \pi_V(y)$, $\forall y \in Y$.

(c) Let Y be countable. Let $S_V \equiv \sum_{y \in Y} \pi_V(y)$. Then $H_{SI}[P(y)]$ contains multiple distributions if $S_V < 1$ and a unique distribution if $S_V = 1$. If $S_V = 1$, the unique feasible distribution is $\eta_V(y) \equiv \pi_V(y)$, $y \in Y$. If $S_V > 1$, then assumption SI does not hold. □

Corollary 2.2.2: Let assumption SI hold. Let Y be a countable subset of R, and let Y contain its lower and upper bounds $y_0 \equiv \inf_{y \in Y}$ and $y_1 \equiv \sup_{y \in Y}$. Let η_0 and η_1 be probability distributions on Y such that, for each $y \in Y$,

$$\eta_0(y) = \pi_V(y) \text{ if } y > y_0 \text{ and } \eta_0(y_0) = \pi_V(y_0) + (1 - S_V) \quad (2.14a)$$

$$\eta_1(y) = \pi_V(y) \text{ if } y < y_1 \text{ and } \eta_1(y_1) = \pi_V(y_1) + (1 - S_V). \quad (2.14b)$$

(a) Let $D(\cdot)$ respect stochastic dominance. Then the smallest and largest elements of $H_{SI}\{D[P(y)]\}$ are $D(\eta_0)$ and $D(\eta_1)$.

(b) The closed interval

$$H_{SI}[E(y)] = [\textstyle\sum_{y \in Y} y\pi_V(y) + (1 - S_V)y_0,\ \sum_{y \in Y} y\pi_V(y) + (1 - S_V)y_1] \quad (2.15)$$

is the identification region for $E(y)$. □

2.5. Mean Independence and Mean Monotonicity

This section studies identification of the expectations of real-valued functions of the outcome. The distributional assumptions considered here are weaker than Assumptions MAR and SI. Throughout this section, $g(\cdot)$ is a real-valued function that attains its lower and upper bounds.

Mean Independence

Assumptions MMAR and MI weaken the forms of statistical independence asserted in assumptions MAR and SI to corresponding forms of mean independence. Assumption MMAR is a non-refutable hypothesis that point-identifies $E[g(y)]$. Assumption MI is a refutable hypothesis that generically shrinks the identification region obtained using the empirical evidence alone, but point-identifies $E[g(y)]$ only in special cases. Propositions 2.3 and 2.4 give the results.

Proposition 2.3: Let assumption MMAR hold. Then $E[g(y)]$ is point-identified with

$$E[g(y)] = \textstyle\sum_{v \in V} E[g(y)|v = v, z = 1]P(v = v). \quad (2.16)$$

Assumption MMAR is non-refutable. □

Proof: The Law of Iterated Expectations gives

$$E[g(y)] = \textstyle\sum_{v \in V} E[g(y)|v = v]P(v = v). \quad (2.17)$$

Assumption MMAR states that

$$E[g(y)|v] = E[g(y)|v, z = 1]. \quad (2.18)$$

Applying (2.18) to (2.17) yields (2.16). The right side of (2.16) is point-identified by the sampling process, so $E[g(y)]$ is point-identified. Assumption MMAR is non-refutable because the empirical evidence reveals nothing about $E[g(y)|v, z = 0]$.

Q. E. D.

Proposition 2.4: (a) Let assumption MI hold. Then the closed interval

$$H_{MI}\{E[g(y)]\} =$$

$$[\max_{v \in V} E[g(y)z + g_0(1 - z)|v = v],\ \min_{v \in V} E[g(y)z + g_1(1 - z)|v = v]]. \quad (2.19)$$

is the identification region for $E[g(y)]$.

(b) Let $H_{MI}\{E[g(y)]\}$ be empty. Then assumption MI does not hold. □

Proof: (a) Application of equation (1.9′) to each conditional expectation $E[g(y)|v = v]$, $v \in V$ gives its identification region using the empirical evidence alone; that is, the closed interval

$$H\{E[g(y)|v = v]\} = [E[g(y)z + g_0(1 - z)|v = v], E[g(y)z + g_1(1 - z)|v = v]]. \quad (2.20)$$

Moreover, the identification region for $\{E[g(y)|v = v], v \in V\}$ is the $|V|$-dimensional rectangle $\times_{v \in V} H\{E[g(y)|v = v]\}$.

Assumption MI states that the expectations $E[g(y)|v = v]$, $v \in V$ coincide, all being equal to $E[g(y)]$. Hence $E[g(y)]$ must lie in $\cap_{v \in V} H\{E[g(y)|v = v]\}$. Any value in this set is feasible, so $H_{MI}\{E[g(y)]\}$ is the identification region.

(b) If assumption MI holds, the set $H_{MI}\{E[g(y)]\}$ is necessarily non-empty. Hence, the assumption cannot hold if $H_{MI}\{E[g(y)]\}$ is empty.

Q. E. D.

As with assumption SI, the identifying power of assumption MI can range from point identification to no power at all, depending on the nature of the instrumental variable. Point identification of $E[g(y)]$ occurs if there exists a $v \in V$ such that $P(z = 1 \mid v = v)$; then $E[g(y)] = E[g(y) \mid v = v]$. There is no identifying power if the pair (y, z) is statistically independent of v. Then $H_{MI}\{E[g(y)]\} = H\{E[g(y)]\}$.

Mean Monotonicity

Although mean independence is a weaker property than statistical independence, empirical researchers often find that assertions of mean independence are still too strong to be credible. There is therefore reason to ask whether Assumptions MMAR and MI may be weakened in ways that enhance credibility while preserving some identifying power. A simple way to do this is to change the equalities in equations (2.4) and (2.5) to the weak inequalities in equations (2.6) and (2.7).

Weakening assumption MMAR in this way yields assumption MMM, which asserts that, for each realization of v, the mean value of $g(y)$ when y is observed is greater than or equal to the mean value of $g(y)$ when y is missing. (The direction of the inequality can be reversed by applying the assumption to the function $-g(y)$.) Weakening assumption MI in this way yields assumption MM, which presumes that the set V has been pre-ordered. Propositions 2.5 and 2.6 characterize the identifying power of these monotonicity assumptions.

Proposition 2.5: Let assumption MMM hold. Then the identification region for $E[g(y)]$ is the closed interval

$$H_{MMM}\{E[g(y)]\} =$$

$$[E[g(y) \mid z = 1]P(z = 1) + g_0 P(z = 0),\ \textstyle\sum_{v \in V} E[g(y) \mid v = v, z = 1]P(v = v)]. \tag{2.21}$$

Assumption MMM is non-refutable. □

Proof: Let $v \in V$. Under assumption MMM, the identification region for $E[g(y) \mid v = v, z = 0]$ is the closed interval

$$H_{MMM}\{E[g(y) \mid v = v, z = 0]\} = [g_0, E[g(y) \mid v = v, z = 1]]. \tag{2.22}$$

Moreover, the joint identification region for $\{E[g(y) \mid v = v, z = 0], v \in V\}$

is the $|V|$-dimensional rectangle $\times_{v \in V} H_{MMM}\{E[g(y) \mid v = v, z = 0]\}$. The Law of Iterated Expectations gives

$$E[g(y)] = \textstyle\sum_{v \in V} E[g(y) \mid v = v, z = 1]P(v = v, z = 1)$$
$$+ E[g(y) \mid v = v, z = 0]P(v = v, z = 0). \tag{2.23}$$

Applying (2.22) to (2.23) yields (2.21). Assumption MMM is non-refutable because the empirical evidence reveals nothing about $\{E[g(y) \mid v = v, z = 0], v \in V\}$.

Q. E. D.

Proposition 2.6: (a) Let V be an ordered set. Let assumption MM hold. Then the identification region for $E[g(y)]$ is the closed interval

$$H_{MM}\{E[g(y)]\} = [\ \sum_{v \in V} P(v = v)\{\max_{v' \le v} E[g(y)z + g_0(1 - z) \mid v = v']\},$$
$$\sum_{v \in V} P(v = v)\ \{\min_{v' \ge v} E[g(y)z + g_1(1 - z) \mid v = v']\}]. \tag{2.24}$$

(b) Let $H_{MM}\{E[g(y)]\}$ be empty. Then assumption MM does not hold. □

Proof: (a) The proof to Proposition 2.4 showed that, using the empirical evidence alone, the identification region for the expectations $\{E[g(y) \mid v = v], v \in V\}$ is the $|V|$-dimensional rectangle $\times_{v \in V} H\{E[g(y) \mid v = v]\}$. Under assumption MM, a point $d \in R^V$ belongs to the identification region for $\{E[g(y) \mid v = v], v \in V\}$ if and only if d is an element of this rectangle whose components $(d_1, d_2, \ldots, d_{|V|})$ form a weakly increasing sequence. Applying this to the Law of Iterated Expectations (2.17) yields (2.24).

(b) If assumption MM holds, the set $H_{MM}\{E[g(y)]\}$ is necessarily non-empty. Hence, the assumption cannot hold if $H_{MM}\{E[g(y)]\}$ is empty.

Q. E. D.

Proposition 2.5 shows that, under assumption MMM, the identification region for $E[g(y)]$ is a right-truncated subset of the region obtained using the empirical evidence alone. The smallest feasible value of $E[g(y)]$ is the same as when using the empirical evidence alone. The largest is the value that $E[g(y)]$ would take under assumption MMAR.

Proposition 2.6 shows that the identification region under assumption MM is a subset of the region obtained using the empirical evidence alone and a superset of the one obtained under assumption MI. The identifying power of assumption MM depends on how the regions $\{H\{E[g(y)|v=v]\}, v \in V\}$ vary with v. The extreme possibilities occur if this sequence of intervals shifts to the left or right as v increases. In the former case, the identification region under assumption MM is the same as under assumption MI. In the latter case, assumption MM has no identifying power.

2.6. Other Assumptions Using Instrumental Variables

This chapter has examined assumptions that help to identify outcome distributions when an instrumental variable is observed. The tension between the credibility and strength of conclusions is especially evident as one weakens assumption SI to assumption MI and then to assumption MM. Each successive assumption is more plausible but has less identifying power.

It is easy to think of other assumptions that make different tradeoffs between credibility and identifying power. For example, assumption MI could be weakened not to the monotonicity asserted in assumption MM but rather to some form of "approximate" mean independence. A way to formalize this is to assert that, for all pairs $(v, v') \in V \times V$,

$$\|E[g(y)|v=v'] - E[g(y)|v=v]\| \leq C, \tag{2.25}$$

where $C > 0$ is a specified constant. Recall that the empirical evidence alone restricts the vector of expectations $E[g(y)|v]$ to the $|V|$-dimensional rectangle $\times_{v \in V} H\{E[g(y)|v=v]\}$. Relationship (2.25) further restricts $E[g(y)|v]$ to points in $R^{|V|}$ that satisfy specified linear inequalities.

Alternatively, assumption MI could be weakened to the zero-covariance assumption

$$E[g(y)\cdot v] - E[g(y)]E(v) = 0, \tag{2.26}$$

which may be rewritten as

$$\sum_{v \in V} P(v=v)\,[v - E(v)]E[g(y)|v=v] = 0. \tag{2.27}$$

The empirical evidence point-identifies $E(v)$. Hence, equation (2.27) establishes a linear constraint among the elements of $E[g(y)|v]$.

Complement 2A. Estimation with Nonresponse Weights

Organizations conducting major surveys commonly release public data files that provide *nonresponse weights* to be used for estimating means and other parameters of outcome distributions when data are missing. Nonresponse weights are distinct from *design weights*, which are used to compensate for planned variation in sampling rates across strata of the population.

The standard construction of nonresponse weights presumes the existence of an instrumental variable v. The standard use of such weights to infer a population mean $E[g(y)]$ yields a consistent estimate if assumption MMAR holds but not otherwise. Hence, empirical researchers contemplating application of nonresponse weights need to take care.

Weighted Sample Averages

Suppose that a random sample of size N has been drawn from population J. Let N(1) denote the sample members for whom $z = 1$, and let N_1 be the cardinality of N(1). Let $s(v): V \to [0, \infty)$ be a weighting function. Consider estimation of $E[g(y)]$ by the weighted sample average

$$\theta_N \equiv \frac{1}{N_1} \sum_{i \in N(1)} s(v_i)\cdot g(y_i).$$

By the Strong Law of Large Numbers, $\lim_{N \to \infty} \theta_N =_{a.s.} E[s(v)\cdot g(y)\,|\,z=1]$.

The standard weights provided by survey organizations have the form

$$s(v) = \frac{P(v=v)}{P(v=v\,|\,z=1)}, \quad v \in V.$$

With such weights,

$$\begin{aligned} E[s(v)\cdot g(y)\,|\,z=1] &= \textstyle\sum_{v \in V} E[s(v)\cdot g(y)\,|\,v=v, z=1]\cdot P(v=v\,|\,z=1) \\ &= \textstyle\sum_{v \in V} E[g(y)\,|\,v=v, z=1]\cdot P(v=v) \\ &= \textstyle\sum_{v \in V} E[g(y)\,|\,v=v, z=1]\cdot P(v=v\,|\,z=1)\cdot P(z=1) \\ &\quad + \textstyle\sum_{v \in V} E[g(y)\,|\,v=v, z=1]\cdot P(v=v\,|\,z=0)\cdot P(z=0) \end{aligned}$$

$$= E[g(y) \mid z = 1] \cdot P(z = 1)$$

$$+ \sum_{v \in V} E[g(y) \mid v = v, z = 1] \cdot P(v = v \mid z = 0) \cdot P(z = 0).$$

The right side of this equation equals $E[g(y)]$ if assumption MMAR holds, but it generically differs from $E[g(y)]$ otherwise.

Endnotes

Sources and Historical Notes

My work on the identifying power of assumptions using instrumental variables began with Proposition 2.4, which was introduced in Manski (1990) and developed more fully in Manski (1994, Proposition 6). The monotonicity ideas in Propositions 2.5 and 2.6 are based on Manski and Pepper (2000, Proposition 1).

The term *instrumental variable* is due to Reiersol (1945) who, along with other econometricians of his time, studied the identification of linear structural equation systems. Goldberger (1972), in a review of this literature, dates the use of instrumental variables to identify linear structural equations back to Wright (1928). Modern econometric research uses instrumental variables to address this and many other identification problems. However, the practice invariably is to assert assumptions strong enough to yield point identification of quantities of interest.

It is revealing to consider some history within economics. Until the early 1970s, empirical researchers confronting missing outcome data essentially always used assumption (2.1), although often without explicit discussion. At that time, the credibility of this assumption was questioned sharply when researchers observed that, in many economic settings, the process by which observations on y become missing is related to the value of y; see, for example, Gronau (1974). Econometricians subsequently developed a variety of models of missing data that do not assert (2.1) but instead use instrumental variables and parametric restrictions on the shape of the distribution $P(y, z)$ to point-identify $P(y)$; see, for example, Heckman (1976) and Maddala (1983). These developments were initially greeted with widespread enthusiasm, but methodological studies soon showed that seemingly minor changes in the assumptions imposed could generate large changes in the implied value of $P(y)$; see, for example, Arabmazar and Schmidt (1982), Goldberger (1983), and Hurd (1979).

Text Notes

1. Proposition 2.1 has long been well-known, so much so that it is unclear when the idea originated. In the survey sampling literature, this proposition provides the basis for construction of sampling weights that aim to enable population inference in the presence of missing data; see Complement 2A. Rubin (1976) introduced the term *missing at random*. In applied econometric research, assumption (2.1) is sometimes called *selection on observables*; see Fitzgerald, Gottschalk, and Moffitt (1998, Section IIIA) for discussion of the history of the concept and term.

2. Empirical researchers sometimes assert that assumption MAR becomes more credible as the instrumental variable partitions the population into more refined sub-populations. That is, if v_1 and v_2 are alternative specifications of the instrumental variable, with $P(v_1 | v_2)$ degenerate, researchers may assert that v_2 is a more credible instrumental variable than is v_1. Unfortunately, this assertion typically is backed up by nothing more than the empty statement that v_2 "controls for" more determinants of missing data than v_1. In principle, assumption MAR could hold for both, either, or neither of v_1 and v_2.