Identifying the Hand of Past: Distinguishing State Dependence from Heterogeneity

By JAMES J. HECKMAN*

A basic problem in social science is to ascertain the importance of initial endowments on subsequent outcomes of a dynamic process. Interest in this topic centers on two distinct issues: (a) Do initial endowments have a temporally persistent effect on outcomes (i.e., is there "heterogeneity")?; (b) Are the effects of initial endowments attenuated or accentuated by subsequent experiences of the process being studied or by related processes (i.e., is there "state dependence")?

These two questions show up in a variety of contexts. 1) The importance of family background on a person's subsequent education and earnings is a hotly debated topic. Do market or nonmarket mechanisms reinforce or diminish initial endowments? 2) The incidence of criminal activity is concentrated among a small population of repeated offenders. Are certain persons "prone" to criminality, or does crime breed crime? 3) Are persons who experience unemployment more likely to experience future unemployment due to loss of work experience or market stigma? 4) Does early entry into an industry confer an advantage of incumbency, or does it merely proxy the basic managerial and innovative ability of early entrants?

Many formal models of state-dependent processes or processes with persistent effects of initial conditions have been formulated. Invariance of steady states to initial endowments was viewed as a desirable feature of an economic model in the 1960's and 1970's. Path-dependent synergism and nonergodicity are fashionable features of models today. Is the choice between models with long-run dependence on initial conditions and those without solely a matter of intellectual esthetics? Can data discriminate between these two classes of models? How many economically extraneous statistical "regularity conditions" have to be imposed in order to distinguish between these models? How many and what kind of *maintained* assumptions are required in order to let the data speak on these important questions?

In this paper I focus on one nonergodic model that receives much attention in the literature—the model of "duration dependence" and "heterogeneity" discussed by Herbert Silcock (1954) and also examined by Tony Lancaster (1979), Chris Elbers and Geert Ridder (1982), myself and Burton Singer (1984), and myself and Bo Honoré (1989).

I. The Problem of Separating Heterogeneity from Duration Dependence

The length of time in a state, L, and its determinants are the objects of interest. For example, the length of time that a new entrant firm survives in an industry, the length of an unemployment spell or a job spell or a strike, the length of time a person spends in school, or the time to a first birth are all objects of interest in current research. For simplicity, assume a continuous time model.

A set of explanatory variables is thought to explain L, but in addition there is some "randomness" or unpredictability given the explanatory variables. Outcomes depend on a larger set of variables than the economic analyst has at his or her disposal. We write (X,Θ) where X is a vector of observed explanatory variables and Θ is a scalar unobserved variable known to the agent but not the econometrician. The assumption that Θ is scalar is made for analytical con-

*University of Chicago, Chicago, IL 60637. This work was supported by NSF-87-10145 and NIH grant HD-19226. I have benefited from discussions with Ricardo Barros and Bo Honoré.

venience. The assumption of "randomness" is that the distribution of L is nondegenerate given (X,Θ), that is,

$$(1) \qquad \Pr(L \leq l|x,\theta) = F(l|x,\theta).$$

If (X,Θ) fully explain durations, the probability on the left would be zero or one. The distinction between Θ and the unobservable giving rise to the nondegenerate conditional distribution of L is arbitrary. Distribution (1) is of economic interest if agents make decisions based on Θ but not on the random components that make the distribution nondegenerate given (X,Θ). Then (1) can properly be called a structural distribution and is an object of economic interest.

In duration analysis it is convenient to work with the survivor function: $S(l|x,\theta) = \Pr(L > l|x,\theta)$, the conditional probability that a spell exceeds length l. To avoid technicalities, assume that L is continuously distributed so the density of L is well defined. The structural conditional hazard rate or exit rate from the spell is denoted $h(l|x,\theta)$. If h is increasing in l, positive duration dependence is said to be present for those values of l. If h is decreasing, negative duration dependence is said to be present. In either case, state dependence (or one form of the hand of the past) is present because the length of time spent in a state determines the conditional exit rate from the state.

For time-invariant (X,Θ) or for time-varying (X,Θ) satisfying the conditions given in my article with James Walker (1990a),

$$(2) \qquad S(l|x,\theta) = \exp - \int_0^l h(u|x(u),\theta(u))\,du.$$

The survivor is the inverse of the exponentiated integrated hazard.[1] For analytical convenience I assume that $\Theta(u) = \Theta$ is constant for all u and is a scalar random variable. Its distribution function is $G(\theta)$. The variable Θ is assumed statistically independent of X, and can thus be interpreted as a time-invariant endowment or propensity.

The problem of distinguishing state dependence from heterogeneity (or initial endowments) can now be stated precisely. The data determine the conditional distribution of L given X which is the *mean* conditional survivor function:

$$(3) \qquad S(l|x) = \int_{\underline{\theta}} S(l|x,\theta)\,dG(\theta)$$
$$= \int_{\underline{\theta}} \exp\left(-\int_0^l h(u|x(u),\theta)\,du\right) dG(\theta)$$

where $\underline{\theta}$ is the support of Θ. Associated with $S(l|x)$ is the empirical hazard

$$(4) \qquad h(l|x).$$

A standard result in the literature establishes that (4) is biased toward negative duration dependence relative to $h(l|x,\theta)$.[2] The more mobility-prone agents exit spells more rapidly producing a distribution of Θ conditional on L and X that is weighted toward less mobile values.

The issue is whether or not it is possible to identify $G(\theta)$ and $h(u|x(u),\theta)$ from $S(l|x)$. Without further structure imposed, it is not. There is no way to separate out the contributions of state dependence from heterogeneity in a completely general model. The data do not "speak for themselves" except in the case where $h(l|x)$ shows positive duration dependence. For this to arise, $h(l|x,\theta)$ must show positive duration dependence at least for certain intervals of the support of Θ. (See myself and Singer, 1985.)

II. Separability as an Avenue of Identifiability

One way to answer the stated question is to impose specific functional forms. My article with Singer (1984) presents identification conditions for a variety of conventional models in which $h(\cdot|\cdot)$ is specified to be-

[1] For general time-varying explanatory variables, relation (2) is not true although it is commonly assumed to be true in applied work.

[2] See Proposition 1 in my book with Singer (1985, p. 53).

long to a parametric family and $G(\theta)$ is assumed to be a proper probability distribution but its functional form is not specified. Identification through explicit functional form assumptions about both h and G is conventional but controversial. See the examples of the consequences of misspecification discussed in my book with Singer (1985).

An alternative and somewhat more general route to *nonparametric* identification is to characterize the hazard in some more abstract way. One starting point is to invoke a separability restriction of the sort often made in consumer and producer theory,

$$h(l|x) = m(l|x)\theta. \tag{5}$$

Writing $M(l|x) = \int_0^l h(u|x(u))\,du$, we conclude that

$$S(l|x) = \int_{\underline{\theta}} [\exp - M(l|x)\theta]\, dG(\theta). \tag{6}$$

In terms of a transformed time scale, $M(L|x)$ is an exponential random variable conditional on θ. The transformed dependent variable has the conditional expectation $E(\ln(M(L|x)|\theta)) = \Gamma'(1) - \ln\theta$, where Γ is the gamma function. Written as a familiar-looking regression equation, the problem of identifying the hand of the past is to determine M and the distribution of Θ from the regression

$$\ln M(L|x) = \Gamma'(1) - \ln\theta + V \tag{7}$$

where $E(V) = 0$, $\text{Var}(V) = \pi^2/6$.[3] The severity of the identification problem is evident from (7).

Representation (6) reveals the nature of the available sample information. Level sets of $S(l|x)$, that is, $\{(l,x)\colon S(l|x) = s\}$ trace out level sets of $M(l|x)$. If $M(l|x)$ is differentiable in x, we know the ratio of partial derivatives of M when the right-hand sides of the following expressions are well defined:

$$\frac{\dfrac{\partial M(l|x)}{\partial x_i}}{\dfrac{\partial M(l|x)}{\partial x_j}} = \frac{\dfrac{\partial S(l|x)}{\partial x_i}}{\dfrac{\partial S(l|x)}{\partial x_j}}, \tag{8}$$

$$\frac{\dfrac{\partial M(l|x)}{\partial x_i}}{\dfrac{\partial M(l|x)}{\partial l}} = \frac{\dfrac{\partial S(l|x)}{\partial x_i}}{\dfrac{\partial S(l|x)}{\partial l}}$$

where x_i, x_j are elements of x. Assuming $M(0,x) = 0$ for all x (so there are no jumps in the hazard at zero) and $E(\Theta) < \infty$, the right derivative of $S(l|x)$ evaluated in the neighborhood of $l = 0$ is

$$\left.\frac{\partial S(l|x)}{\partial l}\right|_{l=0} = -\left.\frac{\partial M(l|x)}{\partial l}\right|_{l=0} E(\Theta).$$

If $E(\Theta) = 1$ or some other known constant, the magnitude of the local hazard is known for all x.

This information does not suffice to identify M or G. A certain class of monotonic transformations of M explains the data equally well. Denote a member of this class by $J(0) = 0$.

$$\begin{aligned} S(l|x) &= \int_{\underline{\theta}} [\exp - [M(l|x)]\theta]\, dG(\theta) \\ &= \int_{\underline{\theta}^*} [\exp - [J(M(l|x))\theta^*]\, dG^*(\theta^*) \end{aligned} \tag{9}$$

where Θ^* is a random variable in general distinct from Θ with distribution G^*.

Observe that $S(l|x)$ is both an exponential mixture of M and an exponential mixture of $J(M)$. As noted by William Feller (1971, p. 452), every mixture of exponentials is an infinitely divisible distribution so by Theorem 1 of Feller (p. 450): $S(M) = e^{-\psi(M)}$ and $S(J) = e^{-\psi^*(J(M))}$ where ψ and ψ^* satisfy the condition that $\psi(0) = \psi^*(0) = 0$ and both have a completely

[3]See the derivation in Chris Flinn and myself (1982, Appendix B).

monotone first derivative.[4] Since both $S(M(l|x))$ and $S(J(M(l|x)))$ equal $S(l|x)$ for the same values of l and x, the equivalence class is defined as containing those J such that

$$(10) \qquad \psi(M) = \psi^*(J(M)).$$

The requirement that ψ' and $(\psi^*)'$ are completely monotone restricts the admissible J. If $\psi'/(\psi^*)'$, $[(\psi^*)' \neq 0]$ is completely monotonic, then so is J'. Repeated differentiation of (10) produces an algorithmic definition of the derivatives of J. My article with Walker (1990b) uses these results for a class of exponential models $(M(l|x) = l)$ and demonstrates the equivalence of a broad class of apparently different duration models within the class of mixture of exponentials models and shows the practical importance of this equivalence for a problem in demography.

Conventional duration models invoke additional separability assumptions:

$$(11) \qquad M(l|x) = Z(l)K(x)$$

to produce the "proportional hazards" model. Z is assumed to be nonnegative, differentiable, and monotone increasing in l. $K(x)$ is nonnegative. The previous assumptions ensure that $Z(0) = 0$.

Identification in this class of models has received much attention for the case when X is time invariant. Elbers and Ridder establish that if $E(\Theta) < \infty$ and X assumes at least two distinct values, Z, K, and G are nonparametrically identified. Singer and I (1984) replace the finite mean of Θ assumption with a tail condition on G. Ricardo Barros and Honoré (1988) show that for this model $J(M) = aM^b$, $0 < a$, $0 < b \leq 1$ so J' is completely monotonic. They prove that the components of proportional hazards models are identified up to an arbitrary normalization and up to arbitrary powers (b).[5]

[4]Thus letting $R = \psi'$, R is completely monotone if $(-1)^n R^{(n)}(M) \geq 0$ for all nonnegative integer n and for all $M > 0$ where (n) denotes the nth derivative.

[5]See also the discussion in my book with Singer (1985, p. 64).

(See also Ridder, 1990.) By restricting the mean of Θ or the tail of G, one can avoid this arbitrariness. But such restrictions are themselves arbitrary. Tails of G are not known in advance. Finiteness of the mean for unobservable Θ is not necessarily a reasonable assumption. Models with infinite means for Θ produce entirely reasonable duration models.

Honoré (1990) establishes that if the elements of X are time varying, are not functionally dependent on l, and if one coordinate of X has discrete jumps for a proportion of the population strictly between zero and one and X is time invariant for the rest of the population, then Z, K, and G are nonparametrically identified without invoking arbitrary tail conditions for G or finite mean conditions.

It is interesting to contrast these results with those achieved from an *accelerated hazard* model:

$$(12) \qquad M(l|x) = Z(lK(x)).$$

For this model, the assumption $E(\Theta) < \infty$ or the other assumptions do not identify Z or G. However, if $Z'(0) > 0$ and $E(\Theta) < \infty$, and adopting the *normalization* $K(x_*) = 1$, then $K(x)$ is nonparametrically identified up to an arbitrary positive scale.

$$\left. \frac{\dfrac{\partial S(t|x)}{\partial t}}{\dfrac{\partial S(t|x_*)}{\partial t}} \right|_{t=0} = K(x)$$

We may rewrite (9) using known $q = lK(x)$, to reach $S(q) = \int_\theta [\exp - Z(q)\theta]\, dG(\theta)$. The terms Z, G are not identified except for the special case when $Z(t) = at^b$ (the Weibull) when the proportional hazard model and accelerated hazard model coincide. In that case it is not necessary to have any regressor in order to identify a, b, and G (see myself and Singer, 1984). My article with Honoré notes that $K(x)$ can be identified under weaker assumptions than are required to identify Z and G in both the proportional hazard and accelerated hazard models. It also notes that the accelerated failure time model is not nonparametrically identified.[6]

Barros (1986) considers the case $M(l|x) = N(Z(l), K(x))$ where Z and K have the properties assumed in the discussion of the proportional hazard model without time-varying variables and where N is assumed *known* and $N(Z, \cdot) = 0$ if and only if $Z = 0$, N is differentiable with respect to K, $N_K \neq 0$ for $Z \neq 0$ and $N_Z(0, \cdot)$ is finite and different from zero with a well-defined inverse. For $E(\Theta) < \infty$, he proves that Z, K, and G are nonparametrically identified.

III. Conclusion

The ability to distinguish between heterogeneity and duration dependence in single-spell duration models rests critically on maintaining explicit assumptions about the way unobservables and observables interact. A general nonseparable model is nonparametrically underidentified. Separable models are identified subject to additional assumptions on the nature of the explanatory variables (as in Honoré) or subject to restrictions on unobservables. Economically extraneous statistical assumptions drive the answer to the stated question. Viewed as a prototype for identification in general nonergodic models, these results are not encouraging.[7]

[6]The same qualitative results hold for the more natural version of the accelerated hazard model $Z(tK(x)\theta)$.

[7]Honoré demonstrates how access to multiple spell data facilitates identification of single spell hazards. My article with Honoré discusses conditions for nonparametric identifiability of the competing risks model.

REFERENCES

Barros, Ricardo, "The Identifiability of Hazard Functions," working paper, University of Chicago, 1986.

______ **and Honoré, Bo,** "Identification of Duration Models with Unobserved Heterogeneity," working paper, Northwestern University, 1988.

Elbers, Chris and Ridder, Geert, "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model," *Review of Economic Studies*, July 1982, *49*, 403–09.

Feller, William, *An Introduction to Probability Theory and Its Applications*, Vol. II, New York: Wiley & Sons, 1971.

Flinn, Chris and Heckman, James, "Models for the Analysis of Labor Force Dynamics," in G. Rhodes and R. Bassman, eds., *Advances in Econometrics*, Vol. 1, 1982, 35–95.

Heckman, James and Honoré, Bo, "The Identifiability of the Competing Risks Model," *Biometrika*, June 1989, *76*, 325–30.

______ **and Singer, Burton,** "The Identifiability of the Proportional Hazard Model," *Review of Economics Studies*, July 1984, *51*, 231–43.

______ **and** ______, "Social Science Duration Analysis" in their *Longitudinal Analysis of Labor Market Data*, Cambridge: Cambridge University Press, 1985, ch. 2.

______ **and Walker, James,** (1990a) "The Relationship Between Wages and Income and The Timing and Spacing of Births: Evidence From Swedish Longitudinal Data," *Econometrica*, November 1990, *58*, 1411–42.

______ **and** ______, (1990b) "Estimating Fecundability from Data on Waiting Times to First Conception," *Journal of the American Statistical Association*, June 1990, *85*, 283–94.

Honoré, Bo, "Identification Results for Duration Models with Multiple Spells or Time-Varying Covariates," unpublished manuscript, Northwestern University, 1990.

Lancaster, Tony, "Econometric Methods for the Duration of Unemployment," *Econometrica*, September 1979, *47*, 939–56.

Ridder, Geert, "The Nonparametric Identification of Generalized Accelerated Failure-Time Models," *Review of Economic Studies*, May 1990, *57*, 167–81.

Silcock, Herbert, "The Phenomenon of Labor Turnover," *Journal of the Royal Statistical Society*, Series A, July 1954, *117*, 429–40.