

Job Matching and Occupational Choice

Author(s): Robert A. Miller

Source: *Journal of Political Economy*, Dec., 1984, Vol. 92, No. 6 (Dec., 1984), pp. 1086-1120

Published by: The University of Chicago Press

Stable URL: <https://www.jstor.org/stable/1831993>

---

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

The University of Chicago Press is collaborating with JSTOR to digitize, preserve and extend access to *Journal of Political Economy*

# Job Matching and Occupational Choice

---

Robert A. Miller

*Carnegie-Mellon University*

This paper presents a model of job matching that generalizes the existing literature by allowing for different job types, or occupations. Such differences affect the value of job-specific experience, inducing a career profile where certain types are sampled before others. More specifically, the analysis shows that it is optimal for the young and inexperienced to gravitate toward jobs exhibiting a certain kind of risk. Then, after deriving the equilibrium job turnover rate for an economy in which people do not switch occupations, panel data are used to estimate its underlying parameters. The hypothesis that people do not switch occupations is rejected against the alternative that they do, thus providing empirical support for the theoretical extension undertaken here.

## I. Introduction

The relationship linking skills that are job specific to turnover decisions has long been considered an important issue in the economics of labor mobility. Earlier writers, most notably Becker (1962, 1975), Mincer (1962), and Oi (1962), perceived a two-way flow between job-specific training and tenure. On the one hand, Becker asserts the “willingness of workers or firms to pay for specific training should . . . closely depend [negatively] on the likelihood of labor turnover” (1962, p. 19). On the other hand, these authors also recognized that the more job-specific capital a worker has, the more destructive a

This paper was drawn from the first chapter of my doctoral dissertation (1982). I am most indebted to William Brock and James Heckman, who jointly chaired my committee; Paul Silver, who programmed the computations and provided many useful comments; and two anonymous referees, whose advice led to significant revisions of a previous draft. The research was supported by NSF grant SES-8107963.

[*Journal of Political Economy*, 1984, vol. 92, no. 6]

© 1984 by The University of Chicago. All rights reserved. 0022-3808/84/9206-0009\$01.50

separation would be, and hence the less likely he is to break the match.

Combined, these two effects yield a decreasing hazard: in other words, the conditional probability of quitting or being laid off falls with greater tenure. For the longer a person has held a job, probably the more specific capital he has acquired (since a worker in whom little capital was invested is more likely to have been filtered out previously). And the higher is his expected output due to job-specific factors, the smaller the chance of a separation occurring.

The decreasing hazard hypothesis is consistent with the observation that certain identifiable groups in the work force turn over more often, experience lower rates of wage growth, and apparently accumulate less specific human capital than others. Furthermore, studying job tenure, several investigators, including Borjas and Rosen (1979), Bartel and Borjas (1981), and Mincer and Jovanovic (1981), have found persistent evidence of positive duration dependence even after they control for socioeconomic characteristics described in the data. That is, among cohorts whom the data measure as being equivalent, the probability of sticking to a job, conditional on past experience, increases with tenure.

Models of job matching and turnover provide a simple framework for analyzing issues related to the accumulation of job-specific capital. An article by Jovanovic (1979) demonstrated that if a person can learn only through specific experience how suited he is to a job, then an (eventually) decreasing hazard emerges from the optimal decision rule. Intuitively, the more experience a person has in a job, the better the assessment of his competence, so the less new information will affect this assessment and prompt the individual to leave it.

The contributions of Jovanovic (1979) and Viscusi (1979) illustrate how optimizing behavior undertaken in learning environments can generate duration dependence that is not merely an artifact of unobserved heterogeneity (defined here as differences in traits across job matches known to the individuals themselves but not recorded in the data and accordingly unobserved by the econometrician). Nevertheless, as Heckman and Borjas (1980) have remarked in another context, data on turnover alone cannot help one to decide whether unobserved heterogeneity or the behavioral rules generated by economic models are causing positive duration dependence found by econometricians. Unless assumptions, typically strong and unjustifiable, specify the probability distribution describing unobserved heterogeneity across the population, a statistical test that chooses between the two alternatives is impossible to devise.<sup>1</sup>

<sup>1</sup> While the discussion of Heckman and Borjas (1980) was concerned with transitions between employment and unemployment, it is clear their conclusions apply more gen-

Thus, work over the last 2 decades has established an interdependence between job-specific capital and labor mobility, formally explained precisely how this interdependence might arise, and produced empirical results that would be broadly consistent with the formal theory but for the presence of unobserved heterogeneity.

The purpose of this paper is to explore the theoretical and empirical ramifications of generalizing the job-matching paradigm. In job-matching models, a Bayesian decision maker combines prior information about each job's characteristics with sample information he has accumulated from working experience on that job in order to forecast its future returns (both pecuniary and nonpecuniary). The next section lays out a framework to accommodate situations where his prior beliefs may differ across jobs. This extension suggests the following definition of an occupation: loosely speaking, two jobs belong to the same occupation if a person's prior beliefs about them are identical. Hence differences between such jobs emerge only through specific experience.

Drawing from the statistical theory of multi-armed-bandit processes, Section III investigates the optimal order to sample jobs and the optimal duration time for each job. (In previous work, the order in which jobs are sampled is immaterial because people have the same prior beliefs over all jobs.) These investigations yield several predictions about life-cycle career patterns. They provide an explanation based on rational behavior for the longstanding empirical regularity, well captured by the expression "young and foolish," that the inexperienced gravitate toward risky ventures. Because younger workers are less experienced than their older counterparts, they more willingly try out jobs where success is rare; older people who have quit them already know they themselves are unsuitable.

A numerical analysis is conducted in Section IV to compute job turnover generated by optimal behavior. The results show that if the quality of job matches is difficult to forecast without specific experience or can be determined quite accurately with only a small amount of experience, turnover is higher.

The empirical parts of this paper, Sections V–VII, focus mainly on two issues. One expects a person's socioeconomic background,

---

erally. The claim is starkly illustrated by considering the sequence of probabilities for breaking a job match at specified tenure levels, derived from sequentially maximizing an intertemporal utility function subject to a production and information technology. Call this sequence  $\{p_i\}_{i=1}^{\infty}$ . Contrast this setup with a very simple model in which a population of identical agents work in various jobs  $\{i\}_{i=0}^{\infty}$ , which are labeled according to the fixed time they take to complete. Suppose the proportion of jobs that last  $i$  periods is  $p_i$ . Then clearly turnover behavior generated by one model is indistinguishable from the other.

through its effects on his prior beliefs, to determine in part how frequently he switches jobs. Thus, differences observed across socio-economic groups in layoffs and quits might be attributable to optimal decisions induced by different levels of schooling acting on prior beliefs. Accordingly the first set of hypotheses tests whether schooling is a significant parameter in determining them. The second issue is whether, from an econometrician's viewpoint, each person spends his entire working career in what is defined here as a single occupation. If so, he has the same prior beliefs about the job-matching characteristics of successive employment positions held, and the theoretical portion of this paper would receive scant empirical support. Similarly the practice of ignoring occupational effects, widespread in related work on labor force participation and search unemployment, would be indirectly upheld. A second set of hypotheses addresses this issue.

Section V provides some informal diagnostics on these matters from the Coleman-Rossi data set. Then in Section VI a maximum likelihood estimator is constructed to mesh the turnover data with the probabilities generated numerically for an economy where nobody switches occupations. Maximum likelihood estimates are obtained and hypothesis testing is conducted in Section VII. Overall, the empirical findings support the view that prior beliefs differ across people and also across the jobs any one person works in.

## II. The Individual's Problem

Perhaps the easiest way to picture the model is to imagine Robinson Crusoe shipwrecked at date zero on an island. With no possibility of rescue foreseen, he maximizes a discounted flow of food output by sequentially choosing one of several available farming techniques to produce it. Through experience, he discovers how well different techniques match the geography of the island, and this knowledge helps him make wiser decisions. It is natural to ask which techniques will be experimented with first, whether all of them will be tried, how long it takes Crusoe to settle down with one (if ever), and so on. The transparent interpretation follows from equating food output with a person's income or utility stream, farming techniques with jobs, and the suitability of those techniques to the island's geography with productivity differences of an individual across different jobs.

To help fix ideas, some notation is introduced at this point. Denote by  $x_{mt}$  the agent's return, which includes both pecuniary and non-pecuniary aspects, from working in the  $m$ th job at time  $t \in T \equiv \{0, 1, 2, \dots\}$ , chosen from the job set  $M$ . Suppose a return is not directly

observed unless the person actually works in the associated job, and furthermore it is the sum of three components:

$$x_{mt} = \psi_t + \xi_m + \sigma_m \epsilon_{mt}. \quad (1)$$

The first term on the right-hand side of (1) denotes a time-trended variable independent of the job and is observed regardless of where the individual works. It may be due to business cycle fluctuations, the effects of age, or general experience and formal education. The second term is a match parameter that is time invariant. The individual does not observe  $\xi_m$  directly, but before acquiring any experience in the job he believes it to be normally distributed  $N(\gamma_m, \delta_m^2)$ . If he had rational expectations, this distribution would describe the latent distribution of talents in that job across the subpopulation of people with socioeconomic characteristics similar to his. The third term is never observed, but  $\sigma_m$  is known and  $\epsilon_{mt}$  is an independent random variable drawn from the standard normal distribution  $N(0, 1)$ . The idea behind the second two terms on the right-hand side of (1) is that match quality cannot be ascertained exactly but must be guessed at by combining prior knowledge about matches of a similar type with the returns history from the current match.

Suppose  $\tau_{mt}$  measures the experience the worker has accumulated in the  $m$ th job by period  $t$ . In other words, letting  $d_{ms}$  be unity if the  $m$ th job is chosen at time  $s$  and zero otherwise,

$$\tau_{mt} \equiv \sum_{s=0}^{t-1} d_{ms}.$$

Using this notation, the total amount of time spent in the  $m$ th job is  $\tau_{m\infty}$ , which is called completed tenure and henceforth abbreviated to  $\tau_m$ .

Both the theoretical and empirical portions of this analysis categorize jobs according to whether outsiders can perceive any differences or not. Individuals evaluate jobs in the same category solely on match-specific grounds, whereas when assessing jobs belonging to different categories, both specific and general attributes are relevant. In this sense the model admits a precise definition of an occupation.

**DEFINITION 1:** Two jobs  $m$  and  $m'$  belong to the same occupation  $n$  if and only if  $(\gamma_m, \delta_m, \sigma_m) = (\gamma_{m'}, \delta_{m'}, \sigma_{m'})$ . For future reference define their common information factor  $\alpha_m \equiv \sigma_m^2 \delta_m^{-2}$ . Also denote the set of jobs within the  $n$ th occupation by  $(n)$ .

The definition implies that a person holds the same prior beliefs about the return stream he would receive if he worked in jobs belonging to the same occupation. Furthermore, the rate at which he learns about such matches is identical.

Given the model's specification, beliefs about the  $m$ th match can be

characterized as a normal distribution  $N(\gamma_{mt}, \delta_{mt}^2)$  where, following Zellner (1971), a Bayesian updating rule implies

$$\gamma_{mt} = \left[ \delta_m^{-2} \gamma_m + \sigma_m^{-2} \sum_{s=0}^{t-1} (x_{ms} - \psi_s) d_{ms} \right] (\delta_m^{-2} + \tau_{mt} \sigma_m^{-2})^{-1},$$

$$\delta_{mt} = (\delta_m^{-2} + \tau_{mt} \sigma_m^{-2})^{-1/2}. \quad (2)$$

Observe from equations (2) that beliefs about a match change only with experience on the job.

The agent's objective is to maximize expected returns from jobs over his lifetime by choosing a decision rule  $d$ , which maps available information into his preferred job choice at each point in time:

$$d(\{\gamma_{mt}, \delta_{mt}\}_{m \in M}) = \times_{m \in M} d_{mt}(\{\gamma_{mt}, \delta_{mt}\}_{m \in M}),$$

where for all  $(m, t) \in M \times T$ ,  $d_{mt} \in \{0, 1\}$  and  $\sum_{m \in M} d_{mt} = 1$ . To summarize, the model is formally characterized by its valuation function,

$$V(\{\gamma_m, \delta_m\}_{m \in M}) = \max_d E_0 \left\{ \sum_{t=0}^{\infty} \sum_{m \in M} \beta^t d_{mt} x_{mt} \right\}, \quad (3)$$

where  $E_t$  stands for expectations conditional on information  $\{\gamma_{mt}, \delta_{mt}\}_{m \in M}$  available at time  $t \in T$ , and  $\beta$  is the discount factor.

### III. Optimal Decision Making

The characterization of the solution to this problem is based on fundamental (but until recently neglected) work by Gittins and Jones (1974).<sup>2</sup> Their theorem on multi-armed-bandit processes implies for this application that each job is associated with an index depending on beliefs about its future returns, and that under certain conditions the optimal decision is to choose the job with the largest index. The optimality of this simple rule is remarkable. In general, which job will be picked depends on the characteristics of all jobs. If there are 10 jobs, 20 parameters must be kept track of (both the mean and the variance for match distributions being updated). The dynamic allocation index (DAI) strategy considerably simplifies the computation of the solution because each index depends on the two characteristics of the job associated with it alone.

The definition below precisely specifies the index for this particular application.

**DEFINITION 2:** Denote by  $DAI_m(\gamma_{mt}, \delta_{mt})$  the dynamic allocation in-

<sup>2</sup> See also Gittins (1979) for a review of work in the area.

dex for the  $m$ th job given beliefs  $(\gamma_{mt}, \delta_{mt})$ . It is the real-valued mapping

$$\text{DAI}_m(\gamma_{mt}, \delta_{mt}) = \sup_{\tau \geq t} \left\{ \frac{E_t \left[ \sum_{r=t}^{\tau} \beta^{r-t} (x_{mr} - \psi_r) \right]}{E_t \left( \sum_{r=t}^{\tau} \beta^{r-t} \right)} \right\}.$$

In other words, the index is the valuation function for the following (fictitious) problem: at time  $t$ , after netting out the general component  $\psi_t$ , choose a stopping time  $\tau$  to maximize the expected return stream from remaining on the  $m$ th job  $E_t[\sum_{r=t}^{\tau} \beta^{r-t}(x_{mr} - \psi_r)]$  divided by the current value of receiving a unitary return over that phase  $E_t(\sum_{r=t}^{\tau} \beta^{r-t})$ .

Proposition 1 formally states how the indices are used in computing the optimal decision rule  $d^o$ . (All propositions are proved in App. A.)

**PROPOSITION 1:** Denoting  $d^o$  as the optimal decision rule for the problem described by equation (3), if  $d_{kt}^o = 1$ , then  $\text{DAI}_k(\gamma_{kt}, \delta_{kt}) = \max_{m \in M} [\text{DAI}_m(\gamma_{mt}, \delta_{mt})]$ . Furthermore, any tie-breaking rule between the maxima is optimal.

The properties of the solution are introduced by briefly characterizing a subcase that appears in the literature on optimal search. Setting  $\sigma_m = 0$  for all  $m \in M$  and  $\psi_t = 0$  for all  $t \in T$  specializes this framework to the normal variant of Weitzman's (1979) analysis of Pandora's problem. (If in addition  $(\gamma_m, \delta_m) = (\gamma_{m'}, \delta_{m'})$  for all  $\{m, m'\} \subseteq M$ , a prototype search model of variety reviewed in Lippman and McCall [1976] is generated.) Equation (1) becomes

$$x_{mt} = \xi_m. \quad (4)$$

Definition 1 now implies that the index value of a sampled job is just its (constant) return. From proposition 1 a very simple career strategy emerges. At date zero rank the indices from highest to lowest and sample jobs, one per period, using this ordering, until the maximum return received in any previous period exceeds the highest-ranking index value among the remaining unsampled jobs; then work in the job associated with this maximal return forever.

Deriving the index values for this specialization is relatively straightforward. Consider a (perhaps hypothetical) job whose match is drawn from a standard normal distribution  $N(0, 1)$ . Call its index value  $D(0, \beta)$ . (The significance of the zero will presently become apparent.)

**DEFINITION 3:** For  $\sigma_m = 0$  define  $D(0, \beta) \equiv \text{DAI}_m(0, 1)$ .

For this specialization all job indices are a linear combination of



$D(0, \beta)$ . Each index is additive in the expected return  $\gamma_m$  and another component  $\delta_m D(0, \beta)$ , the latter increasing proportionately with  $\delta_m$ .

PROPOSITION 2: If  $\sigma_m = 0$  then  $DAI_m(\gamma_m, \delta_m) = \gamma_m + \delta_m D(0, \beta)$ .

The characterization is completed by utilizing proposition 3 to compute  $D(0, \beta)$  numerically. (The top line of table B1 in App. B depicts the mapping  $D(0, \beta)$ , which is monotone increasing in  $\beta$ .)

PROPOSITION 3: Denote by  $\Phi(\epsilon)$  the standard normal cumulative distribution function and  $\Phi'(\epsilon)$  the associated density. Then for each  $\beta \in (0, 1)$  the real number  $D(0, \beta)$  uniquely solves  $D(0, \beta) = \{1 - \beta\Phi[D(0, \beta)]\}^{-1}\beta\Phi'[D(0, \beta)]$ .

The two preceding paragraphs fully operationalize the index rule for a particular specialization. The remainder of this section analyzes its general properties and discusses the economic implications. An implication of proposition 1 is that a person choosing job  $m \in (n)$  will not later switch to another occupation  $n'' \neq n$  at least until all jobs  $m' \in (n)$  have been sampled. The reasoning is as follows: When the worker initially picks  $m \in (n)$ , if at that time the person has no experience in  $m' \in (n)$  either, from definition 1 the index values for  $m$  and  $m'$  are identical. Also, because the worker prefers  $m$  to every job  $m'' \in (n'')$ , by proposition 1 the index value for  $m''$  is less than the index value for  $m$ , and thence  $m'$ . But from equations (2) and definition 2 index values cannot change without specific experience. Therefore, appealing to proposition 1 again,  $m''$  will not optimally be chosen unless the index value for  $m'$  falls as a result of work experience (in  $m'$ ).

COROLLARY 1: Let  $m \in (n)$  and  $m'' \in (n'')$  where  $n \neq n''$ . If  $d_{mt}^o = d_{m''s}^o = 1$  for some  $s > t$ , then for all  $m' \in (n)$ ,  $\tau_{m's} > 0$ .

Suppose, after acquiring experience in  $m \in (n)$ , a new job  $m' \in (n)$  is sampled. The proposition also implies that the worker will not return to  $m$  until every job in  $(n)$  has been sampled. When  $m'$  is first picked, from definition 1, the person is indifferent between  $m'$  and  $m'' \in (n)$  if he has no experience in  $m''$  either: hence by proposition 1 the common index value for  $m'$  and  $m''$  exceeds the index value for  $m$ . Using the same argument applied to derive the first corollary, some experience in  $m''$  must be acquired before its index value could possibly fall below the index value for  $m$ .

COROLLARY 2: Suppose  $\{m, m', m''\} \subseteq (n)$ . If  $d_{mt} = d_{m's} = d_{m''r} = 1$  for some  $t < s < r$ , then  $\tau_{m''r} > 0$ .

Corollary 1 says that whenever a new occupation is entered, all jobs within it are sampled before any occupational shift occurs. From corollary 2 multiple spells within a job will not occur at least until some experience has been acquired in all jobs belonging to the same occupation. So if there are an infinite number of jobs in the first occupation a worker enters, all jobs he ever takes last one spell and belong to that occupation. For this specialization the model essentially

reduces to Jovanovic's (1979) framework; heterogeneity across the population becomes the sole reason for people to work in different occupations.

**COROLLARY 3:** Let  $n(\#)$  denote the number of jobs belonging to occupation  $n$ . If  $m' \in (n)$ ,  $n(\#) = \infty$ , and  $d_{m'0} = 1$ , then

$$\tau_m = \begin{cases} \min_{s \in T} \{\tau_{ms} | d_{ms} = 1, d_{m(s+1)} = 0\} & m \in (n) \\ 0 & m \notin (n). \end{cases}$$

The properties of the optimal decision rule  $d^o$  can be further determined by examining the properties of the mappings  $\text{DAI}_m(\gamma_m, \delta_m)$ . In the present application, an index can be expressed more simply than in definition 2. Central to this simplification is the standard index  $D(\sigma_m^2, \beta)$ .

**DEFINITION 4:** For all  $(\sigma_m, \beta)$  define  $D(\sigma_m^2, \beta) \equiv \text{DAI}_m(0, 1)$ .

Less formally, the standard index is the DAI for a job whose match parameter  $\xi_m$  is drawn from the standard normal distribution and whose return net of the general component  $(x_{mt} - \psi_t)$  is perturbed by a normal random variate  $\epsilon_{mt}$  with variance  $\sigma_m^2$ . Whether such a job actually exists is immaterial; the purpose of inventing the notation is to provide a reference point against which all indices are judged.

**PROPOSITION 4:**  $\text{DAI}_m(\gamma_{mt}, \delta_{mt}) = \gamma_{mt} + \delta_{mt}D(\alpha_m + \tau_{mt}, \beta)$ .

Intuitively, the index for any job  $m \in M$  is just the sum of its expected return (net of the general component) plus a term  $\delta_{mt}D(\alpha_m + \tau_{mt}, \beta)$ , which represents the value of the job as a source of information about itself. (Recall from definition 1 that  $\alpha_m$  is the information factor.) Clearly the properties of  $\text{DAI}_m(\gamma_{mt}, \delta_{mt})$  follow from those of  $D(\alpha_m + \tau_{mt}, \beta)$ , since given  $(\gamma_{mt}, \delta_{mt})$ , the former is just a linear transformation of the latter, and although  $D(\alpha_m + \tau_{mt}, \beta)$  does not admit a closed-form representation, it can be computed numerically.

Appendix B describes the algorithm for computing  $D(\alpha_m, \beta)$ . It is based on the following proposition.

**PROPOSITION 5:** Let  $F$  denote the space of bounded continuous real valued functions taking their domain on the space  $(-\infty, \infty) \times (0, \infty)$ . For all  $f \in F$  define  $C$ , a mapping from  $F$  to itself, as

$$C[f(\gamma, \alpha)] = \beta \int_{-\infty}^{\infty} \max(0, (1 - \beta)\{\gamma + \epsilon[\alpha(\alpha + 1)]^{-1/2} \\ + f\{\gamma + \epsilon[\alpha(\alpha + 1)]^{-1/2}, \alpha + 1\})d\Phi(\epsilon).$$

There exists a unique function  $g(\gamma, \alpha) \in F$  satisfying  $g(\gamma, \alpha) = C[g(\gamma, \alpha)]$ . Moreover, defining  $C_k(f)$  recursively as  $C[C_{k-1}(f)]$ , for all  $f \in F$ ,  $\|C_k(f) - g\| \leq (1 - \beta)^{-1}\|C_k(f) - C_{k-1}(f)\|$ . Finally for all  $(\alpha, \beta)$

$\in (0, \infty) \times (0, 1)$  the standard index solves  $D(\alpha, \beta) = (1 - \beta)^{-1} \alpha^{1/2} g[-\alpha^{-1/2} D(\alpha, \beta), \alpha]$ .

Briefly, the function  $g(\gamma, \alpha)$  is approximated by recursively applying  $C$  to the zero function. Since the final inequality uniformly bounds the approximation error, this method provides the means for computing  $g$  to an arbitrary degree of accuracy. The standard index  $D(\alpha, \beta)$  is then solved for each  $(\alpha, \beta)$  from the equation in  $g$ , using a fixed-point routine.

Figure 1 illustrates the qualitative features obtained from a numerical analysis of  $D(\alpha_m, \beta)$  described in Appendix B (which also reports the results in table B1). In the top right quadrant, the standard index  $D(\alpha_m, \beta)$  is mapped as a function  $\alpha_m$  for a particular value of  $\beta$ . On the left side of the horizontal axis, the product  $\delta_{mt} D(\alpha_m + \tau_{mt}, \beta)$  is graphed, while the lower portion of the vertical axis is, from proposition 4,  $DAI_m(\gamma_{mt}, \delta_{mt})$ .

From an economist's perspective, six main results emerge. The current job yields two benefits: payoff from expected reward for the period and information about future returns within that particular job. Naturally, a person will opt for the job with the bigger expected reward if the value from the informational component is the same for the two alternatives. This first result is reflected in the third quadrant of figure 1. Raising  $\gamma_{mt}$  lowers the intercept on the  $q_4$  axis.

Supposing expected current returns are equal, he would choose the job that provides the most valuable information. Information value is embodied within the term  $\delta_{mt} D(\alpha_m + \tau_{mt}, \beta)$ . It depends on three factors: what is already known, given by the precision  $\delta_{mt}^{-2}$ , the rate of acquisition measured by  $\sigma_m$ , and preferences over time  $\beta$ . The better one's initial information about suitability and hence the lower  $\delta_m$ , the less one is prepared to forgo in order to find out more about the job. Its attraction as a source of knowledge accordingly declines with tighter priors. This is the second result. In the diagram, raising  $\delta_m$  would reduce both the slope of the line in the top left quadrant and also  $\alpha_m$  (from def. 1), the latter effect causing  $D(\alpha_m, \beta)$  to increase. Hence the product  $\delta_m D(\alpha_m, \beta)$  would increase too.

Furthermore, greater variability in the payoff sequence itself retards learning about the underlying process; that is, prior beliefs are not affected as much by any given amount of on-the-job experience when the latter is an unreliable guide for measuring competence in an occupation. That the informational component is therefore less valuable constitutes a third result. Diagrammatically, if  $\sigma_m$  increases,  $\alpha_m$  increases and  $D(\alpha_m, \beta)$  falls.

Fourth, specific experience  $\tau_{mt}$  reduces  $D(\alpha_m + \tau_{mt}, \beta)$  at a decreasing rate. Also, from equation (2),  $\delta_{mt}$  is decreasing in  $\tau_{mt}$ . Accordingly, the ray from the origin in the top left quadrant swivels clockwise as  $\tau_{mt}$

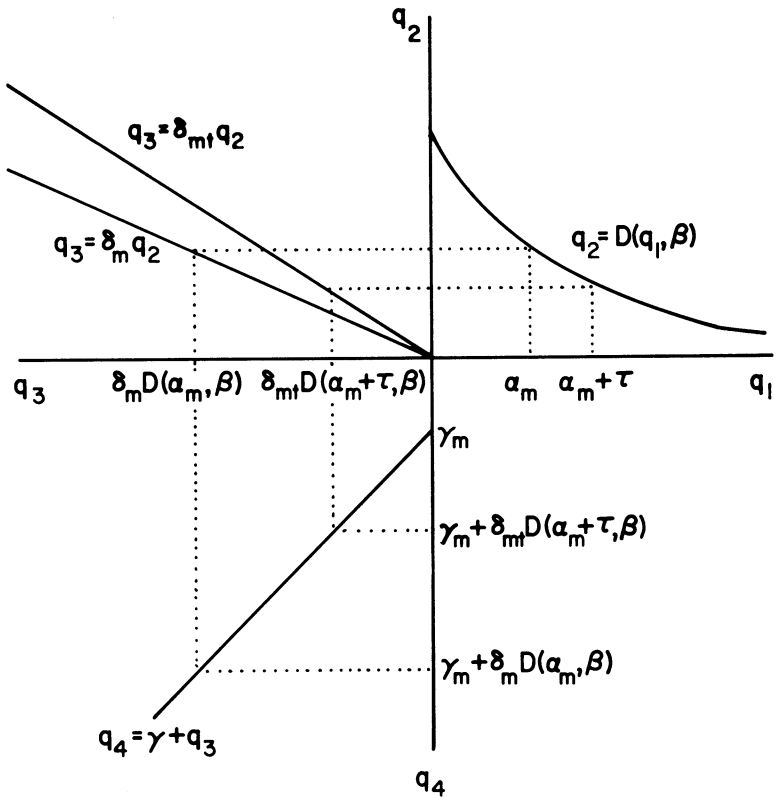


FIG. 1.—The dynamic allocation index

increases. Consequently the product  $\delta_{mt}D(\alpha_m + \tau_{mt}, \beta)$  falls. Thus the value of additional information declines as it is acquired.

In addition, matches about which a precise estimate can be ascertained more quickly lose their attractiveness as an information source sooner. To see this point consider the limiting case when a single-period tryout perfectly reveals match quality, that is, when  $\sigma_m = 0$ . With reference to the top left quadrant, as  $\sigma_m$  approaches zero,  $(\delta_m^{-2} + \sigma_m^{-2})^{1/2}$  approaches infinity, so using (2) the ray from the origin  $q_2 = \delta_{m1}q_3$  becomes vertical, folding into the  $q_2$  axis and making the index value  $DAI_{m1}(\gamma_{m1}, 0)$  equal to the true match parameter  $\xi_m$ . Obviously, this would not occur if  $\sigma_m > 0$ .

Finally, people who discount future payoffs more value information less. Intuitively, this result follows directly from the fact that information about a job match is a form of human capital. In the diagram, decreasing  $\beta$  reduces the standard index  $D(\alpha_m, \beta)$  for all  $\alpha_m$ .

These findings yield predictions about life-cycle behavior and the

composition of labor markets. A theory of equalizing differences would imply that jobs with high informational benefits pay less on average in equilibrium. They attract the inexperienced who relatively quickly discover their personal match. Workers who have passed through the experimentation stage will not be found in such jobs unless their match is uncommonly good. Jobs yielding a great deal of information about matches therefore tend to be filled mainly by a transient inexperienced population, a cross-sectional average of whom earn modest returns, plus a much smaller number of experienced, very successful, permanent workers. The performing arts sector probably fits into this category.

On the other hand, jobs with fewer informational benefits tend to exhibit less turnover, higher-than-average returns to entrants, and a greater concentration of incomes about the mean.

A young worker is therefore more likely to quit his job for two reasons. First, there is a lower probability that within a particular occupation his match will be superior. Typically an older worker has had more time to find one. Second, a greater percentage of the young are found in occupations where high turnover is endemic. Because they are less self-aware than older people, young people are more willing to experiment in activities that increase their self-awareness.

Ever since Adam Smith observed the attraction to youth of some occupations yielding risky returns, he and others following him have explained the phenomenon in terms of a taste for the uncertain, overoptimistic expectations about ability, or status.<sup>3</sup> This section has demonstrated what types of uncertainty attract the inexperienced, explaining their attraction in terms of demand for information, without recourse to tastes or irrational expectations.

#### IV. Job Turnover

Roughly speaking, job turnover is attributable to occupational sources and match-specific factors. Since both considerations determine the ranking of index values, the probability of choosing a certain job at some future time potentially depends on current beliefs about all the jobs. Therefore, a complete characterization of equilibrium turnover

<sup>3</sup> "The contempt of risk and the presumptuous hope of success, are in no period of life more active than at the age at which young people chuse their professions" (Smith 1937, p. 109). In a similar vein, Marshall (1948) writes, "if an occupation offers a few extremely high prices, its attractiveness is increased out of all proportion to their aggregate value. For this there are two reasons. The first is that young men of an adventurous disposition are more attracted by the prospects of a great success than they are deterred by the fear of failure; and the second is that the social rank of an occupation depends more on the highest dignity and the best position which can be attained through it than on the average good fortune of those engaged in it" (1948, p. 554).

rates presents one with a formidable task, requiring the numerical computation of sequences of multiple integrations over the belief space  $\{\gamma_{mt}, \delta_{mt}\}_{m \in M}$ . Nevertheless some progress can be made at moderate cost.

Considering job  $m \in (n)$  held at time  $t \in T$ , suppose there exists another one belonging to the same occupation  $m' \in (n)$  in which the person has no experience. That is,  $\tau_{m't} = 0$ . By corollary 2 the  $m$ th match has not previously been interrupted: at time  $t$  the worker is still in his first spell. By corollary 1 if and when he quits, he will switch to a new job in (the same) occupation  $n$ . This section analyzes the discrete hazard function associated with such transfers. They are especially significant when  $n(\#) = \infty$ . For appealing to corollary 3, the  $m$ th match lasts only one spell, and hence in that case the discrete hazard for the first spell is identically the hazard for the job match.

DEFINITION 6:  $h_{m\tau} \equiv P\{\gamma_{mt} + \delta_{mt}D(\alpha_m + \tau_{mt}, \beta) < \gamma_m + \delta_m D(\alpha_m, \beta) | \tau_{mt} = \tau\}$ .

Let  $m \in (n)$ . The statement  $\tau_{mt} = \tau$  implies from proposition 1 that, if  $\tau_{m't} = 0$  for some  $m' \in (n)$ , given optimization, the inequality  $\gamma_{ms} + \delta_{ms}D(\alpha_m + \tau_{ms}, \beta) \geq \gamma_m + \delta_m D(\alpha_m, \beta)$  holds for all  $\tau_{ms} < \tau_{mt}$ . The fact that at time  $t$  this inequality is reversed implies that it is no longer optimal to choose  $m$  over  $m'$ .

When  $\sigma_m = 0$ , match quality is perfectly revealed after one period's experience; hence  $DAI_m(\gamma_{m1}, \delta_{m1}) = DAI_m(\gamma_{m1}, 0) = \xi_m$ . Noting that beliefs above  $\xi_m$  are initially normally distributed  $N(\gamma_m, \delta_m^2)$ , proposition 2 and definition 6 imply, for this specialization,

$$h_{m1} = P\{\xi_m < \gamma_m + \delta_m D(0, \beta)\} = \Phi[D(0, \beta)].$$

Furthermore, since no information is acquired after working more than one period,  $h_{m\tau} = 0$  for all  $\tau > 1$ .

To derive the discrete hazard for cases where match quality is only imperfectly revealed by returns, the probability distribution  $\Psi_{m\tau}(\rho)$  of the transformed means  $(\gamma_{mt} - \gamma_m)\sigma_m^{-1}$  is required.

DEFINITION 7:  $\Psi_{m\tau}(\rho) \equiv P\{\gamma_{mt} < \gamma_m + \rho\sigma_m | \tau_{mt} = \tau\}$ .

The hazard  $h_{m\tau}$  and the distribution  $\Psi_{m\tau}(\rho)$  may be recursively computed for successive values of  $\tau \in T$ . First the distribution  $\Psi_{m1}(\rho)$  is found by updating the prior transformed mean (of zero) with a realization from the standard normal distribution  $\Phi(\epsilon)$ . If the transformed posterior mean lies below the critical value  $[\alpha_m^{-1/2}D(\alpha_m, \beta) - (\alpha_m + 1)^{-1/2}D(\alpha_m + 1, \beta)]$  the spell ends. Hence the proportion of spells ending after one period's experience  $h_{m1}$  is  $\Psi_{m1}[\alpha_m^{-1/2}D(\alpha_m, \beta) - (\alpha_m + 1)^{-1/2}D(\alpha_m + 1, \beta)]$ . The probability distribution  $(1 - h_{m1})^{-1}[\Psi_{m1}(\rho) - h_{m1}]$  now forms the distribution of prior means for beliefs about remaining, uncompleted spells with one period's tenure. The procedure outlined above is repeated to find  $\Psi_{m2}(\rho)$ , and so on.

PROPOSITION 6: For all  $(m, \tau) \in M \times T$ , if  $\sigma_m > 0$ , then

$$\begin{aligned} h_{m\tau} &= \psi_{m\tau}[\alpha_m^{-1/2}D(\alpha_m, \beta) - (\alpha_m + \tau)^{-1/2}D(\alpha_m + \tau, \beta)] \\ \psi_{m1}(\rho) &= \Phi[\rho\alpha_m^{1/2}(\alpha_m + 1)^{1/2}] \\ \psi_{m(\tau+1)}(\rho) &= (1 - h_{m\tau})^{-1} \left\{ \int_{-\infty}^{\infty} \psi_{m\tau}[\rho - \epsilon(\alpha_m + \tau)^{1/2} \right. \\ &\quad \left. \times (\alpha_m + \tau + 1)^{1/2}] d\Phi(\epsilon) - h_{m\tau} \right\}. \end{aligned}$$

Appendix C describes how the sequence  $\{h_{m\tau}\}_{\tau=1}^5$  was computed over a grid  $(\alpha_m, \beta)$  lying in the space  $[0.1, 17.1] \times [0.1, 0.95]$  from values obtained for the standard index  $D(\alpha_m, \beta)$  and reports the results. For any  $(\alpha_m, \beta)$  coordinate pair the sequence declines monotonically and has a long tail for reasons explained in the Introduction. Three further results are depicted in table C1.

First, the more future payoffs are discounted, the lower the hazard, because the value of finding a superior job match is diminished. Thus, lower values of  $\beta$  uniformly shift the hazard down.

Second, increasing the information factor  $\alpha_m$  lowers the hazard. Recall from definition 1 that  $\alpha_m \equiv \sigma_m^2 \delta_m^{-2}$ . Therefore if the quality of job matches is difficult to forecast without specific experience (meaning  $\delta_m$  is high) or can be determined quite accurately from just a small amount of working experience (which occurs when  $\sigma_m$  is low), turnover is higher.

Third, over 50 percent separate with tenure of 1 period.

COROLLARY 4:  $h_{m1} = \Phi\{[D(\alpha_m, \beta) - D(\alpha_m + 1, \beta)]\alpha_m^{1/2}(\alpha_m + 1)^{1/2}\} > 1/2$ .

The economic intuition behind this corollary is also straightforward. By definition 1 an agent is indifferent between two untried jobs belonging to the same occupation. After one is taken, its value as an information source declines after a period's experience, when a piece of information about match quality is received. In addition (given rational expectations), the normally distributed information received is unfavorable half the time. So the job's index value is more likely to fall than rise. When this happens, the agent completes the match next time he chooses the occupation.

## V. Data

When any theory is confronted with data, questions about how to calibrate its primitives immediately arise. Here, three theoretical concepts have empirical counterparts in the data, namely jobs, occupations, and returns.

The only type of learning admitted by the model is about job matches. Ideally an empirical application should therefore view jobs

in terms of those activities where the most task-specific learning takes place. This point is best made by example. Given the discipline musical scores impose on concert renditions, a violinist's most important match is neither to his orchestra nor to his conductor, but to his instrument. Therefore, treating successive orchestras or conductors as separate job matches would yield poorer empirical results than if the position of violinist itself were regarded as one job match. Needless to say, the detail required to make such an approach operational is not typically available, and the crude employment categories utilized in this study severely limit the inferences one can make.

Two interpretations compatible with the model are that the return  $x_{mt}$  represents the utility a person gets in period  $t$  from working in job  $m$  after purchasing consumption goods with his pecuniary income, or a monetary return plus a value of on-the-job amenities implicitly purchased. The first interpretation is valid if financial markets are closed to agents, while the second applies if the supply of credit is perfectly elastic. Much applied work measures returns with wages, but their link to both interpretations is somewhat tenuous. First, the non-pecuniary aspects of a job, an important consideration in a matching context, are ignored. Second, if financial markets are open, as Jovanovic (1979) notes, the wage contract is nonunique, so wages do not necessarily correspond to realized output. For these reasons wage data were not used; this also means the econometric results are compatible with either theoretical interpretation mentioned above.

The Coleman-Rossi data set was used to quantify the theory and assess its empirical significance. A detailed summary of the data, previously used by Bartel (1979) in her work on migration, has been provided by Ornstein (1976). Very briefly, information was gathered in the winter of 1969 through one interview with each of 822 white and 767 nonwhite men. Among other things, these men were questioned extensively about their entire work history, education, and family background. Appendix D describes the criteria met by the 467 white men making up the sample used in this study.

Table 1 illustrates tenure in and turnover between spells by employment group and education level. Here tenure is measured biennially.<sup>4</sup> The number of people belonging to the sample who were working in each group when the survey was taken is recorded in column 1. Columns 2–4 successively measure the number of people at different tenure levels in their spell by employment group and educational attainment as a percentage of the corresponding total in

<sup>4</sup> Ideally the period length (which determines the arrival rate of new information) should be simultaneously estimated along with the other parameters of interest, but this extension remains a topic for future research.



TABLE 1  
TENURE AND TURNOVER BY EMPLOYMENT AND EDUCATION

	CURRENT POSITION				PAST SPELLS			
	Number	Percentage with Tenure of			Number	Empirical Hazard		
		≥ 2	≥ 3	≥ 4		1	2	3
Employment:								
Professional	67	76	65	31	183	61	49	65
Farm owner	22	95	90	9	44	55	50	30
Manager	80	80	73	33	128	60	55	61
Clerk	40	82	67	35	175	69	55	44
Salesman	27	77	62	29	138	64	51	54
Craftsman	107	81	65	25	379	61	53	59
Operative	84	80	78	39	553	68	59	53
Serviceman	13	92	61	46	60	73	63	33
Farm laborer	6	83	83	33	144	72	54	63
Nonfarm laborer	21	76	57	33	281	78	55	39
Education:								
Grade school	177	84	75	28	779	70	55	64
High school	113	81	67	33	566	68	58	42
College	84	76	67	35	463	61	50	50

column 1. For example, of the 67 professionals interviewed, 76 per cent had current tenure of at least 2 biennial periods (4 years) in 1969. Column 5 is the number of previous, or completed, spells held by the sample population, while the last three columns, the empirical hazards, categorize them by spell length. They show the number of quits at specified tenure levels as a percentage of past spells lasting at least as long. Thus 58 percent of previous spells involving a high school graduate and lasting at least 2 periods ended after 2 periods. Some indication of the underlying characteristics of employment groups is given by these empirical hazards. If there were only a single occupation in an economy populated by identical agents, the corresponding percentages would be roughly coincidental. However, even at this highly aggregated level, there are noticeable divergences.

The apparent differences in table 1 between empirical hazard rates across employment categories could result from the existence of several occupations associated with those groupings. Alternatively, such differences might be wholly attributable to a heterogeneous population. Under the latter hypothesis, job turnover rates depend on lifetime socioeconomic characteristics, for which employment groups serve as a proxy. One interesting specialization of this hypothesis is that people with common socioeconomic characteristics belong to the same employment group. If true, no switching between employment groups would occur. A matrix of relative frequencies depicting transi-

TABLE 2  
TRANSITIONS WITHIN AND BETWEEN EMPLOYMENT GROUPS

	Professional	Farm Owner	Manager	Clerk	Salesman	Craftsman	Operative	Serviceman	Farm Laborer	Nonfarm Laborer
Professional (183)	67	1	11	4	4	5	5	1	0	1
Farm owner (44)	0	25	2	2	2	9	39	2	14	5
Manager (128)	11	2	39	4	20	10	9	1	1	3
Clerk (175)	10	0	14	33	7	11	15	2	0	7
Salesman (138)	1	1	27	6	30	9	17	4	0	5
Craftsman (379)	5	0	7	6	5	48	18	2	2	7
Operative (553)	4	3	5	6	4	19	38	3	4	14
Serviceman (60)	3	0	5	8	7	10	30	18	3	15
Farm laborer (144)	2	8	1	1	2	8	28	2	31	16
Nonfarm laborer (281)	1	2	2	8	2	18	40	3	1	22

tions between employment groups was accordingly constructed from the job changes indicated by the data. In cell  $(i, j)$  of table 2 (where  $i$  designates the row and  $j$  the column) is the percentage of completed spells belonging to the  $i$ th employment category that were followed by a spell in the  $j$ th. On dividing the matrix elements by 100, the specialization above yields an identity matrix, which is plainly rejected by table 2.

A third summary measure of the data is the spell count per job. If there were only one occupation and if jobs (as recorded by the data) were synonymous with matches, from corollary 3 multiple spells would never be observed. However, table 3 exhibits the contrary.

Notwithstanding the highly aggregated form of the data, these

TABLE 3  
SPELLS NOT INVOLVING NEW JOBS (by Employment Group)

Professional	Farm Owner	Manager	Clerk	Salesman	Craftsman	Operative	Serviceman	Farm Laborer	Nonfarm Laborer
10	12	8	3	4	13	19	1	1	18

diagnostics provide weak, informal evidence against the hypothesis of a one-occupation economy (as given in def. 1). The empirical hazards differ markedly across the various categories, people do switch groups, and multiple spells within jobs occur.

Several caveats are, however, in order. Although half the spells recorded as completed end within a single (2-year) period and the empirical hazard for completed matches declines, the relative frequencies are not directly comparable with data in Appendix C, because the former are computed from censored data. Moreover, the top portion of table 1 makes no allowance for population heterogeneity. For this reason the probability transition matrix has limited value too: it is hardly surprising that the specialization is not confirmed by the data, given its unreasonable premises. Finally, the assumption implicit in table 3 that positions are synonymous with matches may be questioned; returning to a position previously held can entail another boss, more modern machinery, and/or different duties. In such cases it might be reasonable to presume a new match is beginning.

VI. Econometrics

The pitfalls of interpreting data in such an unstructured way underscore the value of proceeding more formally. In this latter respect, the aim here is twofold. First, the structural parameters for a one-occupation model are estimated by exploiting the economic and numerical analyses already undertaken. In contrast to the empirical hazards depicted in table 1, the hazards estimated below control for both observed and unobserved heterogeneity. Moreover, they are not subject to censoring bias. Second, likelihood ratio tests are conducted to determine the significance of heterogeneity and also whether incorporating employment group dummies increases the explanatory power of the model; under the null hypothesis of one occupation the collective contribution of the employment dummies would be insignificant.

The discrete hazards, as given by definition 6, form the basis of the

likelihood function  $L$ . Let  $p_\tau(\alpha, \beta)$  denote the unconditional probability of a person with discount factor  $\beta$  working  $\tau$  periods in a new job with information factor  $\alpha$  before switching to another new job in the same occupation. That is,

$$p_\tau(\alpha, \beta) \equiv h_\tau(\alpha, \beta) \prod_{s=1}^{\tau-1} [1 - h_s(\alpha, \beta)].$$

Then the joint probability of spell duration times observed across the sample is

$$L = \prod_{j \in J} \prod_{i \in (j)} \left\{ \sum_{\tau \in s(i)} [p_\tau(\alpha_{1i}, \beta_{1j})\lambda + p_\tau(\alpha_{2i}, \beta_{2j})(1 - \lambda)] \right\}.$$

There are  $J$  workers, and associated with the  $j$ th person are  $(j)$  job spells; in the one-occupation model each spell  $i \in (j)$  corresponds to a different match. Denote by  $s(i)$  the information contained in the data set about the duration of the spell associated with the  $i$ th observation. For the  $i$ th observation, the econometrician might observe a spell ending at tenure  $\tau$ ; then  $s(i) = \tau$ . So far as the current spell is concerned, the econometrician observes it lasting  $\tau$  periods or more; in this case  $s(i) = \{\tau, \tau + 1, \dots\}$ .

Assuming momentarily that the parameters  $(\alpha_i, \beta_j)$  are either  $(\alpha_{1i}, \beta_{1j})$  or  $(\alpha_{2i}, \beta_{2j})$ , for some  $\lambda \in \{0.6, 0.7, 0.8, 0.9, 1.0\}$ , the probability of the event  $s(i)$  is  $\sum_{\tau \in s(i)} [p_\tau(\alpha_{1i}, \beta_{1j})\lambda + p_\tau(\alpha_{2i}, \beta_{2j})(1 - \lambda)]$ . Conditional on  $\{\alpha_{1i}, \alpha_{2i}, \beta_{1j}, \beta_{2j}\}_{i \in (j)}$ , the maximum likelihood estimate for  $\lambda$  can be obtained from  $L$ . This approach explicitly recognizes that certain variables relevant in determining match quality, like physical stamina, emotional makeup, and work environment, are measured by the data only poorly, if at all. Their joint effect is modeled here as a Bernoulli distribution. So with probability  $\lambda$  the  $i$ th spell falls into the first unobserved state, and with probability  $(1 - \lambda)$  into the second.<sup>5</sup>

Naturally, the set  $\{\alpha_{1i}, \alpha_{2i}, \beta_{1j}, \beta_{2j}\}_{i \in (j)}$  is not observed for any  $i \in (j)$ , although vectors of dummy variables  $z_i$  and  $z_j$  describing the socioeconomic characteristics of each spell and each person are observed. Accordingly, the information and discount factors are assumed to be linear mappings of these observed characteristics:

$$\begin{aligned} \alpha_{ki} &= z_i' a_k, & k &= 1, 2 \\ \beta_{kj} &= z_j' b_k, & k &= 1, 2. \end{aligned} \tag{5}$$

<sup>5</sup> This form of unobserved heterogeneity has a mover-stayer interpretation. Alternatively it may be viewed as an approximation to the nonparametric methods proposed by Heckman and Singer (1982, 1984). Given the dimension of the parameter space and the model's structure, a two-point distribution with interactions was found reasonable for the sample size under consideration.

TABLE 4  
MAXIMUM LIKELIHOOD ESTIMATES FOR A ONE-OCCUPATION ECONOMY

EDUCATION	STRUCTURAL PARAMETERS				
	$\alpha_1$	$\alpha_2$	$\beta_1$	$\beta_2$	$\lambda$
Grade school	3.1	16.1	.95	.95	.7
High school	3.1	15.1	.95	.95	.9
College	16.1	16.1	.95	.95	...

Substituting (5) into  $L$ , maximum likelihood estimates for  $a_1, a_2, b_1, b_2$ , and  $\lambda$  are found by evaluating the likelihood at all points belonging to the discrete parameter space depicted in table C1.

VII. Empirical Findings

Table 4 presents the maximum likelihood estimates for a one-occupation economy. Job matches are classified according to the educational attainment of the worker involved, and unobserved heterogeneity is modeled as a binary random variable. Two findings are noteworthy.

First, the effect of education is nonmonotone. After margining out the unobserved heterogeneity, the maximum likelihood estimates of the information factor are 7, 4, and 16 for the respective educational groups. This result is not surprising. Higher educational attainment is conceivably associated with faster learning about match quality, greater opportunities for specialization, and more informed career choices. On the other hand, general education may help compensate for low aptitude in deciphering match quality. These effects respectively lower  $\sigma$ , raise  $\delta$ , lower  $\delta$ , and raise  $\sigma$ , rendering the net effect on  $\alpha$  ambiguous.

Second, the estimates of the discount factor, uniformly high and furthermore equal to the upper bound of the grid, may indicate a misspecification. Because the theory allows only for job severances due to poor matching, turnover attributable to other causes probably induces upward bias in the estimated hazards (which are determined from table C1 using the estimates of table 4). Since the hazard is increasing in the discount factor, such upward bias might be responsible for the unrealistically high estimates of  $\beta$  obtained.

To gauge the importance of education and the viability of the assumption that people never switch occupations, seven other variations on the basic framework were run. Four permutations of the model are generated by successively omitting and including education and/or unobserved heterogeneity. In addition, the information factor was

TABLE 5  
CONTROLLING FOR EDUCATION ( $\chi^2_{.05}(q)$ , LRTS)

EMPLOYMENT	UNOBSERVED HETEROGENEITY	
	No	Yes
No	(9.5, 792.4)	(18.3, 785.9)
Yes	(33.9, 796.8)	(64, 809.0)

TABLE 6  
CONTROLLING FOR EMPLOYMENT GROUP ( $\chi^2_{.05}(q)$ , LRTS)

EDUCATION	UNOBSERVED HETEROGENEITY	
	No	Yes
No	(16.9, 77.3)	(28.9, 69.3)
Yes	(40.1, 81.7)	(66, 81.5)

allowed to depend on the employment group in four of the eight variations.

Table 5 depicts the results of likelihood ratio tests that examine the significance of education. The data support the view that educational attainment is associated with different beliefs and learning speeds, generating systematic differences in turnover behavior. The second entry in each cell LRTS (likelihood ratio test statistic) is minus twice the log likelihood ratio, while the first is the corresponding boundary for a .05 size test calculated from a  $\chi^2$  distribution adjusted for the appropriate degrees of freedom,  $q$ . Regardless of whether unobserved heterogeneity is accounted for or whether employment dummies are included, the null hypothesis of no effect is (asymptotically) rejected.<sup>6</sup>

Perhaps the main finding in this section is that employment group dummies have an impact. Table 6 shows that, regardless of whether unobserved heterogeneity and educational differences are incorporated, the null hypothesis—that the information factor does not depend on the employment category of the match—is rejected. Since individuals are mobile between employment categories (as seen from table 2), explaining this dependence in terms of unobserved

<sup>6</sup> It may also be worth mentioning that when unobserved heterogeneity is introduced in the presence of schooling and employment group dummies the improvement in the maximum value of the log likelihood is negligible (from  $-2,202$  to  $-2,201$ ).

heterogeneity, while a logical possibility, seems implausible. More reasonable is the conclusion that the employment categories, rough as they are, actually indicate something about the nature of the work involved and that the characteristics of matches (or, more generally, job-specific capital) differ across these categories. Given this conclusion, it follows that rational individuals will consider the sorts of issues analyzed by the theory above as they move through their working lives.

VIII. Conclusion

This paper extends the matching literature by formally incorporating the notion that people try out several jobs, perhaps from different occupations, depending on how things go. It expositis the optimal decision rule, showing which jobs should be sampled first and for how long. Then the turnover rate is derived numerically for an economy where, after controlling for socioeconomic traits, all differences between jobs are attributable to match-specific considerations. The parameters of such an economy are estimated using panel data, and the hypothesis that people do not switch occupations is rejected. The empirical findings bolster the theoretical portion of the paper. It argues that since the value of specific experience varies across job types, optimizing behavior induces a stochastic career profile: jobs yielding returns that are subject to a particular form of uncertainty are experimented with first. To the casual social commentator, these jobs seem underpaid, and the young, inexperienced workers engaged in them look gullible or foolhardy; but in point of fact, this may be far from the truth.

Appendix A

This appendix proves the propositions stated in the text. In proving them, frequent use is made of the following lemma.

LEMMA 1: For all  $m \in M$

$$DAI_m(\gamma_{mt}, \delta_{mt}) = \sup_{\tau \geq t} E_t \left[ (1 - \beta) \sum_{r=t}^{\tau} \beta^{r-t} (x_{mr} - \psi_r) + \beta^{\tau-t+1} DAI_m(\gamma_{mt}, \delta_{mt}) \right].$$

PROOF: From definition 2,

$$DAI_m(\gamma_{mt}, \delta_{mt}) = \sup_{\tau \geq t} \left\{ \frac{E_t \left[ \sum_{r=t}^{\tau} \beta^{r-t} (x_{mr} - \psi_r) \right]}{E_t \left( \sum_{r=t}^{\tau} \beta^{r-t} \right)} \right\}$$

$$0 = \sup_{\tau \geq t} \left\{ \frac{E_t \left[ \sum_{r=t}^{\tau} \beta^{r-t} (x_{mr} - \psi_r) \right]}{E_t \left( \sum_{r=t}^{\tau} \beta^{r-t} \right)} - \text{DAI}_m(\gamma_{mt}, \delta_{mt}) \right\}$$

(subtracting  $\text{DAI}_m(\gamma_{mt}, \delta_{mt})$  from both sides)

$$= \sup_{\tau \geq t} E_t \left\{ \left[ \sum_{r=t}^{\tau} \beta^{r-t} (x_{mr} - \psi_r) \right] - \left( \sum_{r=t}^{\tau} \beta^{r-t} \right) \text{DAI}_m(\gamma_{mt}, \delta_{mt}) \right\}$$

(multiplying through by  $[E_t \sum_{r=t}^{\tau} \beta^{r-t}]$ )

$$= \sup_{\tau \geq t} E_t \left\{ \left[ \sum_{r=t}^{\tau} \beta^{r-t} (x_{mr} - \psi_r) \right] - (1 - \beta)^{-1} (1 - \beta^{\tau-t+1}) \text{DAI}_m(\gamma_{mt}, \delta_{mt}) \right\}$$

since  $(1 - \beta)^{-1} (1 - \beta^{\tau-t+1}) = \sum_{r=t}^{\tau} \beta^{r-t}$ . The proof is completed upon multiplying through by  $(1 - \beta)$  and adding  $\text{DAI}_m(\gamma_{mt}, \delta_{mt})$  to both sides.

PROPOSITION 1: Denoting  $d^o$  as the optimal decision rule for the problem described by equation (3), if  $d_{kt}^o = 1$ , then

$$\text{DAI}_k(\gamma_{kt}, \delta_{kt}) = \max_{m \in M} [\text{DAI}_m(\gamma_{mt}, \delta_{mt})].$$

Furthermore, any tie-breaking rule between the maxima is optimal.

PROOF OF PROPOSITION 1: Appealing to lemma 1 and noting that the general component  $\psi_t$  is received regardless of which job  $m \in M$  is chosen, theorem 2 of Gittins and Jones (1974) applies directly.

PROPOSITION 2: If  $\sigma_m = 0$ , then  $\text{DAI}_m(\gamma_m, \delta_m) = \gamma_m + \delta_m D(0, \beta)$ .

PROPOSITION 3: Denote by  $\Phi(\epsilon)$  the standard normal cumulative distribution function. Then for each  $\beta \in (0, 1)$  the real number  $D(0, \beta)$  uniquely solves  $D(0, \beta) = \{1 - \beta \Phi[D(0, \beta)]\}^{-1} \beta \Phi'[D(0, \beta)]$ .

PROOF OF PROPOSITIONS 2 AND 3: From lemma 1,

$$\text{DAI}_m(\gamma_m, \delta_m) = \sup_{\tau \in T} E_0 \left[ (1 - \beta) \sum_{r=0}^{\tau} \beta^r (x_{mr} - \psi_r) + \beta^{\tau+1} \text{DAI}_m(\gamma_m, \delta_m) \right].$$

Supposing  $\sigma_m = 0$ , after one period's experience, the match parameter  $\xi_m$  is perfectly revealed. Therefore if  $\xi_m \geq \text{DAI}_m(\gamma_m, \delta_m)$ , setting  $\tau = \infty$  is optimal. Otherwise  $\tau = 0$  is optimal. Hence

$$\text{DAI}_m(\gamma_m, \delta_m) = (1 - \beta) \gamma_m + \beta E_0 \{ \max[\xi_m, \text{DAI}_m(\gamma_m, \delta_m)] \}.$$

Let  $\Phi(\epsilon)$  denote the standard normal distribution function. Noting that beliefs about  $\xi_m$  are initially distributed  $N(\gamma_m, \delta_m^2)$ :

$$\begin{aligned} \text{DAI}_m(\gamma_m, \delta_m) &= (1 - \beta) \gamma_m + \beta E_0 \{ \max[\gamma_m + \delta_m \epsilon, \text{DAI}_m(\gamma_m, \delta_m)] \} \\ &= \gamma_m + \delta_m \beta E_0 \left\{ \max \left[ \epsilon, \frac{\text{DAI}_m(\gamma_m, \delta_m) - \gamma_m}{\delta_m} \right] \right\}. \end{aligned}$$

Let  $x \equiv \delta_m^{-1} [\text{DAI}_m(\gamma_m, \delta_m) - \gamma_m]$ , subtract  $\gamma_m$  from both sides of the equation, and then divide through by  $\delta_m$  to obtain the recursion



$$\begin{aligned}
x &= \beta E_0[\max(\epsilon, x)] \\
&= \beta \left[ x\Phi(x) + \int_x^\infty \epsilon \Phi'(\epsilon) d\epsilon \right] \\
&= \beta [x\Phi(x) + \Phi'(x)].
\end{aligned}$$

For each  $\beta \in (0, 1)$  consider solutions to the equation above in  $x$ . From the definition of  $x$ ,  $\delta_m^{-1}[\text{DAI}_m(\gamma_m, \delta_m) - \gamma_m]$  solves the recursion for all  $\beta \in (0, 1)$ . Setting  $(\gamma_m, \delta_m) = (0, 1)$ , proposition 3 is established on rearranging terms. Also,  $\partial(\text{LHS})/\partial x = 1$ ;  $\partial(\text{RHS})/\partial x = \beta\Phi(x) < 1$ . Hence  $D(0, \beta)$  is the unique solution to the equation above, implying  $D(0, \beta) = \delta_m^{-1}[\text{DAI}_m(\gamma_m, \delta_m) - \gamma_m]$ , which, on rearrangement, yields proposition 2.

**PROPOSITION 4:**  $\text{DAI}_m(\gamma_{mt}, \delta_{mt}) = \gamma_{mt} + \delta_{mt}D(\alpha_m + \tau_{mt}, \beta)$ .

**PROOF OF PROPOSITION 4:** If  $\sigma_m = 0$ , the proof to propositions 2 and 3 applies. Accordingly, suppose  $\sigma_m > 0$ . From definition 2,

$$\begin{aligned}
\text{DAI}_m(\gamma_{mt}, \delta_{mt}) &\equiv \sup_{\tau \geq t} \left\{ \frac{E_t \left[ \sum_{r=t}^{\tau} \beta^{r-t} (x_{mr} - \psi_r) \right]}{E_t \left( \sum_{r=t}^{\tau} \beta^{r-t} \right)} \right\} \\
&= \gamma_{mt} + \sup_{\tau \geq t} \left\{ \frac{E_t \left[ \sum_{r=t}^{\tau} \beta^{r-t} (x_{mr} - \psi_r - \gamma_{mt}) \right]}{E_t \left( \sum_{r=t}^{\tau} \beta^{r-t} \right)} \right\}
\end{aligned}$$

(adding and subtracting  $\gamma_{mt}$ )

$$= \gamma_{mt} + \sup_{\tau \geq t} \left\{ \frac{E_t \left[ \sum_{r=t}^{\tau} \beta^{r-t} \delta_{mt}^{-1} (x_{mr} - \psi_r - \gamma_{mt}) \right]}{E_t \left( \sum_{r=t}^{\tau} \beta^{r-t} \right)} \right\} \delta_{mt}$$

(multiplying and dividing the bracketed expression by  $\delta_{mt}$ ). But from equations (1) and (2) in the text,  $\delta_{mt}^{-1}(x_{mr} - \psi_r - \gamma_{mt}) = \delta_{mt}^{-1}(\xi_m - \gamma_{mt}) + (\alpha_m + \tau_{mt})^{1/2} \epsilon_{mr}$ . As of time  $t$ , beliefs about the match parameter  $\xi_m$  are normally distributed  $N(\gamma_{mt}, \delta_{mt}^2)$ ; this implies that beliefs about  $\delta_{mt}^{-1}(\xi_m - \gamma_{mt})$  are normally distributed  $N(0, 1)$ . Appealing to definition 4,

$$D(\alpha_m + \tau_{mt}, \beta) = \sup_{\tau \geq t} \left\{ \frac{E_t \left[ \sum_{r=t}^{\tau} \beta^{r-t} \delta_{mt}^{-1} (x_{mr} - \psi_r - \gamma_{mt}) \right]}{E_t \left( \sum_{r=t}^{\tau} \beta^{r-t} \right)} \right\}.$$

Substituting the left-hand side of this equation into the expression derived for  $\text{DAI}_m(\gamma_{mt}, \delta_{mt})$  above, one obtains the result for  $\sigma_m > 0$ .

**PROPOSITION 5:** Let  $F$  denote the space of bounded continuous real-valued functions taking their domain on the space  $(-\infty, \infty) \times (0, \infty)$ . For all  $f \in F$  define  $C$ , a mapping from  $F$  to itself, as

$$C[f(\gamma, \alpha)] = \beta \int_{-\infty}^{\infty} \max(0, (1 - \beta)\{\gamma + \epsilon[\alpha(\alpha + 1)]^{-1/2}\} \\ + f\{\gamma + \epsilon[\alpha(\alpha + 1)]^{-1/2}, \alpha + 1\})d\Phi(\epsilon).$$

There exists a unique function  $g(\gamma, \alpha) \in F$  satisfying  $g(\gamma, \alpha) = C[g(\gamma, \alpha)]$ . Moreover, defining  $C_k(f)$  recursively as  $C[C_{k-1}(f)]$ , for all  $f \in F$ ,  $\|C_k(f) - g\| \leq (1 - \beta)^{-1}\|C_k(f) - C_{k-1}(f)\|$ . Finally, for all  $(\alpha, \beta) \in (0, \infty) \times (0, 1)$ ,  $D(\alpha, \beta) = (1 - \beta)^{-1}\alpha^{1/2}g[-\alpha^{-1/2}D(\alpha, \beta), \alpha]$ .

PROOF OF PROPOSITION 5: Note that  $F$  is complete and therefore, by Blackwell's (1965) sufficiency conditions,  $C$  is a contraction. Hence the contraction mapping theorem applies, simultaneously establishing the existence of a unique fixed point  $g(\gamma, \alpha)$  and the inequality stated in the proposition (see Smart [1974] for details).

Now suppose  $\sigma_m = 1$ , which implies  $\delta_m = \alpha_m^{-1/2}$ . Then, for all  $y \in (-\infty, \infty)$ , define

$$V(y, \gamma_m, \alpha_m) \equiv \sup_{\tau \in T} E_0 \left[ (1 - \beta) \sum_{r=0}^{\tau} \beta^r (x_{mr} - \psi_r) + \beta^{\tau+1} y \right]$$

$$g_1(\gamma_m, \alpha_m) \equiv \beta E_0 \max[0, V(0, \gamma_{m1}, \alpha_m + 1)].$$

One can establish  $g(\gamma, \alpha) \equiv g_1(\gamma, \alpha)$ . To prove this claim, observe that

$$C[g_1(\gamma_m, \alpha_m)] = \beta \int_{-\infty}^{\infty} \max(0, (1 - \beta)\{\gamma + \epsilon[\alpha_m(\alpha_m + 1)]^{-1/2}\} \\ + g_1\{\gamma_m + \epsilon[\alpha_m(\alpha_m + 1)]^{-1/2}, \alpha_m + 1\})d\Phi(\epsilon)$$

(from the definition of  $C$ )

$$= \beta E_0 \max[0, (1 - \beta)\gamma_{m1} + g_1(\gamma_{m1}, \alpha_m + 1)]$$

(using eq. [2])

$$= \beta E_0 \max\{0, (1 - \beta)\gamma_{m1} \\ + \beta E_1 \max[0, V(0, \gamma_{m2}, \alpha_m + 2)]\}$$

(from the definition of  $g_1$ )

$$= \beta E_0 \max[0, V(0, \gamma_{m1}, \alpha_m + 1)]$$

(appealing to Bellman's [1957] principle of optimality)

$$\equiv g_1(\gamma_m, \alpha_m).$$

Thus  $g_1$  is a fixed point for  $C$  on  $F$ . So from the uniqueness of  $g$ , the claimed identity between  $g_1$  and  $g$  follows immediately. Hence, appealing to Bellman's principle once more,  $V(0, \gamma_m, \alpha_m) = (1 - \beta)\gamma_m + g(\gamma_m, \alpha_m)$ . Also, since  $\sigma_m = 1$ , the definition of  $V(y, \gamma_m, \alpha_m)$  implies by lemma 1 that

$$V[\text{DAI}_m(\gamma_m, \alpha_m^{-1/2}), \gamma_m, \alpha_m] = \text{DAI}_m(\gamma_m, \alpha_m^{-1/2}) \\ = \gamma_m + \alpha_m^{-1/2}D(\alpha_m, \beta),$$

the second line following from proposition 4. Setting  $\gamma_m = -\alpha_m^{-1/2}D(\alpha_m, \beta)$ , then,

$$\begin{aligned} 0 &= \text{DAI}_m(\gamma_m, \alpha_m^{-1/2}) \\ &= V[0, -\alpha_m^{-1/2}D(\alpha_m, \beta), \alpha_m^{-1/2}] \\ &= -(1 - \beta)\alpha_m^{-1/2}D(\alpha_m, \beta) + g[-\alpha_m^{-1/2}D(\alpha_m, \beta), \alpha_m]. \end{aligned}$$

PROPOSITION 6: For all  $(m, \tau) \in M \times T$

$$\begin{aligned} h_{m\tau} &= \psi_{m\tau}[\alpha_m^{-1/2}D(\alpha_m, \beta) - (\alpha_m + \tau)^{-1/2}D(\alpha_m + \tau, \beta)] \\ \psi_{m1}(\rho) &= \Phi[\rho\alpha_m^{1/2}(\alpha_m + 1)^{1/2}] \\ \psi_{m(\tau+1)}(\rho) &= (1 - h_{m\tau})^{-1} \left\{ \int_{-\infty}^{\infty} \psi_{m\tau}[\rho - \epsilon(\alpha_m + \tau)^{1/2}(\alpha_m + \tau + 1)^{1/2}]d\Phi(\epsilon) - h_{m\tau} \right\}. \end{aligned}$$

PROOF OF PROPOSITION 6: From definition 6,

$$\begin{aligned} h_{m\tau} &= P\{\gamma_{mt} + \delta_{mt}D(\alpha_m + \tau, \beta) < \gamma_m + \delta_mD(\alpha_m, \beta) | \tau_{mt} = \tau\} \\ &= P\{\gamma_{mt} + \sigma_m(\alpha_m + \tau)^{-1/2}D(\alpha_m + \tau, \beta) < \gamma_m + \sigma_m\alpha_m^{-1/2}D(\alpha_m, \beta) | \tau_{mt} = \tau\} \\ &\text{(noting from [2] that if } \tau_{mt} = \tau \text{ then } \delta_{mt} = \sigma_m[\alpha_m + \tau]^{-1/2}) \\ &= P\{\gamma_{mt} < \gamma_m + \sigma_m[\alpha_m^{-1/2}D(\alpha_m, \beta) - (\alpha_m + \tau)^{-1/2}D(\alpha_m + \tau, \beta)] | \tau_{mt} = \tau\} \\ &\text{(subtracting } \sigma_m[\alpha_m + \tau]^{-1/2}D[\alpha_m + \tau, \beta] \text{ from both sides of the inequality)} \\ &= \psi_{m\tau}[\alpha_m^{-1/2}D(\alpha_m, \beta) - (\alpha_m + \tau)^{-1/2}D(\alpha_m + \tau, \beta)] \end{aligned}$$

(from def. 7). This proves the first equality in proposition 6. Also from definition 7,

$$\begin{aligned} \psi_{m1}(\rho) &= P\{\gamma_{mt} < \gamma_m + \rho\sigma_m | \tau_{mt} = 1\} \\ &= P\{\gamma_{m1} < \gamma_m + \rho\sigma_m | \tau_{m1} = 1\} \end{aligned}$$

(since from eq. [1] net returns  $[x_{mt} - \psi_t]$  are identically and independently distributed over time)

$$= P\{[\delta_m^{-2}\gamma_m + \sigma_m^{-2}(x_{m1} - \psi_1)](\delta_m^{-2} + \sigma_m^{-2})^{-1} < \gamma_m + \rho\sigma_m\}$$

(from eq. [2])

$$= P\{[\alpha_m\gamma_m + (x_{m1} - \psi_1)](\alpha_m + 1)^{-1} < \gamma_m + \rho\sigma_m\}$$

(multiplying the numerator and denominator of the left-hand side by  $\sigma_m^{-2}$  and noting  $\alpha_m = \delta_m^{-2}\sigma_m^2$ )

$$= P\{\sigma_m^{-1}(x_{m1} - \psi_1 - \gamma_m) < \rho(1 + \alpha_m)\}$$

(subtracting  $\gamma_m$  from both sides of the inequality and then multiplying through by  $\sigma_m^{-1}[1 + \alpha_m]$ ). But from (1),  $\sigma_m^{-1}(x_{m1} - \psi_1 - \gamma_m) = \sigma_m^{-1}(\xi_m - \gamma_m) + \epsilon_{m1}$ . Beliefs about the first expression on the right-hand side form the distribution  $N(0, \alpha_m^{-1})$ , while the second is independently distributed  $N(0, 1)$ . Hence beliefs about the left-hand side are distributed  $N(0, 1 + \alpha_m^{-1})$ . Therefore, supposing  $\epsilon$  is a standard normal random variate,

$$\begin{aligned} \psi_{m1} &= P\{\sigma_m^{-1}(x_{m1} - \psi_1 - \gamma_m) < \rho(1 + \alpha_m)\} \\ &= P\{(1 + \alpha_m^{-1})^{1/2}\epsilon < \rho(1 + \alpha_m)\} \\ &= P\{\epsilon < \rho(1 + \alpha_m)^{1/2}\alpha_m^{1/2}\} \\ &\equiv \Phi[\rho\alpha_m^{1/2}(\alpha_m + 1)^{1/2}]. \end{aligned}$$

This proves the second equality in proposition 6. Finally if  $d_{mt} = 1$  for all  $t \leq \tau$ , then from (2)

$$\begin{aligned}\gamma_{m(\tau+1)} &= [(\alpha_m + \tau)\gamma_{m\tau} + (x_{m\tau} - \psi_\tau)](\alpha_m + \tau + 1)^{-1} \\ &= (\alpha_m + \tau)(\alpha_m + \tau + 1)^{-1}\gamma_{m\tau} + (\alpha_m + \tau + 1)^{-1}(x_{m\tau} - \psi_\tau) \\ &= (\alpha_m + \tau)(\alpha_m + \tau + 1)^{-1}\gamma_{m\tau} + (\alpha_m + \tau + 1)^{-1}(\xi_m + \sigma_m \epsilon_{m\tau}).\end{aligned}$$

Since beliefs about  $\xi_m$  are distributed  $N(\gamma_{m\tau}, \delta_{m\tau}^2)$  when  $t_{mt} = \tau$ , and  $\epsilon_{m\tau}$  is distributed  $N(0, 1)$ ,  $\xi_m + \sigma_m \epsilon_{m\tau} = \gamma_{m\tau} + (\delta_{m\tau}^2 + \sigma_m^2)^{1/2} \epsilon$ , where  $\epsilon$  is distributed  $N(0, 1)$ . Hence

$$\begin{aligned}\gamma_{m(\tau+1)} &= (\alpha_m + \tau)(\alpha_m + \tau + 1)^{-1}\gamma_{m\tau} + (\alpha_m + \tau + 1)^{-1}[\gamma_{m\tau} + (\delta_{m\tau}^2 + \sigma_m^2)^{1/2} \epsilon] \\ &= \gamma_{m\tau} + (\alpha_m + \tau + 1)^{-1}(\delta_{m\tau}^2 + \sigma_m^2)^{1/2} \epsilon \\ &= \gamma_{m\tau} + \sigma_m(\alpha_m + \tau + 1)^{-1/2}(\alpha_m + \tau)^{-1/2} \epsilon,\end{aligned}$$

since  $(\alpha_m + \tau) = \delta_{m\tau}^{-2} \sigma_m^2$  if  $\tau_{mt} = \tau$ . Appealing to definition 7 and substituting for  $\gamma_{m(\tau+1)}$ ,

$$\begin{aligned}\psi_{m(\tau+1)}(\rho) &= P\{\gamma_{m(\tau+1)} < \gamma_m + \sigma_m \rho | \tau_{m(\tau+1)} = \tau + 1\} \\ &= P\{\gamma_{m\tau} < \gamma_m + \sigma_m[\rho - (\alpha_m + \tau + 1)^{-1/2} \\ &\quad \times (\alpha_m + \tau)^{-1/2} \epsilon] | \tau_{m\tau} = \tau, d_{m\tau}^o = 1\}.\end{aligned}$$

Using definition 6, observe that

$$\begin{aligned}P\{d_{m\tau}^o = 1 | \tau_{m\tau} = \tau\} &= 1 - P\{d_{m\tau}^o = 0 | \tau_{m\tau} = \tau\} \\ &= 1 - P\{\gamma_{m\tau} + \delta_{m\tau} D(\alpha_m + \tau, \beta) < \gamma_m \\ &\quad + \delta_m D(\alpha_m, \beta) | \tau_{m\tau} = \tau\} \\ &= 1 - h_{m\tau}.\end{aligned}$$

Combining the last equation with the expression obtained for  $\psi_{m(\tau+1)}(\rho)$ , one obtains

$$\begin{aligned}(1 - h_{m\tau})\psi_{m(\tau+1)}(\rho) &= P\{\gamma_{m\tau} < \gamma_m + \sigma_m[\rho - (\alpha_m + \tau + 1)^{-1/2}(\alpha_m + \tau)^{-1/2} \epsilon] | \tau_{m\tau} \\ &\quad = \tau, d_{m\tau}^o = 1\} P\{d_{m\tau}^o = 1 | \tau_{m\tau} = \tau\} \\ &= P\{\gamma_{m\tau} < \gamma_m + \sigma_m[\rho - (\alpha_m + \tau + 1)^{-1/2}(\alpha_m + \tau)^{-1/2} \epsilon], d_{m\tau}^o = 1 | \tau_{m\tau} = \tau\} \\ &= P\{\delta_m D(\alpha_m, \beta) - \delta_{m\tau} D(\alpha_m + \tau, \beta) < (\gamma_{m\tau} - \gamma_m) \\ &\quad < \sigma_m[\rho - (\alpha_m + \tau + 1)^{-1/2}(\alpha_m + \tau)^{-1/2} \epsilon] | \tau_{m\tau} = \tau\}\end{aligned}$$

(applying the DAI rule, given there exists a new job in the same occupation as  $m$ )

$$\begin{aligned}&= P\{(\gamma_{m\tau} - \gamma_m) < \sigma_m[\rho - (\alpha_m + \tau + 1)^{-1/2}(\alpha_m + \tau)^{-1/2} \epsilon] | \tau_{m\tau} = \tau\} \\ &\quad - P\{\delta_m D(\alpha_m, \beta) - \delta_{m\tau} D(\alpha_m + \tau, \beta) < \gamma_{m\tau} - \gamma_m | \tau_{m\tau} = \tau\} \\ &= P\{\gamma_{m\tau} < \gamma_m + \sigma_m[\rho - (\alpha_m + \tau + 1)^{-1/2}(\alpha_m + \tau)^{-1/2} \epsilon] | \tau_{m\tau} = \tau\} - h_{m\tau} \\ &\text{(using def. 6)}\end{aligned}$$

$$\begin{aligned}
 &= \int_{-\infty}^{\infty} P\{\gamma_{m\tau} < \gamma_m + \sigma_m[\rho - (\alpha_m + \tau + 1)^{-1/2}(\alpha_m + \tau)^{-1/2}\epsilon]\tau_{m\tau} \\
 &\quad = \tau, \epsilon\} d\Phi(\epsilon) - h_{m\tau} \\
 &= \int_{-\infty}^{\infty} \Psi_{m\tau}[\rho - (\alpha_m + \tau + 1)^{-1/2}(\alpha_m + \tau)^{-1/2}\epsilon] d\Phi(\epsilon) - h_{m\tau}.
 \end{aligned}$$

The last line follows from definition 7. Dividing through by  $(1 - h_{m\tau})$  the third equation in proposition 6 is obtained, thus completing the proof.

## Appendix B

The procedure below describes how to compute the standard index. Then the properties of the resulting approximation are discussed.

i) Given  $(\alpha_j, \beta_k) \in [0, \infty) \times (0, 1)$ , recursively define the functions  $f_h(\gamma, \alpha_j - h)$ , which map  $[-\infty, \infty)$  into itself, as follows. First,  $f_0(\gamma, \alpha_j) = 0$ . Second,  $f_{h+1}(\gamma, \alpha_j - h - 1)$  is the cubic spline that knots together a finite number of coordinate pairs  $[\gamma_i, f_{h+1}(\gamma_i, \alpha_j - h - 1)]$ , the abscissa  $\gamma_i$  being drawn from a finite set  $\{\gamma_i\}_{i \in I}$  and the ordinate  $f_{h+1}(\gamma_i, \alpha_j - h - 1)$  being determined from  $f_{h+1}(\gamma_i, \alpha_j - h - 1) = C[f_h(\gamma_i, \alpha_j - h)]$ . (A cubic spline is a piecewise cubic polynomial that joins coordinate pairs so that the resulting function is twice differentiable everywhere.)

ii) Given a finite set  $\{\alpha_j\}_{j \in J}$ , let  $\{\alpha_k\}_{k \in K}$  be the linear transformation  $\{\alpha_k\}_{k \in K} \equiv \{\alpha_j + 1\}_{j \in J}$ .

iii) Observing from ii that  $f_{h+1}(\gamma_i, \alpha_k - h - 1) = f_{h+1}(\gamma_i, \alpha_j - h)$ , suppose that  $\nu$  is the lowest counting number that, for all  $(\gamma, \alpha) \in \{\gamma_i\}_{i \in I} \times \{\alpha_j\}_{j \in J}$ , satisfies  $|f_{\nu+1}(\gamma, \alpha_j - \nu) - f_\nu(\gamma, \alpha_j - \nu)| \leq 0.001(1 - \beta_k)$ .

iv) Finally, let  $D_{jk}$  denote the (unique) solution to the equation  $D_{jk} = (1 - \beta_k)^{-1}(\alpha_j - \nu)^{1/2}f_\nu[(\alpha_j - \nu)^{-1/2}D_{jk}, \alpha_j - \nu]$ . The number  $D(\alpha_j - \nu, \beta_k)$  was accordingly approximated by  $D_{jk}$ . The theoretical justification for this approximation comes from proposition 5: if  $\|C[f_\nu(\gamma, \alpha)] - f_\nu(\gamma, \alpha)\| \leq 0.001(1 - \beta)$ , then  $\|g(\gamma, \alpha) - f_\nu(\gamma, \alpha)\| \leq 0.001$ . Upon substituting  $f_\nu(\gamma, \alpha)$  for  $g(\gamma, \alpha)$  in the equation  $D(\alpha - \nu, \beta) = (1 - \beta)^{-1}(\alpha - \nu)^{1/2}g[(\alpha - \nu)^{-1/2}D(\alpha - \nu, \beta), \alpha - \nu]$ , the approximation proposed above is obtained.

This algorithm is subject to two main sources of error. First, the sup norm  $\|\cdot\|$  is applied only to the finite set  $\{\gamma_i\}_{i \in I} \times \{\alpha_j\}_{j \in J}$  instead of the (much larger) set  $[-\infty, \infty) \times [0, \infty]$ . Second, the mapping  $f_\nu(\gamma, \alpha)$  is used in the final step to determine  $D(\alpha, \beta_k)$ , rather than  $g(\gamma, \alpha)$ . In practice both of those errors could, apparently, be cheaply controlled for at least moderate values of  $\beta$ , say less than 0.95. At high levels of  $\beta$ , the number of iterations  $\nu$  must be greatly increased to satisfy the convergence criterion.

Table B1 reports computed values of  $D(\alpha, \beta)$  for 230 coordinate pairs  $(\alpha, \beta)$ . The numbers down the left side of the table indicate values of  $\alpha_m \in [0.1, \dots, 17.1]$ , while those across the top refer to values of  $\beta \in [0.1, \dots, 0.95]$ .

## Appendix C

The following procedure describes how to compute the sequence of hazards  $\{h_t(\alpha, \beta)\}_{t=1}^5$  for different values of  $\alpha$  and  $\beta$ .

i) For each  $\beta_k$  compute a cubic spline for  $D(\alpha, \beta_k)$  by knotting together the points determined in Appendix B.

ii) Appealing to the second equation of proposition 6, compute  $\Psi_{m1}(\rho)$  directly. Hence,  $h_{m1}$  is determined from the first equation in proposition 6 by

TABLE B1  
 $D(\alpha_m, \beta)$

	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95
.0	.0420	.0890	.1422	.2036	.2760	.3644	.4777	.6360	.9015	1.1589
.1	.0396	.0840	.1344	.1926	.2614	.3457	.4538	.6049	.8562	1.0925
.3	.0365	.0775	.1243	.1787	.2433	.3229	.4260	.5712	.8149	1.0458
.7	.0319	.0681	.1096	.1582	.2165	.2890	.3839	.5193	.7512	.9722
1.1	.0287	.0615	.0992	.1436	.1972	.2642	.3525	.4809	.7021	.9144
1.3	.0275	.0588	.0950	.1377	.1893	.2540	.3402	.4648	.6814	.8898
1.7	.0253	.0543	.0879	.1277	.1760	.2368	.3185	.4371	.6455	.8468
2.1	.0236	.0508	.0823	.1197	.1652	.2231	.3005	.4143	.6144	.8085
3.1	.0205	.0442	.0718	.1047	.1450	.1968	.2664	.3704	.5539	.7323
4.1	.0183	.0396	.0644	.0941	.1306	.1780	.2420	.3370	.5077	.6748
5.1	.0167	.0361	.0589	.0862	.1198	.1638	.2236	.3122	.4702	.6269
6.1	.0155	.0334	.0545	.0799	.1113	.1527	.2090	.2926	.4421	.5856
7.1	.0144	.0312	.0509	.0747	.1043	.1434	.1970	.2766	.4189	.5540
8.1	.0136	.0293	.0479	.0703	.0983	.1354	.1866	.2627	.4004	.5282
9.1	.0128	.0278	.0453	.0667	.0931	.1283	.1771	.2500	.3860	.5075
10.1	.0122	.0260	.0430	.0634	.0885	.1220	.1686	.2385	.3719	.4900
11.1	.0116	.0248	.0410	.0605	.0844	.1162	.1606	.2296	.3582	.4736
12.1	.0111	.0237	.0394	.0578	.0806	.1108	.1534	.2202	.3458	.4572
13.1	.0107	.0227	.0377	.0553	.0770	.1066	.1476	.2113	.3336	.4424
14.1	.0103	.0218	.0363	.0532	.0742	.1023	.1415	.2028	.3214	.4277
15.1	.0099	.0210	.0350	.0512	.0713	.0969	.1358	.1947	.3092	.4123
16.1	.0096	.0203	.0338	.0494	.0688	.0932	.1304	.1869	.2969	.3960
17.1	.0093	.0196	.0327	.0478	.0664	.0900	.1257	.1795	.2847	.3791

using step i to evaluate  $\psi_{m1}(\rho)$  at the number  $[\alpha^{1/2}D(\alpha, \beta_k) - (\alpha + 1)^{1/2}D(\alpha + 1, \beta_k)]$ .

iii) Substitute  $\psi_{m1}(\rho)$  and  $h_{m1}$  into the third equation of proposition 6. Then compute  $\psi_{m2}(\rho)$ . Using step i again, evaluate this distribution at the number  $[\alpha^{1/2}D(\alpha, \beta_k) - (\alpha + 2)^{1/2}D(\alpha + \tau, \beta_k)]$  to determine  $h_{m2}$ . Continuing in this fashion,  $h_{m3}$ ,  $h_{m4}$ , and  $h_{m5}$  are successively determined.

Table C1 reports, for 230 coordinate pairs  $(\alpha_m, \beta)$ , the computed sequences  $\{h_{m\tau}(\alpha_m, \beta)\}_{\tau=1}^5$ . The numbers down the left side of the table indicate values of  $\alpha_m \in [0.1, \dots, 17.1]$ , while those across the top refer to values of  $\beta \in [0.1, \dots, 0.95]$ .

## Appendix D

Not all the 822 life histories in the subsample of whites were adequately documented, and a number of conditions were imposed that individual data were required to satisfy for inclusion in the analysis.

i) Information about full-time activities had to be continuously recorded from the time of entry into the labor force through to January 1969. Full-time activities comprised full-time education, full-time employment, and full-time other, all of which were explicitly coded (loss 200).

ii) No overlaps in full-time activities were permitted. Overlaps can occur because of miscoded spell dates, multiple full-time jobs being held simultaneously, or, in some cases, the classification of seasonal work (brought about, e.g., by a schoolteacher's taking a summer job). It was difficult to distinguish these three cases without access to the questionnaires, and so respondents with any overlapping full-time employment spells were dropped (loss 141).

iii) Respondents with missing data on spell dates and relevant variable codes were dropped (loss 5), as were those whose records claimed a spell finished before it started (loss 18). With respect to conditions i–iii, for full-time spells, spell beginning and spell ending dates are available in two forms, a continuous numeric code and a month/year code. When cases failed to satisfy all of the conditions above using the numeric codes, an attempt was made to salvage them using the year/month codes. This succeeded in 11 cases, so, after netting out two who pursued military careers throughout their entire work history, this left the 467 white men that form the basis for the present study.

## References

- Bartel, Ann P. "The Migration Decision: What Roles Does Job Mobility Play?" *A.E.R.* 69 (December 1979): 775–86.
- Bartel, Ann P., and Borjas, George J. "Wage Growth and Job Turnover: An Empirical Analysis." In *Studies in Labor Markets*, edited by Sherwin Rosen. Chicago: Univ. Chicago Press (for N.B.E.R.), 1981.
- Becker, Gary S. "Investment in Human Capital: A Theoretical Analysis." *J.P.E.* 70, no. 5, pt. 2 (October 1962): 9–49.
- . *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. 2d ed. Chicago: Univ. Chicago Press (for N.B.E.R.), 1975.
- Bellman, Richard E. *Dynamic Programming*. Princeton, N.J.: Princeton Univ. Press, 1957.
- Blackwell, David. "Discounted Dynamic Programming." *Annals Math. Statis.* 36 (February 1965): 226–35.
- Borjas, George J., and Rosen, Sherwin. "Income Prospects and Mobility of

TABLE C1

 $\{h_{m\tau}(\alpha_m, \beta)\}_{\tau=1}^5$ 

	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95
.1	.513 .071 .029 .015 .010 .511 .114	.527 .075 .030 .016 .010 .522 .119	.544 .079 .032 .017 .011 .536 .124	.562 .084 .034 .019 .012 .551 .130	.584 .089 .037 .020 .013 .569 .137	.610 .097 .040 .022 .014 .591 .147	.642 .107 .045 .025 .016 .618 .160	.685 .123 .053 .030 .019 .654 .179	.750 .151 .067 .039 .025 .711 .213	.804 .181 .083 .048 .032 .760 .248
.3	.050 .028 .018 .508 .138	.052 .029 .018 .516 .142	.054 .030 .019 .526 .147	.057 .032 .021 .538 .153	.061 .035 .022 .551 .161	.066 .038 .024 .567 .170	.073 .042 .027 .588 .182	.083 .048 .031 .616 .201	.103 .061 .040 .661 .234	.123 .074 .050 .701 .268
.5	.063 .036 .023 .508 .154	.065 .037 .024 .517 .159	.068 .039 .025 .528 .164	.071 .041 .027 .540 .170	.075 .044 .028 .554 .178	.080 .047 .031 .571 .188	.088 .051 .034 .593 .201	.098 .058 .039 .622 .220	.118 .072 .048 .670 .255	.139 .086 .059 .712 .290
.7	.073 .043 .028 .508 .166	.076 .044 .029 .517 .170	.079 .046 .030 .527 .176	.082 .048 .032 .538 .182	.087 .051 .034 .552 .190	.092 .055 .036 .569 .200	.100 .060 .040 .590 .213	.112 .068 .045 .619 .232	.133 .082 .056 .666 .267	.156 .098 .068 .708 .301
.9	.081 .048 .032 .507 .175	.084 .050 .033 .514 .179	.087 .052 .034 .523 .184	.091 .054 .036 .533 .190	.095 .057 .038 .545 .198	.101 .061 .041 .559 .207	.109 .066 .044 .578 .219	.121 .074 .050 .603 .237	.143 .090 .061 .645 .270	.166 .106 .074 .683 .303
1.1	.088 .053 .035	.090 .054 .036	.093 .056 .038	.097 .059 .039	.102 .062 .041	.107 .066 .044	.115 .071 .048	.127 .079 .054	.149 .094 .065	.171 .111 .078



1.5	.506	.512	.520	.529	.539	.551	.568	.591	.630	.665
	.188	.192	.197	.202	.209	.218	.230	.247	.277	.308
	.098	.101	.104	.107	.112	.117	.125	.136	.158	.179
	.060	.062	.064	.067	.069	.073	.079	.087	.102	.119
	.041	.042	.044	.045	.047	.050	.054	.060	.071	.085
2.1	.505	.510	.516	.524	.532	.543	.557	.576	.610	.641
	.201	.204	.209	.214	.220	.228	.238	.254	.281	.309
	.109	.112	.115	.118	.122	.127	.134	.145	.166	.186
	.069	.071	.073	.075	.078	.082	.087	.094	.109	.126
	.048	.049	.050	.052	.054	.057	.061	.067	.078	.091
3.1	.504	.508	.513	.518	.525	.534	.545	.562	.590	.616
	.213	.216	.220	.224	.229	.236	.245	.258	.283	.307
	.122	.124	.126	.129	.133	.137	.144	.153	.171	.192
	.079	.081	.083	.085	.087	.091	.096	.103	.116	.131
	.056	.057	.059	.060	.062	.065	.068	.074	.084	.096
4.1	.503	.506	.510	.515	.521	.528	.537	.551	.577	.600
	.221	.224	.227	.230	.235	.240	.248	.259	.280	.304
	.129	.131	.134	.136	.139	.144	.149	.158	.173	.191
	.086	.088	.089	.091	.094	.097	.101	.108	.119	.133
	.062	.063	.064	.066	.068	.070	.074	.079	.088	.099
5.1	.503	.505	.509	.513	.518	.524	.532	.544	.565	.589
	.226	.228	.231	.234	.238	.243	.250	.260	.278	.298
	.135	.137	.139	.141	.144	.148	.153	.161	.174	.190
	.091	.093	.094	.096	.098	.101	.105	.111	.121	.133
	.066	.067	.069	.070	.072	.074	.078	.083	.091	.100
6.1	.502	.505	.508	.511	.516	.521	.528	.538	.557	.576
	.230	.232	.235	.237	.241	.245	.251	.260	.275	.293
	.139	.141	.143	.145	.147	.151	.156	.163	.173	.187
	.095	.096	.098	.099	.102	.105	.108	.114	.123	.133
	.070	.071	.072	.073	.075	.077	.081	.085	.093	.101
7.1	.502	.504	.507	.510	.514	.519	.525	.535	.550	.567
	.233	.235	.237	.240	.243	.248	.253	.262	.273	.289
	.143	.144	.146	.148	.150	.154	.158	.165	.175	.186
	.098	.099	.101	.102	.104	.107	.111	.116	.125	.134
	.073	.074	.075	.076	.078	.080	.083	.087	.095	.103

$$\{h_{m\ell}(\alpha_m, \beta)\}_{\tau=1}^5$$

	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95
8.1	.502	.504	.506	.509	.513	.517	.523	.532	.543	.559
	.237	.239	.240	.243	.246	.250	.255	.263	.274	.286
	.146	.147	.149	.150	.153	.156	.161	.166	.177	.186
	.101	.102	.103	.105	.107	.109	.113	.118	.126	.135
	.075	.076	.077	.078	.080	.082	.085	.089	.097	.104
9.1	.502	.504	.506	.508	.512	.516	.522	.530	.542	.554
	.240	.241	.243	.245	.248	.252	.257	.263	.275	.286
	.149	.150	.153	.153	.156	.159	.163	.168	.178	.188
	.103	.104	.106	.107	.109	.111	.115	.120	.128	.136
	.077	.078	.079	.080	.082	.084	.087	.091	.098	.105
10.1	.501	.503	.505	.508	.511	.515	.520	.526	.540	.551
	.243	.244	.246	.248	.250	.254	.259	.265	.276	.287
	.152	.153	.155	.156	.159	.161	.165	.171	.181	.190
	.106	.107	.108	.109	.111	.114	.117	.122	.131	.138
	.079	.080	.081	.082	.083	.086	.088	.093	.100	.107
11.1	.501	.503	.505	.507	.510	.514	.519	.526	.538	.549
	.245	.247	.248	.250	.253	.255	.259	.267	.277	.287
	.156	.157	.158	.160	.162	.164	.168	.174	.184	.192
	.109	.110	.111	.112	.114	.117	.119	.124	.133	.141
	.081	.082	.083	.084	.085	.088	.090	.095	.103	.110
12.1	.501	.503	.505	.507	.509	.512	.517	.525	.536	.546
	.247	.249	.250	.252	.254	.257	.261	.268	.278	.287
	.160	.161	.162	.163	.165	.169	.172	.177	.188	.196
	.112	.113	.114	.115	.117	.119	.122	.127	.136	.145
	.084	.084	.085	.086	.088	.090	.092	.097	.105	.113

13.1	.501	.503	.504	.506	.508	.512	.517	.524	.536	.545
	.249	.250	.251	.253	.255	.259	.262	.268	.279	.288
	.164	.165	.166	.167	.169	.172	.175	.181	.192	.201
	.115	.116	.117	.118	.120	.122	.125	.130	.140	.149
	.087	.088	.089	.090	.091	.093	.096	.100	.109	.118
14.1	.501	.502	.504	.506	.508	.513	.516	.523	.535	.545
	.250	.251	.252	.253	.255	.258	.262	.268	.279	.289
	.168	.169	.170	.171	.173	.175	.179	.184	.195	.206
	.120	.121	.121	.123	.124	.126	.129	.134	.144	.154
	.090	.091	.092	.093	.094	.096	.098	.102	.112	.121
15.1	.501	.502	.504	.505	.508	.511	.515	.522	.535	.546
	.250	.251	.252	.253	.255	.257	.261	.267	.278	.289
	.173	.174	.175	.176	.178	.180	.183	.189	.199	.211
	.124	.125	.125	.127	.128	.130	.133	.137	.148	.159
	.093	.094	.095	.096	.097	.098	.101	.105	.114	.125
16.1	.501	.502	.503	.505	.507	.510	.514	.521	.534	.546
	.248	.249	.250	.251	.253	.255	.258	.264	.275	.287
	.176	.177	.178	.179	.181	.183	.186	.190	.202	.214
	.128	.129	.130	.131	.132	.134	.136	.141	.151	.163
	.096	.097	.098	.099	.100	.101	.104	.108	.115	.128
17.1	.501	.502	.503	.505	.507	.509	.513	.520	.533	.547
	.246	.247	.248	.249	.251	.253	.256	.260	.272	.284
	.179	.179	.181	.182	.183	.185	.188	.192	.203	.215
	.132	.132	.133	.134	.136	.137	.140	.144	.152	.166
	.100	.100	.101	.102	.103	.104	.107	.110	.118	.127

- Younger Men." Report no. 7920. Chicago: Univ. Chicago, Center Math. Studies Bus. and Econ., May 1979.
- Gittins, J. C. "Bandit Processes and Dynamic Allocation Indices." *J. Royal Statis. Soc.*, ser. B, 41, no. 2 (1979): 148-77.
- Gittins, J. C., and Jones, D. M. "A Dynamic Allocation Index for the Sequential Design of Experiments." In *Progress in Statistics*, edited by Joseph M. Gani. Amsterdam: North-Holland, 1974.
- Heckman, James J., and Borjas, George J. "Does Unemployment Cause Future Unemployment? Definitions, Questions and Answers from a Continuous Time Model of Heterogeneity and State Dependence." *Economica* 47 (August 1980): 247-83.
- Heckman, James J., and Singer, Burton. "The Identification Problem in Econometric Models for Duration Data." In *Advances in Econometrics*, edited by Werner Hildenbrand. London: Cambridge Univ. Press, 1982.
- . "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." *Econometrica* 52 (March 1984): 271-320.
- Jovanovic, Boyan. "Job Matching and the Theory of Turnover." *J.P.E.* 87, no. 5, pt. 1 (October 1979): 972-90.
- Laird, Nan. "Nonparametric Maximum Likelihood Estimation of a Mixing Distribution." *J. American Statis. Assoc.* 73 (December 1978): 805-11.
- Lippman, Steven A., and McCall, John J. "The Economics of Job Search: A Survey: Part I." *Econ. Inquiry* 14 (June 1976): 155-89.
- Marshall, Alfred. *Principles of Economics*. 8th ed. New York: Macmillan, 1948.
- Miller, Robert A. "Job Specific Capital and Labor Mobility." Ph.D. dissertation, Univ. Chicago, June 1982.
- Mincer, Jacob. "On-the-Job Training: Costs, Returns, and Some Implications." *J.P.E.* 70, no. 5, pt. 2 (October 1962): 50-79.
- Mincer, Jacob, and Jovanovic, Boyan. "Labor Mobility and Wages." In *Studies in Labor Markets*, edited by Sherwin Rosen. Chicago: Univ. Chicago Press (for N.B.E.R.), 1981.
- Oi, Walter Y. "Labor as a Quasi-fixed Factor." *J.P.E.* 70, no. 5, pt. 2 (October 1962): 538-55.
- Ornstein, Michael D. *Entry into the American Labor Force*. New York: Academic Press, 1976.
- Smart, D. R. *Fixed Point Theorems*. London: Cambridge Univ. Press, 1974.
- Smith, Adam. *An Inquiry into the Nature and Causes of the Wealth of Nations*. 1776. Reprint ed. New York: Random House, 1937.
- Viscusi, W. Kip. "Job Hazards and Worker Quit Rates: An Analysis of Adaptive Worker Behavior." *Internat. Econ. Rev.* 20 (February 1979): 29-58.
- Weitzman, Martin L. "Optimal Search for the Best Alternative." *Econometrica* 47 (May 1979): 641-54.
- Zellner, Arnold. *An Introduction to Bayesian Inference in Econometrics*. New York: Wiley, 1971.