workers' cooperation, mutual training, and tenure longevity is another old idea in economics. A recent neoclassical application, with abundant citations, is that of Oliver E. Williamson (1975), who argues that such institutional devices as implicit contracts, collective bargaining, internal promotion ladders, and seniority rights are economically efficient when jobs and workers are heterogeneous and idiosyncratic.

A fixed structure of wages for jobs, which is emphasized by segmentation economists, is descriptively accurate and useful for analysing short-run behaviour, but even in the short run a human capital model of supply-side productivity traits can explain the match of workers to a hierarchy of wage-fixed jobs. In the long run the human capital model can explain changes in workers' productivity traits, and neoclassical models generally would predict changes in the structure of both jobs and wages.

A discussion of empirical work and policy issues concerning segmented labour markets is beyond the scope of this entry (see the bibliography below). It should be stated, however, that the sometime claim that the neoclassical economists ignore the demand side of the market in policy discussions is unfounded.

That labour market outcomes and processes are complex and controversial is evident in the intellectual legacy of the above-listed five sources of inequality. The criticisms and empirical work of the segmented labour market economists have added to this legacy, but they, like the earlier dissenters, the Marxists and the Institutionalists, remain on the bank of the mainstream.

GLEN G. CAIN

BIBLIOGRAPHY

The literature on segmented labour markets is extensive and diversified, and there are disputes about who are the leading theorists and which are the landmark articles. These characteristics make it difficult to provide a brief bibliography. In addition to the items cited in the text, several survey articles and books contain lengthy bibliographies: Taubman and Wachter (1986); Gordon, Edwards and Reich (1982); Wilkinson (1981); Cain (1976). The application of segmented labour market theories to development economics is not, however, covered in these sources, and the author is unaware of any survey or bibliographic sources for this application.

Becker, G.S. 1964. *Human Capital*. New York: Columbia University Press for the National Bureau of Economic Research.

Berger, S. and Piore, M.J. 1980. *Dualism and Discontinuity in Industrial Societies*. Cambridge: Cambridge University Press.

Braverman, H. 1974. *Labor and Monopoly Capital*. New York: Monthly Review Press.

Cain, G. 1976. The challenge of segmented labor market theories to orthodox theory: a survey. *Journal of Economic Literature* 14(4), December, 1215–57.

Cairnes, J.E. 1874. *Some Leading Principles of Political Economy*. New York: Harper & Brothers.

Gordon, D.M., Edwards, R.C. and Reich, M.S. 1982. *Segmented Work, Divided Workers: The Historical Transformation of Labor in the United States*. Cambridge: Cambridge University Press.

Marshall, A. [1890] 1959. *Principles of Economics*. 8th edn, London: Macmillan.

Mill, J.S. [1848] 1900. *The Principles of Political Economy*, Vol. 1. Revised edn, The World's Greatest Classics, New York: Colonial Press.

Myrdal, G. 1944. *An American Dilemma*. New York: Harper & Row.

Taubman, P. and Wachter, M.L. 1986. Segmented labor markets. In *Handbook of Labor Economics*, ed. O. Ashenfelter and R. Layard, Amsterdam: Elsevier Science Publishers.

Werthheim, W.F. 1967. Economy, dual. In *International Encyclopedia of the Social Sciences*, Vol. 4, New York: Macmillan and Free Press, 495–500.

Wilkinson, F. (ed.) 1981. *The Dynamics of Labor Market Segmentation*. New York: Academic Press.

Williamson, O.E. 1975. *Markets and Hierarchies: Analysis and Antitrust Implications*. New York: Free Press.

seigniorage. Full-bodied monies such as gold coin contain metal approximately equal in value to the face value of the coin. Under the gold standard, metal could be brought to the mint and freely coined into gold, less a small *seigniorage* charge for the privilege. Subsidiary or token coin and paper money by contrast cost much less to produce than their face value. The excess of the face value over the cost of production of currency is also called *seigniorage*, because it accrued to the *seigneur* or ruler who issued the currency, in early times.

The use of paper money instead of full-bodied coin by modern governments generates a very large social saving in the use of the resources that would otherwise have to be expended in mining and smelting large quantities of metal. The value of this seigniorage can be measured by considering the aggregate demand curve for currency, as a function of the rate of interest. The area under this demand curve represents the aggregate flow of social benefits from holding currency, under certain assumptions. The social cost of holding currency is measured by the opportunity cost of the resources it takes to produce the currency. If gold were used for currency, its opportunity cost would be measured by the rate of interest that could be earned on those resources if transferred to some other use. Thus the area under the demand curve between the market rate of interest and the cost of providing paper currency represents the flow of seigniorage or social saving that accrues from the use of paper currency instead of gold.

In the international monetary system, gold remains a very large fraction of total holdings of international reserves (about 45 per cent of total reserves valued at market prices at the end of March 1985). Substitution of fiduciary reserve assets such as Special Drawing Rights created by the International Monetary Fund or United States dollars for gold would generate a substantial social gain in the form of seigniorage equal to the excess of the opportunity cost of capital over the costs of providing the fiduciary asset. If interest is paid to the holders of the reserve asset, the seigniorage is split between the issuer and the holder.

The existence of these large seigniorage gains is what led to the development of the gold exchange standard, under which first British sterling, before World War II, and since then United States dollars and other currencies have substituted for gold in international reserve holdings. As interest rates paid on these reserve assets have risen, more of the seigniorage has accrued to holders of reserve assets.

Further substitution of fiduciary reserve assets for gold in the international monetary system has frequently been suggested, and the Second Amendment to the Charter of the International Monetary Fund adopted in 1978 proposed such a goal. Little progress has been made, however, since the underlying issue is one of trust in the financial probity of the issuer and its continued political stability, as well as its continued willingness to convert reserve assets into usable currencies over long periods of time.

S. BLACK

selection. *See* COMPETITION AND SELECTION.

selection bias and self-selection. The problem of selection bias in economic and social statistics arises when a rule other than simple random sampling is used to sample the underlying population that is the object of interest. The distorted representation of a true population as a consequence of a sampling rule is the essence of the selection problem. Distorting selection rules may be the outcome of decisions of sample survey statisticians, self-selection decisions by the agents being studied or both.

A random sample of a population produces a description of the population distribution of characteristics that has many desirable properties. One attractive feature of a random sample generated by the *known rule* that all individuals are equally likely to be sampled is that it produces a description of the population distribution of characteristics that becomes increasingly accurate as sample size expands.

A sample selected by any rule not equivalent to random sampling produces a description of the population distribution of characteristics that does not accurately describe the true population distribution of characteristics no matter how big the sample size. Unless the rule by which the sample is selected is known or can be recovered from the data, the selected sample cannot be used to produce an accurate description of the underlying population. For certain sampling rules, even knowledge of the rule generating the sample does not suffice to recover the population distribution from the sampled distribution.

This entry defines the problem of selection bias and presents conditions required to solve the problem. Examples of various types of commonly encountered sampling frames are given and specific economic selection mechanisms are presented. Assumptions required to use selected samples to determine features of the population distribution are discussed.

The analytical framework developed to understand the inferential problems raised by selection bias is also fruitful in understanding the economics of self-selection. The prototypical choice theoretic model of self-selection is that of Roy (1951). In his model, agents choose among a variety of discrete 'occupational' opportunities. Agents can pursue only one 'occupation' at a time. While every person can, in principle, do the work in each 'occupation', at least at some level of competence, self-interest drives individuals to choose that 'occupation' which produces the highest income (utility) for them. As in the statistical selection bias problem, there is a latent population (of skills). Observed (utilized) skill distributions are the outcome of a selection rule by agents. The relationship between observed and latent skill distributions is of considerable interest and underlies recent work on worker hierarchies (see Willis and Rosen, 1979). The 'occupations' can be: (a) market work or non-market work (b) unemployed and searching or working at the offered wage (c) working in one province or working in another, or (d) any choice among a set of mutually exclusive opportunities.

Because the insights in the Roy model underly much recent research, we present a brief exposition of it and demonstrate how it can be or has been fruitfully extended to a variety of settings. An important issue, closely linked to the problem of identifying population parameters from selected sample distributions, is the empirical content of economic models of self-selection and worker hierarchies. Are they artefacts of distributional assumptions for unobservable skills or are they genuine behavioural hypotheses?

1. A DEFINITION AND SOME EXAMPLES OF SELECTION BIAS

Any selection bias model can be described by the following set-up. Let **Y** be a vector of outcomes of interest and let **X** be a vector of 'control' or 'explanatory' variables. The population distribution of (**Y**, **X**) is $F(\mathbf{y}, \mathbf{x})$. To simplify the exposition we assume that the density is well defined and write it as $f(\mathbf{y}, \mathbf{x})$.

Any sampling rule can be interpreted as producing a non-negative weighting function $\omega(\mathbf{y}, \mathbf{x})$ that alters the population density. Let $(\mathbf{Y}^*, \mathbf{X}^*)$ denote the sampled random variables. The density of the sampled data $g(\mathbf{y}^*, \mathbf{x}^*)$ may be written as

$$g(\mathbf{y}^*, \mathbf{x}^*) = \omega(\mathbf{y}^*, \mathbf{x}^*) f(\mathbf{y}^*, \mathbf{x}^*) \Big/ \int \omega(\mathbf{y}^*, \mathbf{x}^*) f(\mathbf{y}^*, \mathbf{x}^*)\, \mathrm{d}\mathbf{y}^*\, \mathrm{d}\mathbf{x}^* \qquad (1.1)$$

where the denominator of the expression is introduced to make the density $g(\mathbf{y}^*, \mathbf{x}^*)$ integrate to one as is required for proper densities.

Alternatively, the weight may be defined as

$$\omega^*(\mathbf{y}^*, \mathbf{x}^*) = \frac{\omega(\mathbf{y}^*, \mathbf{x}^*)}{\int \omega(\mathbf{y}^*, \mathbf{x}^*) f(\mathbf{y}^*, \mathbf{x}^*)\, \mathrm{d}\mathbf{y}^*\, \mathrm{d}\mathbf{x}^*}$$

so that

$$g(\mathbf{y}^*, \mathbf{x}^*) = \omega^*(\mathbf{y}^*, \mathbf{x}^*) f(\mathbf{y}^*, \mathbf{x}^*). \qquad (1.2)$$

Sampling schemes for which $\omega(\mathbf{y}, \mathbf{x}) = 0$ for some values of (**Y**, **X**) create special problems. For such schemes, not all values of (**Y**, **X**) are sampled. Let indicator variable $i(\mathbf{x}, \mathbf{y}) = 0$ if a potential observation at values **y**, **x** cannot be sampled and let $i(\mathbf{y}, \mathbf{x}) = 1$ otherwise. Let $\Delta = 1$ record the occurrence of the event 'a potential observation is sampled, i.e. the value of **y**, **x** is observed' and let $\Delta = 0$ if it is not. In the population, the proportion that is sampled is

$$\Pr(\Delta = 1) = \int i(\mathbf{y}, \mathbf{x}) f(\mathbf{y}, \mathbf{x})\, \mathrm{d}\mathbf{y}\, \mathrm{d}\mathbf{x}, \qquad (1.3)$$

while

$$\Pr(\Delta = 0) = 1 - \Pr(\Delta = 1).$$

For samples in which $\omega(\mathbf{y}, \mathbf{x}) = 0$ for a non-negligible proportion of the population ($\Pr(\Delta = 0) > 0$), it is clarifying to consider two cases. A *truncated sample* is one for which $\Pr(\Delta = 1)$ is not known and cannot be consistently estimated. For such a sample, (1.1) is the density of all of the sampled **Y** and **X** values. A *censored sample* is one for which $\Pr(\Delta = 1)$ is known or can be consistently estimated. The sampling rule in this case is such that values of **y**, **x** for which $\omega(\mathbf{y}, \mathbf{x}) = 0$ are not known but it is known whether or not $i(\mathbf{y}, \mathbf{x}) = 0$ for all values of **Y**, **X**. In this case it is notationally convenient to define $(\mathbf{Y}^*, \mathbf{X}^*) = (\mathbf{0}, \mathbf{0})$ for values of **y**, **x** such that $\omega(\mathbf{y}, \mathbf{x}) = i(\mathbf{y}, \mathbf{x}) = 0$. Such a definition is innocuous provided that in the population there is no point mass (concentration of probability mass) at (**0**, **0**). (Any value other than (**0**, **0**) can be selected provided that there is no point mass at that value). Given $\Delta = 0$, the distribution of $\mathbf{Y}^*, \mathbf{X}^*$ is

$$G(\mathbf{y}^*, \mathbf{x}^*) = 1 \quad \text{for} \quad \Delta = 0$$

at

$$\mathbf{Y}^* = \mathbf{0} \quad \text{and} \quad \mathbf{X}^* = \mathbf{0}.$$

The joint density of $\mathbf{Y}^*, \mathbf{X}^*, \Delta$ for the case of a censored sample is obtained by combining (1.1) and (1.3). Thus

$$g(\mathbf{y}^*, \mathbf{x}^*, \delta) = \left[\frac{\omega(\mathbf{y}^*, \mathbf{x}^*) f(\mathbf{y}^*, \mathbf{x}^*)}{\int \omega(\mathbf{y}^*, \mathbf{x}^*) f(\mathbf{y}^*, \mathbf{x}^*)\, \mathrm{d}\mathbf{y}^*\, \mathrm{d}\mathbf{x}^*}\right]^{\delta} \times \left[\int i(\mathbf{y}, \mathbf{x}) f(\mathbf{y}, \mathbf{x})\, \mathrm{d}\mathbf{y}\, \mathrm{d}\mathbf{x}\right]^{\delta} \times [1]^{1-\delta} \left[\int (1 - i(\mathbf{y}, \mathbf{x})) f(\mathbf{y}, \mathbf{x})\, \mathrm{d}\mathbf{y}\, \mathrm{d}\mathbf{x}\right]^{1-\delta}. \qquad (1.4)$$

The first term on the right-hand side of (1.4) is the conditional density of $\mathbf{Y}^*, \mathbf{X}^*$ given $\Delta = 1$. The second term is the probability that $\Delta = 1$. The third term is the conditional density of $\mathbf{Y}^*, \mathbf{X}^*$ given $\Delta = 0$. This density assigns unit mass to $\mathbf{y}^* = 0, \mathbf{x}^* = \mathbf{0}$ when $\Delta = 0$. The fourth term is the probability that $\Delta = 0$. Notice that in the case in which $\omega(\mathbf{y}, \mathbf{x}) > 0$ for all $\mathbf{y}, \mathbf{x}, \Delta = 1$ and (1.4) is identical to (1.1).

In a random sample $\omega(\mathbf{y}^*, \mathbf{x}^*) = 1$ (and so $\omega^*(\mathbf{y}^*, \mathbf{x}^*) = 1$). In a selected sample, the sampling rule weights the data differently. Values of $(\mathbf{Y}, \mathbf{X})$ are over-sampled or under-sampled relative to their occurrence in the population. In the case of truncated samples, the weight is zero for certain values of the outcome.

In many problems in economics, attention focuses on $f(\mathbf{y}|\mathbf{x})$, the conditional density of $\mathbf{Y}$ given $\mathbf{X} = \mathbf{x}$. In such problems knowledge of the population distribution of $\mathbf{X}$ is of no direct interest. If samples are selected solely on the $\mathbf{x}$ variables ('selection on the exogenous variables'), $\omega(\mathbf{y}, \mathbf{x}) = \omega(\mathbf{x})$ and there is no problem about using selected samples to make valid inference about the population conditional density. This is so because in the case of selection on the exogenous variables

$$g(\mathbf{y}^*, \mathbf{x}^*) = f(\mathbf{y}^*|\mathbf{x}^*)\frac{\omega(\mathbf{x}^*)f(\mathbf{x}^*)}{\int \omega(\mathbf{x}^*)f(\mathbf{x}^*)\,\mathrm{d}\mathbf{x}}$$

and

$$g(\mathbf{x}^*) = \frac{\omega(\mathbf{x}^*)f(\mathbf{x}^*)}{\int \omega(\mathbf{x}^*)f(\mathbf{x}^*)\,\mathrm{d}\mathbf{x}^*}$$

Thus

$$g(\mathbf{y}^*|\mathbf{x}^*) = \frac{g(\mathbf{y}^*, \mathbf{x}^*)}{g(\mathbf{x}^*)} = f(\mathbf{y}^*|\mathbf{x}^*).$$

For such problems, sample selection distorts inference only if selection occurs on $\mathbf{y}$ (or $\mathbf{y}$ and $\mathbf{x}$). Sampling on both $\mathbf{y}$ and $\mathbf{x}$ is termed *general stratified sampling*.

From a sample of data, it is not possible to recover the true density $f(\mathbf{y}, \mathbf{x})$ without knowledge of the weighting rule. On the other hand, if the weighting rule is known $(\omega(\mathbf{y}^*, \mathbf{x}^*))$, the density of the sampled data is known $(g(\mathbf{y}^*, \mathbf{x}^*))$, the support of $(\mathbf{y}, \mathbf{x})$ is known and $\omega(\mathbf{y}, \mathbf{x})$ is nonzero, then $f(\mathbf{y}, \mathbf{x})$ can always be recovered because

$$\frac{g(\mathbf{y}^*, \mathbf{x}^*)}{\omega(\mathbf{y}^*, \mathbf{x}^*)} = \frac{f(\mathbf{y}^*, \mathbf{x}^*)}{\int \omega(\mathbf{y}^*, \mathbf{x}^*)f(\mathbf{y}^*, \mathbf{x}^*)\,\mathrm{d}\mathbf{y}^*\,\mathrm{d}\mathbf{x}^*} \tag{1.5}$$

and by hypothesis both the numerator and denominator of the left-hand side are known. From the requirement that $(\mathbf{y}^*, \mathbf{x}^*)$ has a well defined density

$$\int f(\mathbf{y}^*, \mathbf{x}^*)\,\mathrm{d}\mathbf{y}^*\,\mathrm{d}\mathbf{x}^* = 1.$$

Integrating the left-hand side of (1.5) it is possible to determine $\int \omega(\mathbf{y}^*, \mathbf{x}^*)f(\mathbf{y}^*, \mathbf{x}^*)\,\mathrm{d}\mathbf{y}^*\,\mathrm{d}\mathbf{x}^*$ and hence to use (1.5) to recover the population density of the data.

The requirements that (a) the support of $(\mathbf{y}, \mathbf{x})$ is known and (b) $\omega(\mathbf{y}, \mathbf{x})$ is nonzero are not innocuous. In many important problems in economics requirement (b) is not satisfied: the sampling rule excludes observations for certain values of $\mathbf{y}, \mathbf{x}$ and hence it is impossible without invoking further assumptions to determine the population distribution of $(\mathbf{Y}, \mathbf{X})$ at those values. If neither the support nor the weight is known, it is impossible, without invoking strong assumptions, to determine whether the fact that data are missing at certain $\mathbf{y}, \mathbf{x}$ values is due to the sampling plan or that the population density has no support at those values. We now turn to some specific sampling plans of interest in economics.

Example 1. Data are collected on incomes of individuals whose income Y exceeds a certain value c (for cutoff value). The rule is to observe Y if $Y > c$. Thus $\omega(y) = 1$ if $y > c$ and $\omega(y) = 0$ if $y \leqslant c$. Because the weight is zero for some values of y, we know that knowledge of the sampling rule does *not* suffice to recover the population distribution. From a random sample of the entire population, the social scientist knows or can consistently estimate (a) the sample distribution of Y above c and (b) the proportion of the original random sample with income below c ($F(c)$ where F is the distribution function of Y). The social scientist does not observe values of Y below c.

In this example, observed income is a *truncated random variable*. The point of truncation is c. The *sample* of observed income is said to be *censored*. If the proportion of the original random sample with income below c is not known and cannot be consistently estimated, the *sample* is *truncated*. In a truncated sample, nothing is known about the proportion of the underlying population that can appear in the sample. A sample is truncated only if $\omega(\mathbf{y}) = 0$ for some intervals of $\mathbf{y}$ (for $\mathbf{y}$ continuous) or if $\omega(\mathbf{y}) = 0$ at values of $\mathbf{y}$ at which there is finite probability mass. In a censored sample, the proportion of the underlying population that can appear in the sample is known, at least to an arbitrarily high degree of approximation, as sample size increases.

Let $Y^* = Y$ if $Y > c$. Define $Y^* = 0$ otherwise (the choice of the value for Y^* when Y is not observed is inessential and any value can be used in place of 0 provided that the true distribution places no mass at the selected value). Define an indicator variable $\Delta = 1$ if $Y > c$. $\Delta = 0$ otherwise. Then the distribution of Y^* is

$$\begin{aligned} G(y^*|Y > 0) &= F(y^*|Y \quad c) = F(y^*|\delta = 1) \\ &= \frac{F(y^*)}{1 - F(c)},\ y^* > c. \end{aligned} \tag{1.6a}$$

$$G(y^*|Y^* > 0) = 1 \qquad \text{for} \qquad Y^* = 0\ (\Delta = 0). \tag{1.6b}$$

Observe that (1.6a) is obtained from (1.1) by setting $\omega(y^*) = 1$ if $y > c$, and $\omega(y^*) = 0$ otherwise, and integrating up with respect to y^*. The distribution of Δ is

$$\mathrm{pr}(\Delta) = [1 - F(c)]^{\delta}[F(c)]^{1-\delta}.$$

The joint distribution of (Y^*, Δ) is

$$\begin{aligned} F(y^*, \delta) &= F(y^*|\delta)\Pr(\delta) \\ &= \left\{\frac{F(y^*)}{(1 - F(c))}\right\}^{\delta}[1 - F(c)]^{\delta}(1)^{1-\delta}[F(c)]^{1-\delta} \\ &= [F(y^*)]^{\delta}[F(c)]^{1-\delta}. \end{aligned} \tag{1.7}$$

Note that (1.7) is obtained from (1.4) by setting $\omega(y) = 0, y < c, \omega(y) = 1$ otherwise, by setting $i(y) = \omega(y)$, and by integrating up with respect to y^*. For normally distributed Y, (1.7) is the 'Tobit' distribution.

The difference between the information in a truncated sample and the information in a censored sample is encapsulated in the contrast between (1.6a) and (1.7). Clearly there is more information in a censored sample than in a truncated sample because one can obtain (1.6a) from (1.7) (by conditioning on $\Delta = 1$) but not vice versa.

Inferences about the population distribution based on assuming that $F(y^*|Y > c)$ closely approximates $F(y)$ are potentially very misleading. A description of population income inequality based on a subsample of high income people may convey no information about the true population distribution.