ECONOMIC RESEARCH REPORTS

A FOUNDATION FOR THREE POPULAR ASSUMPTIONS IN JOB-MATCHING MODELS

by

John Dagsvik

Boyan Jovanovic

and

Andrea Shephard

R.R. #85-13 April 1985

C. V. STARR CENTER
FOR APPLIED ECONOMICS

NEW YORK UNIVERSITY
FACULTY OF ARTS AND SCIENCE
DEPARTMENT OF ECONOMICS
WASHINGTON SQUARE
NEW YORK, N.Y. 10003

A FOUNDATION FOR THREE POPULAR ASSUMPTIONS IN JOB-MATCHING MODELS*

by

John Dagsvik

Central Bureau of Statistics, Norway

Boyan Jovanovic

New York University

and

Andrea Shepard

AT&T Bell Laboratories and

Columbia University

November 1984

*The authors are solely responsible for the views expressed in this paper. Technical assistance provided by the C. V. Starr Center for Applied Economics.

ABSTRACT

Matching models usually assume an exogeneously given distribution of match-productivity, and the act of changing jobs then has the worker taking a new, independent sample from this distribution. Using a "characteristics" approach to matching two heterogeneous populations, this paper shows that assumptions concerning the normality and serial independence of match-productivity (across successive matches) follow from some simple axioms. Moreover, the normality assumption receives support from an empirical test that uses data on the output of a large group of workers.

1. Introduction

This paper examines three assumptions commonly used in matching models of the labor market. The assumptions are roughly the following. First, the population distribution of match-quality coincides with the distribution facing each member of the heterogeneous populations of workers and employers. Second, in spite of the implicit heterogeneity of the two populations, evidence about the output of a match is entirely specific to that match--it is of no value in predicting either party's performance on other matches. Third, the distribution of match-quality is normal. The conclusion of the paper is that if one takes a Lancasterian approach to matching, then a reasonable set of axioms does imply all three assumptions as natural approximations. The paper also presents empirical evidence supporting the normality assumption. Although our results are favorable to these standard assumptions, we do not regard this exercise as a substitute for a thorough analysis of robustness with respect to changes in the assumptions.

Information-theoretic approaches to unemployment and turnover. There are two major explanations for the existence of turnover and unemployment. According to the first, re-allocation of labor takes place in response to a changing composition of product demand and a changing technology. This approach requires neither factor heterogeneity nor uncertainty: even if labor and capital were homogeneous, and even if demand-shifts and technology-shifts were perfectly foreseeable, labor turnover and unemployment would still take place. Consider, for instance, a deterministic version of the Lucas-Prescott (1974) model with perfectly foreseeable demand shifts across markets; the authors assume that a job change must be accompanied by a period of unemployment, and this leads to the existence of unemployment.

The second explanation relies on heterogeneity of jobs and workers, and on uncertainty. According to this approach, each worker would like to be in a job where he matches best. Since he does not know where his best match is, he samples jobs until he finds an acceptable one. In contrast to the first explanation, this approach requires neither shifting tastes nor changing technology. Its crucial aspects are the heterogeneity of jobs and workers, and incomplete information about the location of the optimal match. Without heterogeneity, all matches would be equally productive. With complete information, optimal matches would form immediately, and no further search would be needed.[1]

Of these two theories of match-dissolution, the term "matching model" has come to be associated only with the second, and it is with the latter class of models that the present paper is concerned.[2] Here, two distinct approaches have recently emerged. MacDonald (1982), following up on work by Rosen (1978), is explicit about the properties of the technology and about the nature of workers' endowments, which interact to cause a good match or a bad one. A second, more implicit approach to matching is taken by Flinn (1984), Jovanovic (1979a), and Miller (1984). Here the productivity of a match is "a draw" from a distribution whose origins are not explained. Though less explicit, this approach has gone farther towards empirical implementation than the first. At a high enough level of generality the two approaches might be equivalent, but it is clear that some of the standard working assumptions of the implicit approach[3] are <u>not</u> consistent with the MacDonald-Rosen approach. Such assumptions require explicit justification, and that is the aim of the present paper.

Standard assumptions of the implicit approach. The implicit approach to matching defines

$$x_t = \mu + \varepsilon_t \tag{1}$$

to be the period-t output of the match between a worker and an employer. The quality of the match is μ, and ε_t is a i.i.d. zero-mean disturbance.[4] Let $P(\cdot)$ be the distribution of μ among all matches: it is the distribution ot match-quality for a match between randomly chosen members of each population. The following three assumptions are commonly found in the matching literature; the first two bring tractability to the theoretical analysis, the third to the empirical analysis.

(i) Everyone is sampling from $P(\cdot)$. Whenever a new match forms, both parties regard μ as having been drawn from $P(\cdot)$. This implies an absence of hierarchical differences among members of either population.

(ii) Successive samples of μ are independent draws from $P(\cdot)$. Without this assumption, the history of the outcomes on past matches would be relevant for how one should view the future.

(iii) $P(\cdot)$ normal, and a normal distribution of ε_t. Jovanovic (1979a) and Miller (1984) assume this structure which is convenient for analyzing learning. Flinn (1984) assumes log-normal μ and ε.[5] Flinn and Heckman (1982) report estimates for an exponential specification; they find[6] that their estimates are extremely sensitive to the nature of the specification.

Flinn, Miller and others have used these assumptions (along with other assumptions about measured and unmeasured heterogeneity in general productive abilities) in implementing the matching model. Although these assumptions serve to identify the econometric models, they have not been tested, however, and no theoretical justification has yet been provided for them. One exception to this is Flinn and Heckman's finding (1982, p. 150) that accepted wages are better represented by a truncated normal distribution than by an exponential distribution. If one assumes (as Flinn and Heckman do) that wages are linear in μ, the same conclusions follow for $P(\cdot)$. However, while the Nash-bargain approach to wage determination (e.g., Pissarides (1983)) does under some circumstances imply that wages are linear in μ, many other approaches do not imply linearity. For example, the non-uniqueness result in Theorem 2 of Jovanovic (1979a) admits a wide variety of wage behavior. Moreover, risk-sharing also generally leads to non-linearities. Without some strong assumptions, therefore, it is impossible to recover (or test hypotheses about) the distribution of μ without direct measures of the productivity of the match, x_t.

Moreover, recent findings indicate that estimates of other parameters are likely to be highly sensitive to the way in which unobservables are assumed to be distributed (Heckman and Singer, 1984), and they too motivate this paper. We ask first whether there are "natural" theoretical considerations that imply such conditions, and, second, whether normality is supported by data on the physical output of some newly formed matches. The paper gives affirmative answers to both of the questions posed above. The theoretical justification relies on letting the number of unobserved characteristics (of the two parties to the match) get large, and appealing to a central limit theorem. The argument is not trivial, however, because we require that properties (i) - (iii) all hold in the limit. In the empirical test of normality, because the data are on physical output and not wages, they free us from the need to assume anything about the way that μ or x are related to wages.

Following some preliminary arguments in the next section, Section 3 gives the theoretical arguments, while Section 4 contains the empirical results.

2. The "characteristic" approach to matching

All the properties listed in the introductory section can be obtained quite easily by assuming that the characteristics of the parties to the match determine the output of the match, and that there is just a single characteristic per partner. Let λ be a characteristic of the firm, and θ a characteristic of the worker. Suppose that $\lambda, \theta \in S$, when S is the perimeter of the unit circle (see Marshall (1983) for such an interpretation). Let y be the absolute distance between λ and θ found by traversing the perimeter of the circle in the shortest direction. Suppose furthermore that λ and θ are both uniformly distributed along the circle, and that $\mu = f(y)$ is a deterministic production function ($f : [0,\pi] \to R^1$) (one could easily extend this and allow for the presence of a stochastic component). Then the distribution of the quality of the match is

$$\Pr\{\mu \leq r\} = \Pr\{y \leq f^{-1}(r)\} = f^{-1}(r)/\pi$$

because y is uniformly distributed on $[0,\pi]$. Since f is arbitrary, <u>any</u> distribution of μ can be obtained in this way. Moreover, since the distribution of agents of both types is uniform and on the circle, no agent can expect (upon a random match) to be able to do better than any other--property (i) holds too. Finally, if neither firms nor workers know their indices (λ and θ respectively) either before or after a meeting, then property (ii) holds as well.

The argument could be repeated for vector-valued λ and θ. The difficulty with it is that it does not lead to a natural restriction $P(\mu)$; if such restrictions are to emerge, further assumptions must be made, and this is what we turn to next.

3. Some assumptions and a limiting argument

The worker's skill is assumed to be fully characterized by a vector of attributes $\theta \in \Theta \subseteq R^n$. This vector includes traits that affect the worker's productivity. This vector denotes the average level of these characteristics. At any one time, the actual level of these characteristics is given by $z \in R^n$, where

$$z_t = \theta + u_t, \tag{2}$$

and where $u_t \in R^n$ is a vector of zero-mean random variables which are serially uncorrelated.

(A.1) Aggregation within each skill class. Let the production function be[7]

$$q = \lambda(z) \tag{3}$$

where q is the firm's total output, and $z \in R^n$ is a vector of unobserved latent factors of production available to the firm. The z vector is obtained by adding up the skill endowment vectors of the workers employed by the firm. Imbedded in this construction is the assumption that only the total amount of a given characteristic matters, not its distribution over the workers employed.[8] Formally, $\lambda \in \Lambda \equiv \{\lambda : R^n \to R^1, \lambda \text{ continuous}\}$.

(A.2) No interaction among workers. For each z, $z' \in R^n$, $\lambda(z + z') = \lambda(z) + \lambda(z')$. This requirement is stronger than constant returns to scale. It implies that λ is linear:

$$\lambda(z) = \sum_i \lambda_i z_i \tag{4}$$

for all z. This not only means that there are constant returns, but that in addition, the quality of the match between the worker and the firm depends only on the worker's characteristics and on the firm's technology, and not on the skills of the other workers in the firm, or on their number. Formally, $\Lambda \subseteq R^n$.

The economic meaning of this assumption is that the quality of the worker's match with his employer does not depend on the identity of his co-workers. If it did, the function $\lambda(\cdot)$ could not be linear.

Equation (4) gives us the linear decomposition of the output of the match given in equation (1): The match between firm λ and worker θ yields a period-t output equal to[9]

$$x_t = \sum_i \lambda_i \theta_i + \sum_i \lambda_i u_{it}, \tag{5}$$

which is the same as the expression for x_t given in equation (1), with the obvious definitions:

$$\mu = \sum \lambda_i \theta_i, \tag{6}$$

and

$$\varepsilon_t = \sum \lambda_i u_{it}. \tag{7}$$

This expression for ε_t is not symmetric because λ is deterministic, while z is random. The introduction of a random component in λ is straightforward, and is not central to our main interest.

(A.3) <u>No hierarchical differences across workers or firms</u>. The first assumption requires that all workers expect to produce the same amount with a randomly selected employer:

$$\underset{\lambda}{E} \sum \lambda_i \theta_i = \alpha, \tag{8}$$

for all $\theta \in \Theta$, where α is a positive constant. Similarly, assume that all firms expect to produce the same amount with a randomly selected worker:

$$\underset{\theta}{E} \sum \lambda_i \theta_i = \alpha, \tag{9}$$

for all $\lambda \in \Lambda$.

With this assumption we are abstracting from differences in "general ability" across workers. The abstraction is harmless if differences in general ability are generally known and observed by workers and employers alike; in that case, all our results can be said to apply to a submarket for workers with a given level of general ability. If general ability is not known with accuracy but if the worker knows no more about it than the employer, signal-extraction problems arise, for the match-quality and general ability must both be inferred on the basis of observing the output of the match. Moreover, if the worker knows more about his general ability than the employer, self-selection problems arise (see Weiss (1980), for instance).

Since λ is here interpreted as a property of the firm, the absence of hierarchical differences is really necessitated by the constant-returns assumption implicit in (A.2). Without (A.3) the best firm would absorb all the workers. On the other hand, if λ were to be interpreted as an index for some other factor whose quantities were in limited supply, then one could have hierarchical differences in this factor. For an analysis of matching two populations with members of each population being hierarchically different (see Sattinger (1975)).

(A.4) Random matching. If the matching process is one that pairs randomly selected workers with randomly selected firms, then the θ vector must be independent of the λ vector. This does not, of course, rule out correlation within the θ vector or within the λ vector (although the next assumption rules our perfect linear correlation within the two vectors). One should note that the "random matching process" assumption refers to the process by which initial contacts are made. This does not deny the existence of correlation between the θ's and the λ's on successful matches. Indeed, one expects that any matching process will lead to systematic sorting, albeit imperfect.

(A.5) No perfect linear correlation within θ or within λ. This assumption merely ensures that the spaces of skills and technology are really n-dimensional.

We shall now draw some implications of these assumptions. Equations (8) and (9) can be rewritten as

$$\sum\bar{\lambda}_i\theta_i = \sum\lambda_i\bar{\theta}_i = \alpha \qquad (\text{all } \lambda,\theta), \tag{10}$$

where $\bar{\theta}_i$ and $\bar{\lambda}_i$ denote the means of the marginal distributions of θ_i and λ_i respectively. Assumption (A.4) implies that the $\bar{\lambda}_i$ do not depend on the θ_i, and that the $\bar{\theta}_i$ do not depend on the λ_i. Then, (A.5) implies that at most one of the $\bar{\lambda}_i$ and at most one of the $\bar{\theta}_i$ can be non-zero. (If more than one $\bar{\lambda}_i$ were non-zero, then eq. (10) would represent an exact linear restriction between two or more θ's, in violation of (A.5). Similar reasoning applies if more than one $\bar{\theta}_i$ were non-zero.) Furthermore, eq. (10) implies that not all $\bar{\lambda}_i = 0$ and not all $\bar{\theta}_i = 0$. Therefore, exactly one θ_i and one λ_i must be degenerate random variables, and the index i must be the same. Therefore, we assume henceforth that that λ_1 and θ_1 are degenerate random variables, identically equal to $\bar{\lambda}_1$ and $\bar{\theta}_1$ respectively. We have thus obtained the restrictions:

$$\bar{\lambda}_i = \bar{\theta}_i = 0 \qquad i = 2,\ldots,n \tag{11}$$

and

$$\bar{\lambda}_1\bar{\theta}_1 = \alpha. \tag{12}$$

<u>Taking limits as $n \to \infty$</u>. We emphasize that properties (i) - (iii) of the introduction will, under certain mild restrictions to be noted below, hold only in the limit as $n \to \infty$. Before taking limits, however, we write down the conditions which we would like to be true in the limit. These conditions are (14) - (16) below. Note first that (11) and (12) imply that the unconditional variance of μ is[10]

$$\sigma_\mu^2(n) = \underset{\theta}{E}\left[\begin{matrix}E\{[\mu - E(\mu)]^2|\theta\} \\ \lambda\end{matrix}\right] = \underset{\theta}{E}\sum_{i,j} \theta_i\theta_j\sigma_{ij}(\lambda) = \sum_{i,j} \sigma_{ij}(\theta)\sigma_{ij}(\lambda), \tag{13}$$

where $\sigma_{ij}(\theta) \equiv \mathrm{cov}(\theta_i,\theta_j)$ and $\sigma_{ij}(\lambda) \equiv \mathrm{cov}(\lambda_i,\lambda_j)$. If the components of θ and λ are uncorrelated then $\sigma_\mu^2(n) = \sum_i \sigma_{ii}(\theta)\sigma_{ii}(\lambda)$. These results reflect the obvious fact that dispersion in the quality of the match reflects directly the dispersion in the characteristics of workers and of firms.

At a given firm, the variance of the output of the match (conditional on θ and λ) is

$$\sigma_\varepsilon^2(\lambda,n) = \sum_{i,j} \lambda_i\lambda_j\sigma_{ij}(u) \tag{14}$$

If no firm is to have an advantage in learning about μ, $\sigma_\varepsilon^2(\lambda,n)$ would have to be identical over firms. Thus eq. (14) is a nonlinear restriction on the admissible λ vectors.

Eq. (10) requires that the mean value of the random match be the same for each worker, and for each firm. A further requirement must be met, however. The variance of the distribution of μ must be the same for all firms and for all workers. Conditional on λ, this variance is

$$\sigma_\mu^2(\lambda,n) = \sum_{i,j} \lambda_i \lambda_j \sigma_{ij}(\theta) = \sigma_\mu^2(n), \tag{15}$$

and it must be a constant independent of λ. Conditional on θ, this variance is

$$\sigma_\mu^2(\theta,n) = \sum_{i,j} \theta_i \theta_j \sigma_{ij}(\lambda) = \sigma_\mu^2(n), \tag{16}$$

and it must be a constant independent of θ. The failure of the right-hand side of eq. (15) to hold would imply the existence of hierarchical differences among firms, thus contradicting (A.3). A firm with a larger $\sigma_\mu^2(\lambda,n)$ would do better in the long run than a firm with a lower $\sigma_\mu^2(\lambda,n)$, because it would, in effect, be searching from a distribution with a higher variance. The same considerations apply to $\sigma_\mu^2(\theta,n)$. That the total rents to a random match grow with σ_μ^2 is, of course, a standard result in the matching literature. On the other hand, if $\sigma_\varepsilon^2(\lambda,n)$ in eq. (14) was different across firms, some firms would learn faster than others, which again would given them an advantage.

To summarize: equations (10), (15) and (16) imply that for all λ and θ, the first and second moments of $P(\mu|\lambda,n)$ and $P(\mu|\theta,n)$ are equal to the first and second movements of $P(\mu|n)$, namely, α and $\sigma_\mu^2(n)$

respectively. In addition, eq. (14) implies that the second moment of ε (as well as the first) is the same for all λ (and, of course, for all θ as well). At first glance this may appear to be insufficient to guarantee the truth of condition (i), which states that the entire distribution $P(\mu|\lambda,n)$ and $P(\mu|\theta,n)$ must, for all λ and θ, be the same as $P(\mu|n)$. Although these distributions will generally not coincide for finite n, they will do so in the limit, because the limiting distributions $P(\mu|\lambda,n)$, $P(\mu|\theta,n)$ and $P(\mu|n)$ will all be normal. Since the normal distribution is fully characterized by its first two moments, all the higher moments will coincide as well.

We now turn to the arguments that yield (14) - (16) in the limit. For $i = 2,\ldots,n$, let

$$\theta_i = n^{-1/4}\eta_i$$

and (17)

$$\lambda_i = n^{-1/4}\xi_i$$

where $\{\eta_i\}$ and $\{\xi_i\}$ are stationary sequences of zero-mean variates. Although the η_i and ξ_i are independent of one another, there may be arbitrary correlation among the η's and among the ξ's. Let $\sigma_{ij}(\eta) \equiv \operatorname{cov}(\eta_i,\eta_j)$ and $\sigma_{ij}(\xi) \equiv \operatorname{cov}(\xi_i,\xi_j)$. Assume that for some $s \in (0,\infty)$,

$$\lim_{n\to\infty} n^{-1} \sum_{i,j} \sigma_{ij}(\eta)\sigma_{ij}(\xi) = s. \tag{18}$$

(This is not a strong requirement and we shall comment on it following the proof of the theorem.) Let

$$w_k \equiv \sum_{i=2}^{k} [\eta_i \eta_{k-i} - \sigma_{i,k-i}(\eta)]\sigma_{i,k-i}(\xi)$$

and

$$y_k \equiv \sum_{i=2}^{k} [\xi_i \xi_{k-i} - \sigma_{i,k-i}(\xi)]\sigma_{i,k-i}(\eta)$$

Assume the existence of a constant K such that

$$\text{var}(w_k) < K \quad \text{and} \quad \text{var}(y_k) < K, \tag{19}$$

for all k. In addition, assume that there exist strictly positive constants M and δ such that

$$\frac{1}{n} E(w_n \sum_{k=2}^{n} w_k) < \frac{M}{n^\delta} \quad \text{and} \quad \frac{1}{n} E(y_n \sum_{k=2}^{n} y_k) < \frac{M}{n^\delta} . \tag{20}$$

Finally, we assume a mixing condition directly on the sequence $\{\eta_i \xi_i\}$. For a sequence $v_1, v_2, \ldots$ of random variables, let α_r be a number such that

$$|P\{(v_1,\ldots,v_k) \in A,\ (v_{k+r}, v_{k+r+1},\ldots) \in B\} -$$

$$P\{(v_1,\ldots,v_k) \in A\} P\{(v_{k+r}, v_{k+r+1},\ldots) \in B\}| \leq \alpha_r$$

for k, $r \geq 1$, and for all possible sets A of realizations of $(v_1,\ldots,v_k)$ and B of $(v_{k+r}, v_{k+r+1},\ldots)$ [Billingsley 1978, p. 315]. We assume that for $v_i = \eta_i \xi_i$ $(i = 1,2,\ldots)$,

$$\alpha_r = O(r^{-5}). \tag{21}$$

Theorem. If (A1) - (A5) and (17) - (21) hold, and if η_i and ξ_i are uniformly bounded w.p. 1,

$$\left.\begin{array}{l} P(\mu|n) \\ P(\mu|\theta,n) \\ P(\mu|\lambda,n) \end{array}\right\} \xrightarrow{W} N(\alpha,s)$$

for all $\theta \in \Theta$, $\lambda \in \Lambda$.

Proof. Substitution from (17) into (15) and (16) followed by a rearrangement yields

$$\sigma_\mu^2(\theta,n) = n^{-1} \sum_{i,j} [\eta_i\eta_j - \sigma_{ij}(\eta)]\sigma_{ij}(\xi) + n^{-1} \sum_{i,j} \sigma_{ij}(\eta)\sigma_{ij}(\xi)$$

and

$$\sigma_\mu^2(\lambda,n) = n^{-1} \sum_{i,j} [\xi_i\xi_j - \sigma_{ij}(\xi)]\sigma_{ij}(\eta) + n^{-1} \sum_{i,j} \sigma_{ij}(\eta)\sigma_{ij}(\xi).$$

Therefore, in view of (18),

$$\lim_{n\to\infty} [\sigma_\mu^2(\theta,n) - s] = \lim_{n\to\infty} n^{-1} \sum_{i,j} [\eta_i\eta_j - \sigma_{ij}(\eta)] \tag{22}$$

and

$$\lim_{n\to\infty} [\sigma_\mu^2(\lambda,n) - s] = \lim_{n\to\infty} n^{-1} \sum_{i,j} [\xi_i\xi_j - \sigma_{ij}(\xi)]. \tag{23}$$

The next step in the proof is to show that the right-hand sides of (22) and (23) converge to zero almost surely. The definition of w_k and y_k imply that

$$\sum_{i,j} [\eta_i\eta_j - \sigma_{ij}(\eta)]\sigma_{ij}(\xi) = \sum_{i=2}^{n} w_k$$

and

$$\sum_{i,j} [\xi_i \xi_j - \sigma_{ij}(\xi)]\sigma_{ij}(\eta) = \sum_{i=2}^{n} y_k.$$

Conditions (19) and (20) and the result of Parzen [(1960), p. 420, Thm 2B] imply that $P[\lim_{n\to\infty} n^{-1}\sum w_k = 0] = 1$ and, similarly, $P[\lim_{n\to\infty} n^{-1}\sum y_k = 0] = 1$. Therefore, $\sigma_\mu^2(\theta,n)$ and $\sigma_\mu^2(\lambda,n)$ converge to s almost surely, for all θ and λ. The mean of the limiting distributions of μ is α by construction. It remains to show asymptotic normality. But this follows from eq. (21) which, along with the boundedness assumptions, ensures that the conditions of Billingsley's theorem 27.5 (1978, p. 316) are met. This completes the proof.

Remark. The theorem establishes properties (i) and (iii). Property (ii) also follows because a particular draw of μ says nothing (in the limit) about either λ or θ. This is established by noting that since the limiting distribution of μ is normal, independence of two successive draws μ and μ' follows for each θ if μ and μ' are uncorrelated:

$$\mathrm{Cor}(\mu,\mu'|\theta) = \frac{1}{n^2} \sum_{i,j\geq 2} \theta_i \theta_j E(\lambda_i \lambda_j') = 0,$$

since λ and λ' are independent because of (A.4). Similarly, $\mathrm{Cor}(\mu,\mu'|\lambda) = 0$ for all λ. Hence in the limit, (ii) holds as well. Of course, lack of linear correlation implies independence only in the limit, when μ and μ' are normal.

Example. Let $\sigma^2_{ij}(\eta) = \rho_\eta^{|i-j|}\sigma^2_\eta$ and $\sigma^2_{ij}(\xi) = \rho_\xi^{|i-j|}\sigma^2_\xi$, where $\rho_\eta, \rho_\xi \in [0,1)$, and let $E(\eta_i\eta_j)^2 < \infty$ and $E(\xi_i\xi_j)^2 < \infty$, uniformly in i,j. We now show that (18) - (20) are satisfied. For (18), we note that $n^{-1}\sum_{i,j}\sigma_{ij}(\eta)\sigma_{ij}(\xi) = n^{-1}\sigma^2_\eta\sigma^2_\xi\sum_{i,j}(\rho_\eta\rho_\xi)^{|i-j|} \to \sigma^2_\eta\sigma^2_\xi(1 + \rho_\eta\rho_\xi)/(1 - \rho_\eta\rho_\xi)$, so that (18) is satisfied.

For (19), let $u_{ik} \equiv \eta_i\eta_{k-i} - \sigma_{i,k-i}(\eta)$, so that the u_{ik} have zero mean. Then $\mathrm{var}(w_k) = \sigma^2_\xi\sum_{i,j}\mathrm{cov}(u_{ik},u_{jk})\rho_\xi^{|k-i|+|k-j|}$. The uniform boundedness assumption on $E(\eta_i\eta_j)^2$ implies that there exists a constant A such that $\mathrm{cov}(u_{ik},u_{jk}) \leq A$, all i,j,k. Then $\mathrm{var}(w_k) \leq \sigma^2_\xi A\sum_{i,j}\rho_\xi^{|k-i|+|k-j|} \to \sigma^2_\xi A(1 - \rho_\xi)^{-2}$. An identical bound can be established for $\mathrm{var}(y_k)$, so that (19) is satisfied.

For (20), note that for $1 < k \leq n$, $Ew_nw_{n-k} \leq \sigma^2_\xi\rho^k_\xi Ew^2_n$ follows straightforwardly from the definition of w_k. Therefore $n^{-1}Ew_n\sum_{k=2}^{n}w_k < \sigma^2_\xi n^{-1}Ew^2_n/(1 - \rho_\xi)$. Since $Ew_n = 0$ for each n, $Ew^2_n = \mathrm{var}(w_n)$, and the latter is uniformly bounded because (19) holds. Therefore (20) holds as well.

Condition (21) is not implied by the above covariance conditions, although the latter would be implied by α-mixing on $\{\eta_i\}$ and $\{\xi_i\}$ individually with $\alpha_r = O(b^r_\eta)$ and $O(b^r_\xi)$ respectively, which in turn would imply α-mixing on $\{\eta_i\xi_i\}$ with $\alpha_r = O(\max[b^r_\eta, b^r_\xi])$. Such α-mixing on $\{\eta_i\}$ and $\{\xi_i\}$ would be implied if, for instance, $\eta_i = \rho_\eta\eta_{i-1} + \varepsilon_i$ and $\xi_i = \rho_\xi\xi_{i-1} + \varepsilon'_i$ $(i \geq 2)$ [Billingsley 1979, example 27.5, p. 315).

The case of uncorrelated η's and ξ's results, of course, if $\rho_\eta = \rho_\xi = 0$. We do not regard this as a realistic case, especially if the characteristics are narrowly defined.

Economic interpretation. In what sense (if any) are these limiting results natural? Acceptance of (A1) - (A5) still leaves open the question of whether firms and workers can be viewed as entities composed of sizable numbers of characteristics that are "sufficiently" independent. Large numbers of characteristics can, of course, be obtained simply by narrowing down the classification of characteristics. In the process of doing so, however, the degree of dependence among the characteristics would increase. In terms of the above example, an increase in n resulting from a change in the way characteristics are defined would be accompanied by an increase in the $\sigma_{ij}(\eta)$ and $\sigma_{ij}(\xi)$ terms.

Intuitively, our results say that with enough random variability in firms' technologies and in workers' characteristics, and with enough dimensions in which such heterogeneity can manifest itself, one expects assumptions (i), (ii) and (iii) to be close approximations to reality.

4. Testing for normality

In this section we report the results of our analysis of a data set that contains direct measures of output of newly hired workers of a large manufacturing company. Data on the workers' characteristics and those of their supervisors is also available, and the output data cover the employees' first month of employment.[11] An advantage of using the first month of tenure is that no significant selecting out through separations has yet taken place.

If μ and ε are indeed normally distributed as argued in the previous section, then x should also be normally distributed (see eq. (1)). In other words, a finding of non-normality for x would constitute a rejection of the development of the previous section. Strictly speaking, with a finite number of characteristics, n, exact normality is not expected, of course, but something close to it should be true if n is large.

Flinn (1984) and Miller (1982) assume that x is the residual component of an output equation, a residual that cannot be predicted ex-ante on the basis of observable characteristics of the worker and of his employer. Assuming a linear relationship relating worker w's output y_w to the observable characteristics of worker z_w, and those of his supervisor $z_{s(w)}$, we estimate the equation

$$y_{s(w),w} = \beta_0 + \beta_1 z_w + \beta_2 z_{s(w)} + x_{s(w),w} \quad w = 1,\ldots,n \tag{17}$$

Here s(w) denotes the identity of the supervisor assigned to worker w, and n is the size of the sample.

The assignment of workers to supervisors (the s(w) function) seems to have been entirely random. Table 2 (in the appendix) reports the correlation coefficients between, on the one hand, the characteristics of the workers and, on the other, the characteristics of the supervisors that those workers are matched with. This suggests the absence of any effects on output from the matching of these observable characteristics, for if some were present, the firm would assign workers to supervisors in a systematic way so as to take advantage of such effects. Our attempts to get at these effects directly by including in (17) some interaction terms such as black•black, female•female, and so forth, did not lead to any significant effects.[12]

In addition to the variables on the characteristics of workers and their supervisors, a "job complexity" variable was entered as a regressor. This variable was constructed by engineers in the two plants as a measure of the average performance (by previous workers) on the job presently occupied by the worker in question.

Measurement error in the dependent variable is not usually thought to be a problem. In the present case, however, a normally distributed measurement error could tend to lead us to accept normality of x even if it were false. We do not believe, however, that this is a problem here: Daily records of each worker's output were kept, and each supervisor was on average responsible for some twenty or so workers, so that it certainly was feasible for accurate records to be kept.

The regression results are reported in Table 1.[13] Our main interest is not in the estimates of the coefficients, but rather in the Kolmogorov-Smirnov statistics and the Cramér-Von Mises statistics, also reported in Table 1.

In Figure 1 we plot the cumulative standard normal distribution evaluated at the standardized values of the residuals (standardized by dividing through by their standard deviation). If the distribution of the residuals were normal, the plots would indicate a 45° line, which they almost do, with plant 1 doing somewhat better than plant 2.

Table 1 and Figure 1 about here

Using the tables reported in Stephens (1974) we find that for plant 1, the Kolmogorov-Smirnov statistic of .046 does not permit rejection of the null hypothesis of normality at the 5% level of significance. On the other hand, the Cramér-Von Mises statistic (which is based not on the most extreme deviation of the empirical CDF from the theoretical CDF but rather on their average deviation) allows rejection at the 5% level of significance, but not at the 1% level. For plant 2, however, both statistics lead to rejection even at the 1% level.

Since the sample size for both plants is quite large, we had realistically expected a more decisive rejection of normality for both plants than in fact turned out to be the case. At any rate, Figure 1 makes it clear that an assumption of normality for the distribution of μ is not far off the mark. There remains the possibility, however, that the residuals include both unmeasured <u>hierarchical</u> heterogeneity as well as measurement error, and so our conclusions on this score remain tentative.

Appendix: Data Description

The data used in this analysis are composed of two datasets collected on employees from two locationally separate manufacturing plants of a large U.S. firm. The data contain actual productivity measures of semi-skilled production workers. The data also match these individuals with additional information on their supervisors. The lowest or level 1 supervisor is called a section chief, while his immediate superior is called a department chief.

The semi-skilled production workers' data base was created using information obtained from personnel records on all new employees hired in 1979 at the two plants. Each worker's immediate supervisor was then identified, and their personnel records were then obtained. Table 2 gives the pattern of correlation of the assignment by characteristics. Assignment seems to have been entirely random--all the correlation coefficients are negligible.

References

Billingsley, P. Probability and Measure. New York: Wiley, 1979.

Diamond, P. "Aggregate Demand Management in Search Equilibrium." Journal of Political Economy 90, no. 5 (October 1982): 881-94.

Flinn, C. "Wages and Mobility of Young Men." Journal of Political Economy (1984), forthcoming.

Flinn, C., and J. Heckman. "New Methods for Analyzing Structural Models of Labor Force Dynamics." Journal of Econometrics 18 (1982): 115-68.

Heckman, J., and B. Singer. "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." Econometrica 52, no. 2 (March 1984): 271-320.

Jovanovic, B. "Job-Matching and the Theory of Turnover." Journal of Political Economy 87, no. 5 (October 1979): 972-90. (1979a)

Jovanovic, B. "Firm-Specific Capital and Turnover." Journal of Political Economy 87, no. 6 (December 1979): 1246-60. (1979b)

Lucas, R. E., and E. C. Prescott. "Equilibrium Search and Unemployment." Journal of Economic Theory 7, no. 2 (February 1974): 188-209.

MacDonald, G. "A Market Equilibrium Theory of Job Assignment and Sequential Accumulation of Information." American Economic Review 72, no. 5 (December 1982): 1038-55.

MacLeish, D. L. "Dependent Central Limit Theorems and Invariance Principles." Annals of Probability 2, no. 4 (1974): 620-28.

Marshall, J. "An Equilibrium Model of Search and Non-Price Rationing." U. C. Santa Barbara working paper no. 231 (December 1983).

Miller, R. "Job-Matching and Occupational Choice." Journal of Political Economy 92 (December 1984).

Parzen, E. Modern Probability Theory and Its Applications. New York: Wiley, 1960.

Pissarides, C. "Efficient Employment with Endogenous Jobs and Job Rejection." London School of Economics working paper (1983).

Rosen, S. "Substitution and the Division of Labor." Economica 45 (August 1978): 235-50.

Sattinger, M. "Comparative Advantage and the Distributions of Earnings and Abilities." Econometrica 43, no. 3 (May 1975): 455-68.

Stephens, M. "EDF Statistics for Goodness of Fit and Some Comparisons." Journal of the American Statistical Association 69 (1974): 730-37.

Weiss, A. M. "Job Queues and Layoffs in Labor Markets with Flexible Wages." Journal of Political Economy 88, no. 3 (June 1980): 526-38.

Welch, F. "Linear Synthesis of Skill Distributions." Journal of Human Resources 4, no. 3 (Summer 1969): 311-27.

Wilson, C. "A Model of Job-Search and Matching." University of Wisconsin paper (1980).

Footnotes

1. In this approach, with a fixed, infinitely-lived population of workers and jobs, all job change would eventually cease. For positive job change in the steady state, either workers or jobs must have finite lives.

2. Recent work on frictions in the meeting technology [e.g., Diamond (1982)] imposes an exogenous limit on the duration of matches and cannot be placed in either of the above categories.

3. For instance, assumption (ii) below.

4. In some of the theoretical papers $\varepsilon_t \equiv 0$ so that the quality of the match can be discovered immediately upon contact [Jovanovic (1979b), Wilson (1980), Flinn and Heckman (1982), Pissarides (1983), Marshall (1983)].

5. Flinn also makes certain other assumptions which make agents behave as though the structure were normal rather than log-normal.

6. See their Section 3.3.

7. Looking ahead, the form of the function $\lambda(\cdot)$ will differ across firms, reflecting the heterogeneity in the technology employed by firms. Thus while θ indexes workers, λ indexes firms.

8. Welch (1969) makes a similar assumption.

9. We use x to denote the output of the match between the firm and a given worker, and q to denote the firm's total output.

10. Because of the degeneracy of θ_1 and λ_1, all the covariance expressions involve summation over i,j strictly greater than one.

11. A more detailed description of the data is contained in the appendix.

12. One interaction-type variable showed up significantly positive in the regression for plant 1 but not for plant 2. This was a dummy variable which was defined to be unity if the worker's age was within five years of the age of his supervisor and if the worker was younger than his supervisor, and zero otherwise. However, in plant 2, the same variable had a t-ratio of .16. Moreover, if the relationship were true in plant 1, the assignment should have reflected this and Table 2 should have shown a significant positive relation between the ages of the worker and of his supervisor. In fact, the correlation is negative. Had we found important interactive effects with either race or sex variables (which we did not) while at the same time finding that the company did not act on this by assigning systematically by race or sex, this could possibly be explained as grounds of a possible fear of being sued for discrimination. No such interpretation is possible for the different age interaction effect in plant 1 and for the company's failure to act on it.

13. The regression for plant 1 is somewhat "better" than that for plant 2, both in terms of R^2 and in terms of the t-ratios, which are generally higher for plant 1 in spite of the smaller sample size. Although the nature of the work performed in the two plants was roughly the same, the nature of the job was somewhat less complex in plant 2.

Table 1. Output regressions (t-statistics in parentheses).

variable		Plant 1	Plant 2
intercept		84.1 (3.8)	37.6 (2.1)
workers' characteristics	black	-8.3 (2.6)	-4.7 (1.6)
	female	-15.2 (5.8)	-12.2 (4.0)
	education	2.6 (2.2)	1.2 (1.3)
	age	-.02 (.1)	-.08 (.5)
	married	4.8 (2.7)	3.7 (1.6)
supervisors' characteristics	black	-3.6 (.6)	-5.0 (1.2)
	female	1.2 (.2)	-.8 (.2)
	education	.7 (1.0)	.4 (.6)
	age	-.4 (2.3)	.4 (2.1)
Job complexity		.9 (3.3)	.6 (1.2)
Sample size		355	672
R^2		.16	.05
Kolmogorov-Smirnov statistic		.046	.062
Cramér-Von Mises statistic		.15	.46

Table 2. Assignment correlations.

Plant 1

Supervisors' characteristics	Employees' Characteristics				
	Black	Female	Education	Age	Married
Female	-0.02	-0.03	0.09	0.00	0.06
Black	0.06	-0.03	0.11	0.01	-0.00
Education	0.03	-0.15	-0.07	-0.02	-0.03
Age	-0.03	-0.06	-0.02	-0.03	-0.05
Job complexity	-0.09	0.09	-0.03	-0.03	0.04

Plant 2

Supervisors' characteristics	Employees' Characteristics				
	Black	Female	Education	Age	Married
Female	0.04	0.06	-0.03	-0.00	-0.03
Black	0.05	0.02	-0.04	-0.02	-0.04
Education	0.01	0.10	0.01	0.01	0.01
Age	-0.04	-0.02	-0.00	-0.00	-0.01
Job complexity	0.08	0.05	0.00	0.00	0.09

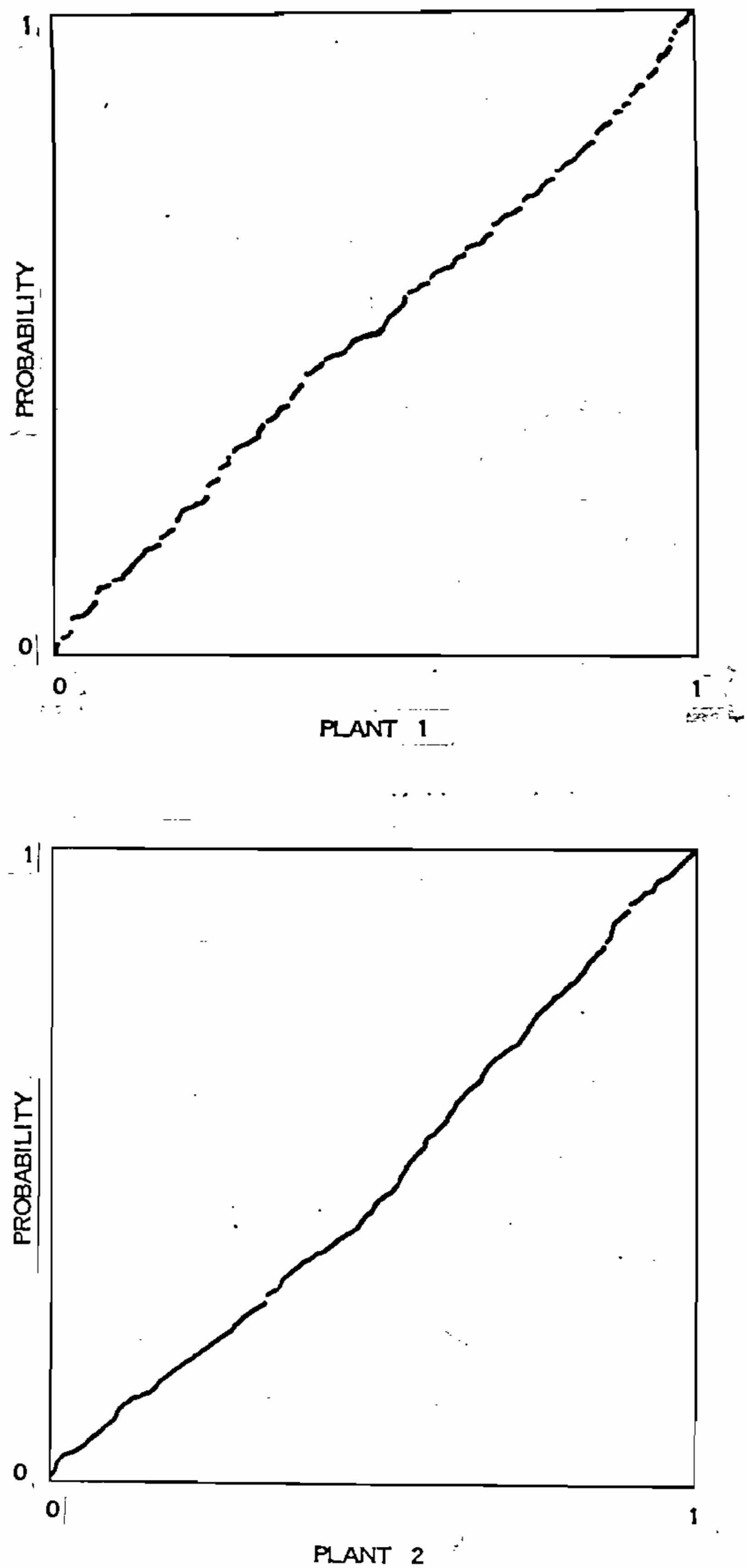

FIGURE 1 : PROBABILITY OF NORMALIZED RESIDUALS