Econometrica, Vol. 58, No. 5 (September, 1990), 1121–1149

THE EMPIRICAL CONTENT OF THE ROY MODEL

BY JAMES J. HECKMAN AND BO E. HONORÉ[1]

This paper clarifies and extends the classical Roy model of self selection and earnings inequality. The original Roy model, based on the assumption that log skills are normally distributed, is shown to imply that pursuit of comparative advantage in a free market reduces earnings inequality compared to the earnings distribution that would result if workers were randomly assigned to sectors. Aggregate log earnings are right skewed even if one sectoral distribution is left skewed. Most major implications of the log normal Roy model survive if differences in skills are log concave. However few implications of the model survive if skills are generated from more general distributions. We consider the identifiability of the Roy model from data on earnings distributions. The normal theory version is identifiable without regressors or exclusion restrictions. Sectoral distributions can be identified knowing only the aggregate earnings distribution. For general skill distributions, the model is not identified and has no empirical content. With sufficient price variation, the model can be identified from multimarket data. Cross-sectional variation in regressors can substitute for price variation in restoring empirical content to the Roy model.

KEYWORDS: Roy model, log concave random variables, log convexity, self selection and earnings inequality, identifiability, total positivity.

INTRODUCTION

THIS PAPER EXAMINES the empirical content of Roy's model (1951) of the distribution of earnings and the robustness of conclusions drawn from it when its normality assumptions are relaxed. Roy's model was developed to explain occupational choice and its consequences for the distribution of earnings when individuals differ in their endowments of occupation-specific skills.

Placing a broad interpretation on the "occupations" in his model, Roy's framework is a prototype for a whole class of models of self selection in labor markets. A woman's choice between market and nonmarket work and its consequences for observed wages have been analyzed within the framework of the Roy model (Gronau (1974), Heckman (1974)). A worker's choice between union and nonunion sectors has been modeled using Roy's setup or extensions of it (Lee (1978)). The analysis of choice of geographical location (Robinson and Tomes (1983)), schooling levels (Willis and Rosen (1979)), occupational choice with endogenous specific human capital (Miller (1984)), choice of industrial sectors (Heckman and Sedlacek (1985)), choice of marital status (McElroy and Horney (1981)), and the consequences of these choices for earnings inequality all fall within the general framework of the Roy model.

[1] This research was supported by NSF Grants SES-84-11242, SES-87-11596, and SES-88-09352. The first draft was prepared in September, 1987. This draft was prepared in August, 1989. We thank Ricardo Barros, Steve Cameron, Kyungsoo Choi, Angus Deaton, Joe Hotz, Ken McLaughlin, Jim Walker, and three anonymous referees for comments as well as workshop participants at Brigham Young University, Northwestern University, Pennsylvania State University, the University of Rochester, and Yale University for comments on previous drafts. Kyungsoo Choi also provided valuable research assistance on this project.

That model and virtually all of its progeny assume that log latent skills or log potential wages in alternative sectors are normally distributed. The normality assumption is convenient and fruitful in producing theoretical implications which we summarize below. Yet empirical support for it is weak. Heckman and Sedlacek (1985, 1990) test for and reject the log normality assumption in U.S. data on wages and sectoral choice. Lee (1982) rejects normality as a description of log wages in his study of union-nonunion wage differentials.

This evidence motivates our paper. We examine the robustness of the qualitative predictions of the normal theory Roy framework when it is assumed that log wages in alternative sectors are not normal. Many conclusions are not robust. However, we produce a broad class of nonnormal models based on log concave random variables which preserves the essential features of the original Roy model. We also derive new implications about aggregate income inequality for the log-normal-skill Roy model. Of special interest is our result that in an economy described by a Roy model, self selection leads to *reduced* inequality in earnings compared to an economy in which individuals are randomly assigned to jobs. An application of this analysis is that in a Roy economy a volunteer army reduces earnings inequality compared to inequality in an economy with forced military conscription. We further note that the *distribution* of aggregate log earnings produced from a log normal Roy economy is log concave although the density is not. This result imposes testable restrictions on the means and variances of aggregate log earnings measured below threshold values.

We also consider the identifiability of the Roy model. Its log normality assumption has exceedingly powerful and counter-intuitive empirical consequences that have not been fully appreciated in the literature. If the Roy model is valid, it is possible to estimate the population distribution of latent skills or potential wages from a single cross-section of data on wages knowing only the wages of persons who select a sector. A stronger result can also be proved. It is not necessary to know which sector a person selects in order to estimate the population distribution of latent skills. No regressors need appear or conventional econometric exclusion restrictions be invoked to identify all parameters of the Roy model. As a consequence of Roy's normality assumption, it is possible to determine the correlation of latent skills or potential wages of persons even though one observes only one skill of any person. The correlation of workers' skills in different sectors plays a key role in Roy's original model and is a cornerstone of the Willis-Rosen (1979) "hierarchical model" of the labor market.

This strong identifiability feature of the Roy model is not robust when log skills are nonnormal. Unless specific distributional assumptions are invoked, it is possible to rationalize any cross-sectional distribution of wages by a model with skills that are independent across sectors or skills that are arbitrarily highly correlated. Accordingly, the notion of skill hierarchies has no empirical content in a general nonnormal model. A single skill model or a multiple skill model will explain cross-section data on wages equally well.

These negative conclusions are fundamentally altered, however, if there are data on wage distributions from different markets with different relative prices of skills. Provided that there is sufficient price variation in a sense that will be made precise below, it is possible to identify latent skill distributions from cross-section wage distributions from different markets. It is possible to recover population skill distributions even though the observing economist does not know which particular skill a worker is using (i.e. which sector is actually chosen) and even if one skill is never observed (e.g. an agent's skill in the nonmarket sector). Thus if there is access to data on wages from different markets, notions of skill hierarchies have empirical content. We also demonstrate how variation in regressors that shift location parameters of skill distributions in a single cross section can substitute for multimarket data in identifying latent skill distributions.

The analysis in this paper has implications for the closely related problem of identification in the competing risks model of duration analysis. (See Kalbfleish and Prentice (1980) or Cox and Oakes (1984).) A competing risks model is one in which distinct causes of an event compete with each other to cause a person's death. For example, cancer, stroke, heart failure, etc., may compete with each other as the cause. The *minimum* of the potential times of death from each cause determines the cause of death. The problem posed in the competing risks literature is to determine the joint distribution of times to death from all possible causes when one observes only the time to death due to a particular cause for any observation. The identification problem raised in that literature is thus isomorphic to the identification problem in the Roy model in which only one skill, of many, is observed for any economic agent. In the Roy model it is the *maximum* potential sectoral wage (defined as skill times skill price) that determines the chosen sector. The identification problem for the Roy model is to derive the potential wage or skill distribution from observed wages or skills. A companion paper (Heckman and Honoré (1989)) considers identification for a class of nonparametric competing risks models widely used in duration analysis.

The plan of this paper is as follows. Section 1 presents the Roy model and new implications of it not previously noted in the literature. We show that many features of the Roy model carry over to a more general class of models based on log concave random variables. However, without some distributional structure imposed, most implications of the model can be reversed. Section 2 considers the identifiability of the Roy model and the benefits of multimarket data in securing identification and the additional benefits of having access to longitudinal data and regressors that shift the location of the skill distribution. The paper concludes with a summary and suggestions for future research.

1. THE ROY MODEL

This section exposits the two-skill Roy model and establishes new properties of it. Income maximizing agents possess two skills S_1 and S_2 with associated positive skill prices π_1 and π_2. Skills are assumed to be nonnegative. Individuals differ in their skill endowment and each person knows his own endowment. The population distribution of skills is $F(s_1, s_2)$. Throughout this paper we use upper-case letters for random variables and lower-case letters for their realizations.

Skill i is useful only in sector i. Agents can work in only one sector in any period. There are no mobility costs and sector-specific skills cannot be augmented by personal investment decisions. As a consequence of these assumptions, skill prices may be uncertain or certain without affecting any of our conclusions provided that sectoral choice decisions are made after skill prices are revealed.

An agent chooses sector one if his earnings are greater there,[2] i.e.,

$$(1) \qquad \pi_1 S_1 > \pi_2 S_2.$$

Otherwise the agent chooses sector two. If (S_1, S_2) are continuously distributed in the population and are nondegenerate random variables, the population proportion of persons who are indifferent between sectors is negligible and can be set to zero. Essentially all persons have a preference for sector one or sector two. Throughout, we assume that (S_1, S_2) has a well-defined density $f(s_1, s_2)$.

The proportion of the population working in sector one, P_1, is the proportion for whom (1) is true:

$$(2) \qquad P_1 = \int_0^\infty \int_0^{\pi_1 s_1/\pi_2} f(s_1, s_2)\, ds_2\, ds_1.$$

An immediate consequence of (2) is that as the relative price of skill one increases, a greater proportion of the population works in sector one.

The density of skill employed in sector one differs from the population density of skill. The latter density may be written as

$$(3) \qquad f(s_1) = \int_0^\infty f(s_1, s_2)\, ds_2.$$

这是总体的技能分布

The former density is

这是能够看到的技能功能分布(因为自选择导致observed不等于population distribution)

$$(4) \qquad g_1(s_1 | \pi_1 S_1 > \pi_2 S_2) = \frac{1}{P_1} \int_0^{\pi_1 s_1/\pi_2} f(s_1, s_2)\, ds_2.$$

In general, the two densities will differ unless π_1/π_2 becomes so large that P_1 becomes one and everyone works in sector one, or unless $\pi_1 S_1 - \pi_2 S_2$ is independent of S_1. The distribution of skills observed in sector one differs from the population distribution of skills. As individuals pursue their comparative advantage, the observed distribution of skills employed in sector one differs from the population skill distribution.

A person using skill S_1 in sector one earns $W_1 = \pi_1 S_1$. Given our efficiency units assumption for sector-specific skills, the distribution of earnings in sector one is obtained from (4) by an elementary transformation of variables. An analogous argument produces the density for earnings in sector two.

The density of earnings in the economy at large, $g(w)$, is a weighted average of the densities in each sector where the weight applied to the sector i density is

[2] Throughout this section, we focus on sector one. Because of the symmetry between the sectors it is clear that all the results derived for sector one can be derived for sector two as well.

the proportion of the population in the sector:

$$g(w) = P_1 g_1(w) + P_2 g_2(w).$$

As skill prices vary, individuals shift sectors in pursuit of their comparative advantage. Such shifts alter the distribution of earnings within sectors and in the economy at large.

Does the pursuit of comparative advantage increase or decrease earnings inequality within sectors and in the overall economy? Do the people with the highest i skill levels actually work in sector i? As people enter a sector in response to an increase in the demand for its services, does the average skill level employed there rise or fall? These are the questions addressed by the Roy model.

In this subsection we assume that $(\ln S_1, \ln S_2)$ has finite mean, (μ_1, μ_2) and variance $\mathbf{\Sigma}$. Define $U_i = \ln S_i - \mu_i$. (U_1, U_2) will then have mean $\mathbf{0}$ and variance $\mathbf{\Sigma}$. We further assume that the distribution of (U_1, U_2) is not degenerate.

As a consequence of these assumptions,

$$\ln W_i = \ln \pi_i + \mu_i + U_i, \qquad i = 1, 2. \tag{5}$$

The expressions that follow will be simplified if we define

$$c = \ln(\pi_1/\pi_2) + \mu_1 - \mu_2, \qquad \sigma^2 = \sigma_{11} + \sigma_{22} - 2\sigma_{12},$$
$$a_1 = \frac{\sigma_{11} - \sigma_{12}}{\sigma^2}, \qquad a_2 = a_1 - 1 = \frac{-\sigma_{22} + \sigma_{12}}{\sigma^2},$$
$$D_* = U_1 - U_2, \qquad V = a_1 U_2 - a_2 U_1,$$
$$c_* = c/\sigma, \qquad D_* = D/\sigma,$$
$$\rho = \operatorname{corr}[D, U_1] = \frac{\sigma_{11} - \sigma_{12}}{\sigma\sqrt{\sigma_{11}}} = a_1\sigma/\sqrt{\sigma_{11}}.$$

Then

$$U_i = a_i D + V, \tag{6}$$

and

$$E[D] = 0, \qquad E[V] = 0,$$
$$\operatorname{Var}[D] = \sigma^2, \qquad \operatorname{Var}[V] = \frac{\sigma_{11}\sigma_{22} - \sigma_{12}^2}{\sigma^2} = \sigma_{11}(1 - \rho^2),$$
$$\operatorname{Cov}[D, V] = 0.$$

Hence (5) can be written as

$$\begin{aligned} \ln W_i &= \ln \pi_i + \mu_i + a_i D + V \\ &= \ln \pi_i + \mu_i + a_i \sigma D_* + V \end{aligned} \tag{7a}$$

and

$$\begin{aligned} \ln W_1 - \ln W_2 &= \ln \pi_1 + \mu_1 + a_1 D + V - \ln \pi_2 - \mu_2 - a_2 D - V \\ &= c + D = \sigma(c_* + D_*). \end{aligned} \tag{7b}$$

We now derive the expressions for the sectoral moments of log earnings:

(8)
$$E[\ln W_1|\ln W_1>\ln W_2] = \ln \pi_1 + \mu_1 + E[U_1|D>-c] = \ln \pi_1 + \mu_1 + a_1 E[D|D>-c] + E[V|D>-c]$$

and

(9)
$$\operatorname{Var}[\ln W_1|\ln W_1>\ln W_2] = \operatorname{Var}[U_1|D>-c] = a_1^2 \operatorname{Var}[D|D>-c] + \operatorname{Var}[V|D>-c] + 2a_1 \operatorname{Cov}[D,V|D>-c].$$

If D and V *are independent*,[3] then $E[V|D>-c]=0$, $\operatorname{Var}[V|D>-c]=\operatorname{Var}[V]$ and $\operatorname{Cov}[D,V|D>-c]=0$ so

一二三阶矩

(10)
$$E[\ln W_1|\ln W_1>\ln W_2] = \ln \pi_1 + \mu_1 + a_1 E[D|D>-c] = \ln \pi_1 + \mu_1 + a_1\sigma E[D_*|D_*>-c_*],$$

(11)
$$\operatorname{Var}[\ln W_1|\ln W_1>\ln W_2] = a_1^2 \operatorname{Var}[D|D>-c] + \operatorname{Var}[V] = \sigma_{11}(\rho^2 \operatorname{Var}[D_*|D_*>-c_*] + (1-\rho^2))$$

and

(12)
$$E[(\ln W_1 - E[\ln W_1|\ln W_1>\ln W_2])^3|\ln W_1>\ln W_2] = a_1^3 E[(D-E[D|D>-c])^3|D>-c] + E[V^3] = a_1^3\sigma^3 E[(D_*-E[D_*|D_*>-c_*])^3|D_*>-c_*] + E[V^3].$$

Likewise

(13)
$$E[\ln S_1|\ln W_1>\ln W_2] = \mu_1 + a_1\sigma E[D_*|D_*>-c_*],$$

(14)
$$\operatorname{Var}[\ln S_1|\ln W_1>\ln W_2] = \sigma_{11}(\rho^2 \operatorname{Var}[D_*|>-c_*] + (1-\rho^2)),$$

and

(15)
$$E[(\ln S_1 - E[\ln S_1|\ln W_1>\ln W_2])^3|\ln W_1>\ln W_2] = a_1^3\sigma^3 E[(D_*-E[D_*|D_*>-c_*])^3|D_*>-c_*] + E[V^3].$$

It is clear that $E[D|D>-c] \geqslant E[D]=0$, where the inequality is strict unless c is to the left of all points in the support of D. If $\sigma_{11}>\sigma_{12}$, the mean of log skills selected into sector one exceeds the population mean μ_1. If $\sigma_{11}<\sigma_{12}$, the opposite occurs. Since $\sigma^2>0$, $\sigma_{11}<\sigma_{12}$ implies $\sigma_{22}>\sigma_{12}$. These inequalities play a critical role in the Roy model since they determine the sign of a_1. For lack of a better term, we designate $\sigma_{11}>\sigma_{12}$ the "standard" case because it gives the intuitively plausible result that people who work in sector one have

[3] Notice that D and V are uncorrelated by construction.

higher than population mean sector one specific skills. We designate $\sigma_{11} < \sigma_{12}$ the "nonstandard" case and have just established that it can occur in at most one sector of a two sector economy.

From equations (10)–(15), it is clear that in order to answer questions about the effects of changes in π_i, we must investigate the conditional distribution of D given $D > c$. In Roy's original paper it was assumed that log skills are normally distributed which implies that D and V are independent. In the following subsection, we will study a log concave Roy model that is more general than the log normal model, but which preserves many features of Roy's original model. Most of the results about the sectoral income distribution presented in subsection 1.1, appear elsewhere for the special case of normally distributed log skills.[4] In subsection 1.2 we present new results about aggregate income inequality under the assumption of normally distributed log skills. In subsection 1.3 we show that many of the conclusions made for the log concave Roy model in subsection 1.1 do not generalize to arbitrary distributions.

1.1. *Log Concavity and the Roy Model*

In this subsection we investigate conditions on D that will allow us to characterize conditional moments (10), (11), (13), and (14). One assumption that will allow us to characterize the truncated distribution of D is that D is *log concave*. We give the definition for vector valued random variables.

DEFINITION 1: A *log concave random variable* $\boldsymbol{X}$ is one for which the density f satisfies the condition that $f(\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2) \geqslant [f(\boldsymbol{x}_1)]^{\lambda}[f(\boldsymbol{x}_2)]^{1-\lambda}$, $0 \leqslant \lambda \leqslant 1$, for $\boldsymbol{x}_1, \boldsymbol{x}_2$ in the support of $\boldsymbol{X}$.

A density is strictly log concave if the first inequality is strict for $0 < \lambda < 1$. The class of log concave densities include normal densities, uniform densities, Beta densities, and extreme value densities. See Pratt (1981) for a more extensive list. Prekopa (1973) establishes that the marginals of log concave densities are log concave and that convolutions of log concave random variables are log concave.

Brascamp and Lieb (1975) establish that if the density is log concave then so is the distribution function $F(\boldsymbol{x})$ and the right tail probability or survivor function $(1 - F(\boldsymbol{x})) = S(\boldsymbol{x})$. It is important to remark that if scalar random D is log concave, then so is $-D$. This means that we can talk about a log concave Roy model without having to worry about which sector has been labeled sector one.

The following facts about log concave random variables allow us to generalize many of the results that are known for the log normal Roy model to the class of models for which D is log concave. Notice that we do not assume that log skills are log concave—just their difference.

[4] See Heckman and Sedlacek (1985).

PROPOSITION 1: *If D is a log concave random variable, then*

$$0 \leqslant \frac{\partial E[D \mid D > d]}{\partial d} \leqslant 1, \tag{16a}$$

$$0 \leqslant \frac{\partial E[D \mid D \leqslant d]}{\partial d} \leqslant 1, \tag{16b}$$

and

$$\frac{\partial \operatorname{Var}[D \mid D > d]}{\partial d} \leqslant 0, \tag{17a}$$

$$\frac{\partial \operatorname{Var}(D \mid D \leqslant d)}{\partial d} \geqslant 0. \tag{17b}$$

The inequalities are strict (except possibly at the boundary of the support) if D is strictly log concave.

PROOF: Appendix B. *Q.E.D.*

COROLLARY 1: *If D is log concave, then* $\operatorname{Var}[D \mid D \geqslant d] \leqslant \sigma^2$.

PROOF: Follows from (17), and $\operatorname{Var}(D) = \sigma^2$. *Q.E.D.*

Differentiating (10) and (13), we can now find the effect of an increase of π_i on the mean log skill and earnings in sector one. As noted in Appendix B, log concavity implies that these derivatives exist almost everywhere:

$$\frac{\partial E[\ln W_1 \mid \ln W_1 > \ln W_2]}{\partial \ln \pi_i} = \begin{cases} 1 - a_1 \left. \dfrac{\partial E[D \mid D > d]}{\partial d} \right|_{d=-c} & \text{if } i = 1, \\ a_1 \left. \dfrac{\partial E[D \mid D > d]}{\partial d} \right|_{d=-c} & \text{if } i = 2, \end{cases} \tag{18}$$

$$\frac{\partial E[\ln S_1 \mid \ln W_1 > \ln W_2]}{\partial \ln \pi_i} = \begin{cases} -a_1 \left. \dfrac{\partial E[D \mid D > d]}{\partial d} \right|_{d=-c} & \text{if } i = 1, \\ a_1 \left. \dfrac{\partial E[D \mid D > d]}{\partial d} \right|_{d=-c} & \text{if } i = 2. \end{cases} \tag{19}$$

In interpreting (18) and (19), there are three cases to consider. If both sectors are standard, then $0 < a_1 < 1$. If sector one is nonstandard, then $a_1 \leqslant 0$. If sector two is nonstandard,[5] $a_1 \geqslant 1$. If both sectors are standard, then mean log earnings increase in response to an increase in either price. Put differently, an increase in π_1 increases the mean log earnings in both sectors. The mean skill

[5] Recall that at most one of the two sectors can be nonstandard.

level decreases in sector one, and increases in sector two. The higher π_1 attracts into sector one previous sector two workers whose comparative advantage was sector two at the old price, and leaves behind in sector two workers with comparative advantage there. Notice that if both sectors are standard, all the derivatives in (18) and (19) are less than 1 in absolute value.

If sector two is nonstandard ($a_1 \geqslant 1$), then an increase in π_1 may decrease the mean log income in sector one. All other derivatives have the same sign as if both sectors were standard. If sector one is nonstandard ($a_1 \leqslant 0$), then an increase in π_1 will increase the skill level in both sectors and the earnings in sector one will go up proportionally more than the increase in π_1.

The impact of comparative advantage on earnings inequality within sectors is easily derived. Equations (11), (14), and Corollary 1 imply that unless $\rho = 1$, selection reduces sectoral earnings inequality over the case of random selection of persons into the sector. As π_1 increases, c increases, and $\mathrm{Var}[D|D > -c]$ increases. So inequality in sector one increases with π_1. By the same argument, inequality in sector one decreases as π_2 increases. To our knowledge, there is no general result about the behavior of the third central moment of a truncated log concave random variable.

If we strengthen our assumption about the joint distribution of $\ln S_1$ and $\ln S_2$ and assume that they are jointly log concave random variables (not just their difference), then the aggregate income distribution is log concave as well. This follows easily from a theorem of Prekopa (1971).

THEOREM 1: *If* $\ln S_1$, $\ln S_2$ *are joint log concave random variables with log concave densities, the aggregate log income distribution*

$$G(\ln w) = P_1 G_1(\ln w) + P_2 G_2(\ln w)$$

is log concave.

PROOF: If ($\ln S_1$, $\ln S_2$) is log concave with density $f(\ln s_1, \ln s_2)$, then so is ($\ln W_1$, $\ln W_2$) because translations of log concave random variables are log concave (Prekopa (1973, Theorem 7)). By the Brascamp-Lieb (1975) theorem, the distribution function $F(\ln w_1, \ln w_2)$ is log concave. The observed wage is $\ln W = \max(\ln W_1, \ln W_2)$ with cdf $F(\ln w, \ln w)$ which is obviously log concave if the distribution of ($\ln W_1$, $\ln W_2$) is log concave. *Q.E.D.*

Karlin and Reinott (1983, Theorem 23) first proved this result. For another proof, see Caplin and Nalebuff (1989). Our proof is new. The theorem is also true for a Roy economy with log concave skills with more than two sectors. Note that the *density* of $\ln W$ need not be log concave for the distribution to be log concave. (See Pratt (1981).) Moreover the survivor function or right tail probability $\Pr(\ln W > \ln w)$ need not be log concave in $\ln w$. It follows from our discussion in Appendix B that because $G(\ln w)$ is log concave,

$$0 \leqslant \frac{\partial E(\ln W | \ln W \leqslant d)}{\partial d} \leqslant 1$$

and

$$\frac{\partial \operatorname{Var}(\ln W|\ln W \leqslant d)}{\partial d} \geqslant 0.$$

The survivor function $1 - G(\ln w)$ need not be log concave so that the other relationships in (16) and (17) need not hold. In the strict log concave case (where the inequalities are strict) as we start from the bottom and progressively raise the inclusion level of earnings, the variance of log earnings increases and mean of log earnings increases at a rate less than the inclusion level.

1.2. *Consequences of Log Normality*

The properties of the Roy model depend on the conditional distribution of D given $D > -c$. In Roy's original paper, it was assumed that the log skills were jointly normally distributed. Thus log skills are log concave random variables with log concave densities and *all* the results of Section 1.1 apply to the Roy model including the result that the distribution of $\ln W$ is log concave. With the normality assumption, it is possible to produce exact expressions for the truncated distribution of D, and it is possible to characterize the aggregate earnings distribution in a Roy economy.

The first thing to notice about the log normal Roy model is that if (U_1, U_2) is normally distributed, then so is (V, D). As V and D are uncorrelated by construction, V and D are independent. The normality assumption allows us to derive exact expressions for the moments of the truncated normal distribution. For completeness, we summarize key properties of these moments.

Let Z be a standard normal random variable and let $\lambda(d) \overset{\text{def}}{\equiv} E[Z|Z > d]$; then for $d \in (-\infty, \infty)$, we prove the following results in Appendix A:

$$\text{(R-1)} \qquad \lambda(d) = \frac{\frac{1}{\sqrt{2\pi}} \exp\{-d^2/2\}}{\Phi(-d)} > \max\{0, d\},$$

$$\text{(R-2)} \qquad 0 < \frac{\partial \lambda(d)}{\partial d} = \lambda'(d) = \lambda(d)(\lambda(d) - d) < 1,$$

$$\text{(R-3)} \qquad \frac{\partial^2 \lambda(d)}{\partial d^2} > 0,$$

$$\text{(R-4)} \qquad 0 < \operatorname{Var}[Z|Z > d] = 1 + \lambda(d)d - \lambda^2(d) < 1,$$

$$\text{(R-5)} \qquad \frac{\partial \operatorname{Var}[Z|Z > d]}{\partial d} < 0,$$

$$\text{(R-6)} \qquad E\left[(Z - \lambda(d))^3 \middle| Z > d\right] = \lambda(d)\left(2\lambda^2(d) - 3d\lambda(d) + d^2 - 1\right) = \frac{\partial^2 \lambda(d)}{\partial d^2},$$

$$\text{(R-7)} \qquad E[Z|Z > d] \geqslant \operatorname{mode}[Z|Z > d].$$

Furthermore,

$$\text{(R-8)} \qquad \lim_{d \to -\infty} \lambda(d) = 0, \qquad \lim_{d \to \infty} \lambda(d) = \infty,$$

$$\text{(R-9)} \qquad \lim_{d \to -\infty} \frac{\partial \lambda(d)}{\partial d} = 0, \qquad \lim_{d \to \infty} \frac{\partial \lambda(d)}{\partial d} = 1,$$

$$\text{(R-10)} \qquad \lim_{d \to -\infty} \mathrm{Var}[Z \mid Z > d] = 1, \qquad \lim_{d \to \infty} \mathrm{Var}[Z \mid Z > d] = 0.$$

Results (R-2), (R-4), and (R-5) are implications of log concavity. (R-7) is an implication of symmetry and log concavity. (R-1) and (R-3) are consequences of normality. The left hand side limits of (R-8) and (R-10) are true for any distribution. So is the right hand limit of (R-8) provided that the support of Z is not bounded on the right. The right hand limits of (R-9) and (R-10) are consequences of normality.

Besides being of interest in their own right, results (R-1) to (R-10) allow us to prove that self selection reduces earnings inequality as measured by the variance of logarithms compared to what it would be if individuals were randomly assigned to sectors of the economy.[6] More precisely, the following theorem can be proved.

THEOREM 2: *For a log normal Roy economy, any random assignment of persons to sectors with the same proportion of persons in each sector as in the Roy economy has higher variance of log earnings provided the proportions lie strictly in the unit interval. This is true whether or not skill prices in the two economies are the same.*

讲自选择相较于random assignment降低了收入差距。

PROOF: See Appendix C. *Q.E.D.*

In a Roy economy market forces operate to reduce income inequality compared to the random assignment economy. This result implies that in a log normal Roy economy, drafting people into the military as compared to recruiting the same proportion of the population into the army will result in greater income inequality. The theorem is true irrespective of the differences in the skill prices that would arise in the two different allocations. Note further that Theorem 2 remains true if in decomposition (6) D is normal but V (assumed independent of D) is a nonnormal distribution with finite second moments. Thus the theorem is true for a somewhat more general class of models.

The assumption of log normality implies that D is normal. Using this, (R-3), and (R-6), we can infer from (12) that the sectoral log earnings distributions are skewed with the direction of the skew depending on the sign of a_i. In the standard case ($a_1 > 0$) the skill level in sector one is higher than it would be with

[6] The variance of the logarithm of earnings is a widely used measure of income inequality. Atkinson (1970) demonstrates that it violates Dalton's Principle of Transfers, at least for certain transfers. Creedy (1985) questions the practical relevance of Atkinson's criticism of the variance of the logarithm of earnings as a welfare measure. He notes that for empirically relevant earnings distributions less than one percent of all possible transfers from rich to poor result in an increase in the variance of log earnings and thus violate the Dalton Principle.

random assignment. The best sector one people tend to work in sector one, exaggerating the right tail and leaving the left tail evacuated. In the nonstandard case the opposite is the case. The best sector one people tend to work in sector two, exaggerating the left tail and leaving the right tail evacuated. However, the aggregate log earnings distribution is right skewed even though one sectoral distribution may be left skewed.

THEOREM 3: *In a log normal skill Roy economy, aggregate log earnings distributions are right skewed as long as some positive fraction of the population works in each sector.*

总体对数收入分布右偏

PROOF: The proof is due to K. Choi and is available on request from the authors. *Q.E.D.*

Again we note that the theorem is true even if V (in (6)) is nonnormal with zero or positive third moment provided that D is normal. We note further that the density of $\ln W$ may be multimodal and that the variance of $\ln W$ need not be monotonic in π_i. (These results are established in a supplement available on request from the authors.) The multimodality of the density does not contradict the log concavity of the distribution function.

1.3. *Nonrobustness of the Roy Model to Non-Log Concavity*

It is natural to ask whether the results obtained for the (large) class of log concave models can be generalized to all distributions of (U_1, U_2). As might be expected, the answer is "no." The same apparatus that was used to derive the results for the log concave model can be used to derive counterexamples to these results. For log convex random variables, inequalities 16(a) and 17(a) in Proposition 1 and the Corollary of the Proposition are reversed.

DEFINITION 2: A *log convex random variable* is one for which the density of f satisfies the condition that $f(\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2) \leqslant (f(\boldsymbol{x}_1))^{\lambda}(f(\boldsymbol{x}_2))^{1-\lambda}$, $0 \leqslant \lambda \leqslant 1$, for $\boldsymbol{x}_1, \boldsymbol{x}_2$ in the support of $\boldsymbol{X}$.

The class of log convex densities include Pareto distributions with finite means and variances and gamma densities with a coefficient of variation greater than one. Corresponding to Proposition 1 we have Proposition 2 for nonnegative random variables.

PROPOSITION 2: *If D is a log convex random variable and $D \geqslant 0$, with support in* $[0, \infty)$,

$$\text{(20)} \qquad \frac{\partial E[D \mid D > d]}{\partial d} \geqslant 1$$

and

$$(21) \qquad \frac{\partial \operatorname{Var}[D \mid D > d]}{\partial d} \geqslant 0.$$

PROOF: Appendix B. *Q.E.D.*

Propositions 16(b) and 17(b) have no counterpart in the log convex case for reasons discussed in Appendix B.

COROLLARY 2: *If D is log convex with support in $[0, \infty)$, then $Var[D \mid D > d] \geqslant \sigma^2$.*

PROOF: Follows from (21) and $\operatorname{Var}[D] = \sigma^2$. *Q.E.D.*

Proposition 2 implies that in a log convex Roy model, the earnings level in sector one may decrease in response to an increase in π_1 even if sector one is standard. The previously established result for the log concave Roy economy that the variances of sectoral log earnings and log skills are less (strictly, not greater) than the population variances is reversed in a log convex Roy economy. This follows from Corollary 2 and equations (11) and (14).[7]

Note further that a result comparable to Theorem 1 for log convex random variables can be proved if the *distribution* of $(\ln W_1, \ln W_2)$ is log convex. By the logic of Theorem 1, $\ln W = \max(\ln W_1, \ln W_2)$ has cdf $F(\ln w, \ln w)$ which is obviously log convex.

2. THE IDENTIFIABILITY OF THE ROY MODEL AND ITS NORMAL EXTENSIONS

We next consider conditions that must be satisfied in order to identify the population distribution of skills from earnings data and data on sectors chosen by individuals. The problem of recovering skill distributions from earnings distributions is nontrivial because individuals use only one skill at any time and observed earnings distributions within sectors are distorted representations of population skill distributions by virtue of self-selection decisions of economic

[7] Two propositions are true for any two sector model. First, as π_i increases so does P_i. (Strictly, as π_i increases P_i does not decrease.) Second, the nonstandard case can occur in at most one sector, i.e. if $E(\ln W_i \mid \ln W_i > \ln W_j) < E(\ln W_i)$, then $E(\ln W_j \mid \ln W_i \neq \ln W_j) > E(\ln W_j)$, $i \neq j$. The first result is trivial. The second result is established by making an argument by contradiction. Assume that (a) $E(\ln W_i \mid \ln W_i > \ln W_j) \leqslant E(\ln W_i)$ and (b) $E(\ln W_j \mid \ln W_i \leqslant \ln W_j) \leqslant E(\ln W_j)$ where the inequality is strict in one of the two cases. With no loss of generality we may assume that $E(\ln W_i) \geqslant E(\ln W_j)$, $i \neq j$. Since $E(\ln W_i - \ln W_j \mid \ln W_i \leqslant \ln W_j) \leqslant 0$, we may use (b) to establish that $E(\ln W_i \mid \ln W_i < \ln W_j) \leqslant E(\ln W_j \mid \ln W_i \leqslant \ln W_j) \leqslant E(\ln W_j) \leqslant E(\ln W_i)$, $i \neq j$. But $E(\ln W_i) = E(\ln W_i \mid \ln W_i > \ln W_j) P_i + E(\ln W_i \mid \ln W_j \leqslant \ln W_i)(1 - P_i) < E(\ln W_i)$, $i \neq j$, since both expectations are less than or equal to $E(\ln W_i)$ and one is strictly less than $E(\ln W_i)$. This forces a contradiction and proves the claim. A similar proposition can be proved for the sectoral means of log skills. Note that the second claim was already proved for the log concave model discussed in the text. The result established here is for *any* two sector model with finite means.

agents. Unless the underlying skill distributions can be identified many of the theoretical distinctions derived from the Roy model have no empirical content.

Identification is a necessary first step toward estimation of a model. Identification theorems are about unobserved population distributions. In general, additional assumptions must be invoked to guarantee the existence of consistent estimators based on sample data. In this paper we only explore the necessary first step. We leave the development of consistent estimators as a task for the future.

We establish that the log normal skills Roy model is identified using a single cross-section of data on earnings and sectoral choices of agents. Thus it is possible to identify $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ of Section 1 (skill prices are normalized to unity). We also establish that it is not necessary to know sectoral choice decisions by agents in order to recover the distribution of skills. We further establish that it is possible to recover the distribution of skills (subject to a normalization) when earnings in one sector are not observed as occurs in the housewife version of the Roy model in which nonmarket output is not observed.

These results are not robust when examined in the context of a general nonnormal model of skill distribution. Even with the knowledge of sectoral choices of agents, it is not possible from a cross-section of data on earnings to distinguish a model in which skills are highly correlated (the Willis-Rosen (1979) "one factor" model) from a model in which skills are independent or even negatively correlated. These results highlight the strong implications of a normality assumption for log skills and the lack of empirical content of the general nonnormal Roy model.

These negative conclusions can be reversed if the observing economist has access to data on earnings distributions from markets with the same distributions of skills but different relative skill prices, due, say, to variation in demand for the products produced by the skills. With sufficient price variation, it is possible to identify the population skill distribution even if the particular sector chosen by an agent is not known and even if earnings in one sector are not observed. Access to multimarket data solves the identification problem in the general nonnormal Roy model, and gives empirical content to the framework. We also establish how the introduction of regressors in a controlled way can secure identification of the Roy model. This analysis has implications for the identifiability of the closely related competing risks duration model.

2.1. *Identifiability of the Log Normal Roy Model from Cross-Section Data*

Provided that there are nondegenerate regressors Z that shift the mean of log skill distributions:

$$\mu_i = Z\beta_i, \quad i = 1, 2,$$

where β is a vector of parameters, and provided that sectoral choices are known, it is possible to identify $\boldsymbol{\Sigma}$, β_1, and β_2 from a single cross section using standard sample selection methods (Heckman (1976)). Skill prices are absorbed into model intercepts.[8]

Using population moments, Heckman and Sedlacek (1981) establish that it is not necessary for there to be any regressors in the model in order to identify $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ provided that the agent's choices are known. Basu and Ghosh (1978) establish the same proposition which for completeness we record as a theorem.

THEOREM 4: *Under the conditions postulated for the log normal Roy model,* $\boldsymbol{\mu}$ *and* $\boldsymbol{\Sigma}$ *can be identified from data on wages paid in each sector and sectoral choices.*

PROOF: See Heckman and Sedlacek (1981) or Basu and Ghosh (1978). *Q.E.D.*

Somewhat more surprisingly, it is possible to identify $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, except for their subscripts, from the knowledge of the aggregate earnings distribution. It is neither necessary to know the sectoral earnings distributions nor the proportion of the population in each sector. Thus it is possible to identify the elements of (μ_1, μ_2) and $(\sigma_{11}, \sigma_{22})$ but it is not possible to assign the means or variances to a particular sector. It *is* possible to associate the mean of an unidentified sector with its variance. The covariance σ_{12} is uniquely identified.

THEOREM 5: *Under the conditions postulated for the log normal Roy model,* $\boldsymbol{\mu}$ *and* $\boldsymbol{\Sigma}$ *can be identified, except for their subscripts, from knowledge of the aggregate earnings distribution.*

PROOF: See Basu and Ghosh (1978). *Q.E.D.*

In the analysis of female labor supply, it is often assumed that one of the sectors is home production. In this case only the proportion working in each sector and the sectoral earnings distribution in one sector are observed. Nonetheless, it is still possible to identify certain combinations of the parameters. More precisely, the following theorem can be proved.

THEOREM 6: *If only the earnings density in sector one,* $f(\ln w_1 | \ln W_1 > \ln W_2)$ *and* $\Pr(\ln W_1 > \ln W_2)$ *are known in the log normal Roy model, it is possible to identify* μ_1 *and* σ_{11}. *It is also possible to identify* $(\mu_1 - \mu_2)/\sigma$ *and* ρ, *where* σ *and* ρ *are as defined in Section* 1. *Without further restrictions it is not possible to identify* μ_2 *or* σ.

PROOF: See Appendix C. *Q.E.D.*

If $\mu_2 = 0$ is adopted as a normalization, σ_{12} and σ_{22} are identified. If $\sigma_{12} = 0$ is adopted as a normalization, μ_2 and σ_{22} are identified. If $\sigma_{22} = 1$ is adopted as

[8] This means that we can normalize $\ln \pi_1 = 0$ and $\ln \pi_2 = 0$.

a normalization, μ_2 and σ_{12} are identified. The inability to observe wage distributions in both sectors impairs the identifiability of parameters.

Theorems 4, 5, and 6 demonstrate the power of the log normality assumption. For the original Roy model it is possible to estimate the parameters of skill distributions without specifying functional forms of equations relating means to regressors or, indeed, without any regressors at all. Given a normality assumption, the model has empirical content.

2.2. *The Nonidentifiability of a General Nonnormal Roy Model in a Single Cross Section*

The conclusions of the previous subsection are fundamentally altered when Roy's log normality assumption is dropped and a general skill distribution is assumed. Any cross-section wage distribution can be rationalized by a model with one skill, two independent skills, or two negatively correlated skills. In the general case, notions of skill hierarchy have no empirical content when applied to single cross-sections of data.

We first establish that any cross-section of data on wages with sectoral choices observed can always be rationalized by an independent skill model. As a convention, we normalize skill prices to unity, so wages and skills are identical.

THEOREM 7: *It is possible to rationalize sectoral wage data in a single cross-section by a two skill model with independence. More precisely, it is possible to rationalize data on* $f(s_1|S_1 > S_2)$, $f(s_2|S_2 > S_1)$, *and* $P(S_1 > S_2)$ *by an independent skill model* $f(s_1, s_2) = f_1(s_1)f_2(s_2)$.

PROOF: The result is a straightforward modification of the proof of the nonidentifiability of the competing risks model by Cox (1962) and Tsiatis (1975). Replace "min" with "max" in their theorems. *Q.E.D.*

Let Y be the observed wage and let R be an indicator giving the sector associated with that wage. Theorem 7 informs us that any distribution of (Y, R) can be explained by a Roy model with independent skills. On the other hand, imagine that for each (Y, R) we define

$$(S_1, S_2) = \begin{cases} (Y, Y-\varepsilon), & \text{if } R = 1, \\ (Y-\varepsilon, Y), & \text{if } R = 2. \end{cases}$$

The correlation (provided that it exists) between S_1 and S_2 constructed in this manner will depend on ε, but it can be made arbitrarily close[9] to 1 by making ε close to 0. We state this as a theorem.

[9] The reason why we talk about making ε close to 0 rather than equal to 0, is that at $\varepsilon = 0$, the worker is indifferent between the two sectors with probability 1.

THEOREM 8: *For any* $c < 1$, *it is always possible to rationalize sectoral wage data in a single cross-section by a two-skill model with correlation greater than* c, *provided that* $\mathrm{Var}[\max\{S_1, S_2\}]$ *exists.*

PROOF: Follows from the discussion preceding the theorem. *Q.E.D.*

We next demonstrate that it is possible to rationalize any finite cross-section data on wages by a hierarchical model. Letting $S_1^{(i)}$ be skill one of person i and $S_2^{(i)}$ his amount of skill two, a hierarchical model is one for which $S_1^{(i)} > S_1^{(j)}$ for $i \neq j$ implies that $S_2^{(i)} > S_2^{(j)}$, i.e. if i is better in skill one, he is also better in skill two, and all skill vectors are ordered.

Imagine that we have N observations on (Y, R), (y_i, r_i) for $i = 1, \ldots, N$. Array the observations in increasing order of y. Now suppose that the true realizations of $(S_1^{(i)}, S_2^{(i)})$ had been

$$\left(s_1^{(i)}, s_2^{(i)}\right) = \begin{cases} \left(y_i, \dfrac{y_i + y_{i-1}}{2}\right), & \text{if } r_i = 1, \\ \left(\dfrac{y_i + y_{i-1}}{2}, y_i\right), & \text{if } r_i = 2, \end{cases}$$

where $y_0 = 0$. These constructed data are hierarchical. These values of $(S_1^{(i)}, S_2^{(i)})$, are consistent with observed (y_i, d_i). We stress that the data set is finite. In the limit, the constructed model would have $(S_1^{(i)} = S_2^{(i)})$; everybody would be indifferent between sectors. So in the limit the constructed model would be one that is ruled out a priori. Note that this construction *induces* dependence across observations and strictly speaking violates the independence assumption invoked in the rest of the paper. If y_i is *measured* in fixed units of a minimum size, it is always possible to produce an independent hierarchical model:

$$\left(s_1^{(i)}, s_2^{(i)}\right) = \begin{cases} (y_i, y_i - \varepsilon) & \text{if } r_i = 1, \\ (y_i - \varepsilon, y) & \text{if } r_i = 2, \end{cases}$$

provided that ε is smaller than the minimal unit of observation used *by the econometrician*. This is a nearly perfectly positively correlated skill model, as above. Using similar methods, it is possible to construct a nearly perfectly negatively correlated skill model.

2.3. *Identification from Multi-market Data in a General Nonnormal Roy Model*

If there is access to data on earnings and sectoral choices from economies with different skill prices but with the same population skill distributions, the pessimistic conclusions of subsection 2.2 are reversed. With sufficient variation in skill prices it is possible to recover the underlying population skill distributions even if the sectors chosen by persons are unknown to the observing economist and even if the earnings of one sector are never observed as in the housewife version of the Roy model. Access to panel data greatly reduces the required amount of price variation. Panel data from two different price situations suffices to identify population skill distributions in a subset of the sample space even when the sectoral choice decisions of agents are unknown. Access to regressors that affect the location of the log-skill distribution substitutes for price variation and secures identification in a single cross-section. This subsection verifies these claims.

The intuition as to why price variation can break the nonidentifiability result in Theorem 7 is clear. Suppose that we have cross-sectional data for two different price settings. Theorem 7 can be applied to rationalize each cross-section by an independent skill distribution. Unless skills are truly independent, different functional forms for skill distributions will, in general, be required to rationalize different cross-sections.[10] This also suggests a test of independence. For each cross-section, we may apply Theorem 7 and generate independent skill distributions. If the constructed skill distributions are not equal we can reject the hypothesis of independence.

The next theorem shows that this intuition is correct. If there is enough price variation, then the joint distribution of (S_1, S_2) is identified.

THEOREM 9: *Let S_1 and S_2 be positive random variables with distribution function $F(s_1, s_2)$. If we only observe $Z = \max\{S_1, \pi_2 S_2\}$ and π_2 takes all possible values in the interval $(0, \infty)$, then F is identifiable.*

PROOF: By assumption we know $\Pr(\max\{S_1, \pi_2 S_2\} \leqslant x)$ for all x and π_2, but

$$\begin{aligned}\{\max\{S_1, \pi_2 S_2\} \leqslant x\} &= \{S_1 \leqslant x, \pi_2 S_2 \leqslant S_1\} \cup \{\pi_2 S_2 \leqslant x, S_1 \leqslant \pi_2 S_2\} \\ &= \{S_1 \leqslant x, S_2 \leqslant x/\pi_2\},\end{aligned}$$

so for any $s_1, s_2 > 0$ we have (setting $x = s_1$ and $\pi_2 = s_1/s_2$),

$$\begin{aligned}F(s_1, s_2) = \Pr(S_1 \leqslant s_1, S_2 \leqslant s_2) &= \Pr(S_1 \leqslant s_1, S_2 \leqslant s_1/\pi_2) \\ &= \Pr(\max\{S_1, \pi_2 S_2\} \leqslant x),\end{aligned}$$

which completes the proof. *Q.E.D.*

The intuition underlying this example is simple. For any π_2, we determine $F(s_1, s_2)$ along the ray $s_1 = \pi_2 s_2$. As π_2 varies over the interval $(0, \infty)$, we determine F over all rays. Note that if additional functional form conditions are imposed on $F(s_1, s_2)$ which permit it to be determined everywhere from knowledge in a region (e.g., in the cone between $s_1 = \pi_2 s_2$ and $s_1 = \pi_2' s_2$ for $\pi_2 \neq \pi_2'$), it is not necessary for π_2 to range the interval $(0, \infty)$. One such condition is that F is real analytic in its arguments so that it can be continued outside of its region of determination. Note further that Theorem 9 also applies

[10] This is also the intuition behind the reason why access to covariates or regressors breaks the nonidentification theorem in competing risks models in duration analysis (see Heckman and Honoré (1989)).

to the case in which sectoral earnings distributions are observed, as the latter case contains more information than is necessary for Theorem 9. The theorem can be proved for economies with more than two sectors (available on request from the authors).

Similar reasoning can be applied to identify population skill distributions when earnings in one sector are never observed as in the case of nonmarket earnings of housewives. This proposition is established in Theorem 10. It is a nonparametric analogue to Theorem 5.

THEOREM 10: *If we observe the distribution of Z given by*

$$Z = \begin{cases} \pi_2 S_2 & \text{if } S_1 < \pi_2 S_2, \\ 0 & \text{if } S_1 \geqslant \pi_2 S_2, \end{cases}$$

and π_2 *traverses the interval* $(0, \infty)$, *then* $F(s_1, s_2)$ *is identified from multimarket data on aggregate earnings.*

PROOF: Let s_1, s_2 be given. We will then show that we can find $F(s_1, s_2)$.

Let $\varepsilon > 0$ be given. For given π_2, we know the probability of events of the type $\{S_1 < \pi_2 S_2 \leqslant x\}$ for all $x \in (0, \infty)$. This means that we know the probability of the event given by OAB in Figure 1. By the same argument we also know the probability of the event given by the set OCD. We therefore know the probability of the difference $DCAB$.

By exactly the same reasoning we know the probability of the event $FGED$, and therefore of the event $FGECAB$. If we continue this process, we will

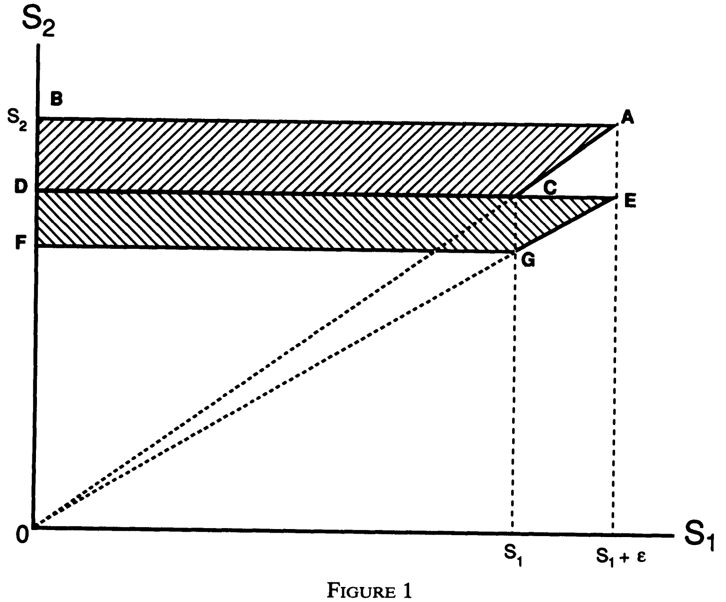

FIGURE 1

converge to a number μ satisfying

$$F(s_1, s_2) \leqslant \mu \leqslant F(s_1 + \varepsilon, s_2).$$

Do this for each ε, and take the limit as $\varepsilon \to 0$, and we obtain $F(s_1, s_2)$. *Q.E.D.*

This theorem can be proved for a general multiple sector economy in which the wage is not observed in one sector (available on request from the authors).

Thus far we have assumed access only to repeated cross-section data. Access to panel data on earnings that follow the same persons over time greatly facilitates identification if individual skills do not change over time.

THEOREM 11: *Suppose that we have panel data on aggregate earnings of individuals and that individual skills do not change over time. If we observe* $(Z, Z') = (\max\{S_1, \pi_2 S_2\}, \max\{S_1, \pi_2' S_2\})$ *for* $\pi_2 < \pi_2'$, *then we can identify* $F(s_1, s_2)$ *over the region* $\pi_2 s_2 \leqslant s_1 \leqslant \pi_2' s_2$.

PROOF: By hypothesis we know $\Pr(Z' \leqslant z', Z \leqslant z)$ for all z', z. Let $s_1, s_2 > 0$ be given such that $\pi_2 s_2 \leqslant s_1 \leqslant \pi_2' s_2$. Now over this region

$$\begin{aligned} F(s_1, s_2) &= \Pr(S_1 \leqslant s_1, S_2 \leqslant s_2) \\ &= \Pr(S_1 \leqslant s_1, S_1 \leqslant \pi_2' s_2, S_2 \leqslant s_2, S_2 \leqslant s_1/\pi_2) \\ &= \Pr(S_1 \leqslant s_1, \pi_2 S_2 \leqslant s_1, S_1 \leqslant \pi_2' s_2, \pi_2' S_2 \leqslant \pi_2' s_2) \\ &= \Pr(\max\{S_1, \pi_2 S_2\} \leqslant s_1, \max\{S_1, \pi_2' S_2\} \leqslant \pi_2' s_2) \end{aligned}$$

which, by hypothesis, is known. Hence $F(s_1, s_2)$ is identified for all $\pi_2 s_2 \leqslant s_1 \leqslant \pi_2' s_2$. *Q.E.D.*

In Theorem 9, it is necessary to observe the distribution of Z *for each value* of π_2 in the interval in order to determine F in the region $\pi_2 s_2 \leqslant s_1 \leqslant \pi_2' s_2$ with panel data. Theorem 11 instructs us that, with panel data, observing the same persons in different markets for skills we can acquire the same information knowing the distribution for Z under only two prices π_2 and π_2'. This theorem can be extended to a general multisectoral economy.

Another avenue for securing identification of general skill distributions from a single cross-section of data on aggregate earnings is to postulate regressors shifting the locations for the skill distributions. This approach to identification is more traditional and requires exclusion restrictions. Given the validity of the assumptions underlying this method, it is possible to recover skill distributions from a single cross-section of data. We record this in the following theorem.

THEOREM 12: *Let* $S_1 = g_1(X_1, X_0) + \varepsilon_1$ *and* $S_2 = g_2(X_2, X_0) + \varepsilon_2$ *where* $(\varepsilon_1, \varepsilon_2)$ *is independent of* (X_0, X_1, X_2). *Assume that* (a) $(\varepsilon_1, \varepsilon_2)$ *is continuously distributed with distribution function* G *and support equal to* R^2; (b) *support*$(g_1(X_1, x_0), g_2(X_2, x_0)) = R^2$ *for all* x_0 *in the support of* X_0; (c) *the marginal distributions of* ε_1 *and* ε_2 *both have medians equal to* 0. *Then* g_1, g_2, *and* G *are identified.*

PROOF: By assumption we know

(A) $\Pr(S_1 > S_2) = \Pr(g_1(x_1, x_0) + \varepsilon_1 > g_2(x_2, x_0) + \varepsilon_2)$,

(B) $\Pr(S_1 \leqslant y, S_1 > S_2) = \Pr(g_1(x_1, x_0) + \varepsilon_1 \leqslant y,$
$g_1(x_1, x_0) + \varepsilon_1 > g_2(x_2, x_0) + \varepsilon_2)$,

(C) $\Pr(S_2 \leqslant y, S_2 \geqslant S_1) = \Pr(g_2(x_2, x_0) + \varepsilon_2 \leqslant y,$
$g_2(x_2, x_0) + \varepsilon_2 > g_1(x_1, x_0) + \varepsilon_1)$,

for all (x_0, x_1, x_2) in the support of (X_0, X_1, X_2) and for all y.

Fix x_0. Let $\bar{x}_1$ and $\bar{x}_2$ be in the support of X_1 and X_2, respectively. From (A), we can then find

$$\begin{aligned} \{(x_1, x_2)&: \Pr(g_1(x_1, x_0) + \varepsilon_1 > g_2(x_2, x_0) + \varepsilon_2) \\ &= \Pr(g_1(\bar{x}_1, x_0) + \varepsilon_1 > g_2(\bar{x}_2, x_0) + \varepsilon_2) \\ &= \{(x_1, x_2): g_1(x_1, x_0) + l = g_2(x_2, x_0)\} \end{aligned}$$

for some unknown constant l.

For any point in that set we can use (B) to find

$$\Pr(g_1(x_1, x_0) + \varepsilon_1 \leqslant y, \varepsilon_1 > \varepsilon_2 + l)$$

for all y. This identifies $g_1(\cdot, x_0)$ except for an additive constant. In a similar way $g_2(\cdot, x_0)$ is identified (except for an additive constant). G is then identified by Theorem 9, except for the location. The location of G is determined by exploiting the fact that the medians of ε_1 and ε_2 are zero. Having determined the location of G, we can determine the additive constants in $g_1(\cdot, x_0)$ and $g_2(\cdot, x_0)$.

Since x_0 was arbitrary, this completes the proof. *Q.E.D.*

This theorem can easily be extended to a multisector economy. Changing "max" to "min" makes Theorem 12 applicable to the closely related competing risks problem. Related theorems for competing risks versions of the proportional hazard and accelerated hazard models are presented in Heckman and Honoré (1989). With regressors that affect the mean it is possible to substitute variation in individual characteristics at a point in time for variation in market price over time or over markets in order to recover population skill distributions.

3. SUMMARY

This paper examines the empirical content of Roy's model of self-selection and earnings. We summarize his model and derive new conclusions from it. Of special interest is our conclusion that the pursuit of comparative advantage reduces inequality in log earnings compared to what it would be if persons were randomly assigned to jobs.

Roy's model is based on the assumption that the logs of sector-specific skills are normally distributed in the population. We examine the robustness of conclusions drawn from the Roy model when log skills are not normal. Most conclusions are not robust. However, we derive a general class of nonnormal models for which the main conclusion of the Roy model remains valid. The general class of models assumes that skills can be decomposed into two components: a log concave random variable and an independent, additive component that can be freely specified. Selection depends on the log concave component but not on the other component. We briefly consider the implications of specifying both components and hence their sum to be log concave.

We also consider the empirical content of Roy's model. We present conditions that guarantee identification of his model from knowledge of population earnings densities. Identifiability is a necessary condition for consistent estimation of the Roy model on sample data. In this paper we only investigate the necessary first step, leaving development of consistent estimators for future research.

Under Roy's normality assumptions, it is possible to recover underlying skill distributions from a single cross-section of earnings even though only one skill is observed for any cross-section. Thus prior notions about skill hierarchies, selection from the bottom, and correlations among sector-specific skills can all be subject to rigorous empirical tests. No regressors or conventional exclusion restrictions are required to secure identification. It is not necessary to postulate functional forms connecting means or location parameters to covariates. Even if the sector chosen by agents is unknown, it is possible to identify the parameters of the Roy model up to their labels (i.e., we know the collection of sectoral means and variances but we do not know which element in the collection characterizes a particular sector). If the earnings in one sector are not observed, it is still possible to identify many parameters of interest.

These strong identification results vanish in a general nonnormal model. In the general case, any cross-section distribution of wages can be rationalized by a model with independent, positively correlated or negatively correlated skills. Hierarchical models of worker skills and inverse hierarchical models can be constructed that fit the data equally well. A general nonnormal Roy model has no empirical content when applied to a cross-section of wage data. The problem of nonidentification in the nonnormal Roy model is essentially the same as the problem of nonidentification in competing risks models in duration analysis.

Access to data from markets with different relative skill prices facilitates identification. With enough price variation, it is possible to recover underlying skill distributions from aggregate data. It is not necessary to know the sectors chosen by individuals. Skill distributions can be recovered even if earnings are not observed for persons employed in one sector. Panel data greatly facilitate identification.

We also demonstrate that if independent regressors are available and the assumption is made that they affect only the location of the log-skill distribution, it is possible to substitute cross-section variation in regressors for multimarket variation in skill prices and recover the underlying skill distributions.

We then demonstrate how access to regressors or variation in skill prices solves the identification problem in the Roy model and gives it empirical content. With sufficient variation in skill prices or regressors, notions about skill hierarchies and association among worker skills have empirical content. In a companion paper (Heckman and Honoré (1989)) we show how access to regressors can be used to identify the competing risks model.

All of the analysis in this paper and in the original Roy model assumes that agents make sectoral choices by comparing incomes in alternative sectors. It would be valuable to analyze models with more general decision rules that allow for utility maximization as in Lee (1978). It is plausible that nonwage dimensions of jobs also influence sectoral choice. See Heckman (1990) for a discussion of nonparametric identification of this model. It would also be valuable to consider multisectoral versions of the Roy model in which the skills are endogenous.

Department of Economics, Yale University, P.O. Box 1972 Yale Station, New Haven, CT 06520, U.S.A.,

and

Department of Economics, Northwestern University, 2003 Sheridan Rd., Evanston, IL 60208-2400, U.S.A.

Manuscript received December, 1987; final revision received December, 1989.

APPENDIX A

Proofs of Results (R-1) to (R-10)

The moment generating function for a truncated normal distribution with truncation point d is

$$\mathrm{mgf}(\beta)=e^{\beta^2/2}\frac{\int_{d-\beta}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}u^2\right)du}{\int_{d}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}u^2\right)du}.$$

The equality in (R-1) follows from

$$\lambda(d)=E[Z|Z>d]=\left.\frac{\partial\,\mathrm{mgf}}{\partial\beta}\right|_{\beta=0}.$$

The inequality is obvious. By direct calculation, $\lambda'(d)=\lambda(d)(\lambda(d)-d)$. Now note that

$$E\left[Z^2|Z>d\right]=\left.\frac{\partial^2\,\mathrm{mgf}}{\partial\beta^2}\right|_{\beta=0}=1+\lambda(d)d,$$

$$E\left[Z^3|Z>d\right]=\left.\frac{\partial^3\,\mathrm{mgf}}{\partial\beta^3}\right|_{\beta=0}=\lambda(d)(2+d^2).$$

Therefore

$$\mathrm{Var}[Z|Z>d] = 1 + \lambda(d)d - \lambda^2(d) = 1 - \frac{\partial\lambda(d)}{\partial d}.$$

As $\mathrm{Var}[Z|Z>d] > 0$ and $\lambda(d)(\lambda(d) - d) > 0$ by (R-1), this proves (R-2) and (R-4).

To prove (R-3) notice that $\mathrm{Var}[Z|Z>d] = 1 - (\partial\lambda(d)/\partial d)$, and therefore

$$\frac{\partial^2\lambda(d)}{\partial d^2} = -\frac{\partial\,\mathrm{Var}[Z|Z>d]}{\partial d} > 0$$

where the inequality follows from Proposition 1.

(R-5) follows from Proposition 1, whereas (R-6) follows by direct calculation from the expression for $E[(Z-\lambda(d))^3|Z>d]$. (R-7) is trivial.

(R-8) is obvious. The first part of (R-9) follows directly from l'Hospital's rule. (R-2) and (R-3) imply that $(\partial\lambda(d)/\partial d)$ is increasing and bounded by 1. Therefore $\lim_{d\to\infty}(\partial\lambda(d)/\partial d)$ exists and does not exceed 1. If $\lim_{d\to\infty}(\partial\lambda(d)/\partial d) < 1$, then $\lambda(d)$ would eventually be less than d, contradicting (R-1). (Details of this proof are available on request from the authors.) This proves the second part of (R-9). (R-9) and (R-4) imply (R-10).

APPENDIX B

Total Positivity, Log Convexity, and Log Concavity

We use results from the theory of total positivity (Karlin (1968)). A function $L(x,y)$ of two real variables is TP_2 if $L(x,y) \geqslant 0$ for all x, y and if for all $x_1 < x_2$, $y_1 < y_2$,

$$\begin{vmatrix} L(x_1,y_1) & L(x_1,y_2) \\ L(x_2,y_1) & L(x_2,y_2) \end{vmatrix} \geqslant 0.$$

By a theorem of Polya and Szego (1972, Vol. 1, Part II, Problem 68) as used by Karlin (1968, p. 17), if $K(x,q)$ is TP_2 and $L(q,y)$ is TP_2, then the *composition formula* states that

$$M(x,y) = \int_{q\in \underline{Q}} K(x,q)L(q,y)\,dq$$

($\underline{Q}$ is an interval that does not depend on x and y) is TP_2. Note that

$$R(x,q) \overset{\text{def}}{=} \begin{cases} 1, & x \geqslant q, \\ 0, & x < q, \end{cases}$$

and

$$R^*(x,q) \overset{\text{def}}{=} \begin{cases} 1, & x \leqslant q, \\ 0, & x > q, \end{cases}$$

are both TP_2.

Let J be a real valued function and define $L(q,y) = J(q-y)$. $L(q,y)$ TP_2 is then equivalent to J log concave. Likewise $L(q,y) = J(q+y)$ TP_2, $q \geqslant 0$, $y \geqslant 0$ is equivalent to J log convex. If $J(z)$ is log concave (convex), positive, and twice differentiable, then $J(q-y)$ TP_2 ($J(q+y)$ TP_2) is equivalent to

$$\frac{J''}{J} - \left(\frac{J'}{J}\right)^2 \lesseqgtr 0 \qquad \begin{matrix} (J(q-y)\ TP_2), \\ (J(q+y)\ TP_2), \end{matrix}$$

or for $J' \neq 0$

$$\frac{J''J}{(J')^2} \lesseqgtr 1 \qquad \begin{matrix} (J(q-y)\ TP_2), \\ (J(q+y)\ TP_2). \end{matrix}$$

Setting $K(x,q)=R(x,q)$ or $K(x,q)=R^*(x,q)$, we obtain that for $x=a$, a constant, the composition formula implies that

(B-1) $$M(a,y)=\int_{q\in \underline{Y}} R(a,q)J(q\pm y)\,dq=\int_{\{q|q\leqslant a,\, q\in \underline{Y}\}} J(q\pm y)\,dq,$$

(B-2) $$M^*(a,y)=\int_{q\in \underline{Y}} R^*(a,q)J(q\pm y)\,dq=\int_{\{q|q\geqslant a,\, q\in \underline{Y}\}} J(q\pm y)\,dq$$

are TP_2 if $J(q\pm y)$ is TP_2 (Y is the range of definition of J).

If Z has a log concave density with support in (m,n), then if either m or n is finite we can extend Z to $(-\infty,\infty)$ by defining $f(z)=0$ for $z\in(-\infty,m]$ and $f(z)=0$ for $z\in[n,\infty)$. The redefined density is log concave. The same extension of the support for a log convex model does not produce a log convex function. It is not possible to have a log convex random variable defined with support on the interval $(-\infty,\infty)$. We follow convention and assume that the support of Z is $[0,\infty)$.

For the case of log concave random variable Z with density J we may use the extension argument to write (B-1) and (B-2) as

(B-1)′ $$M(a,y)=N(a-y)=\int_{-\infty}^{a-y} J(t)\,dt,$$

(B-2)′ $$M^*(a,y)=N^*(a-y)=\int_{a-y}^{\infty} J(t)\,dt.$$

Thus by the Polya-Szego theorem, $\Pr(Z\leqslant a)$ and $\Pr(Z\geqslant a)$ are log concave functions of a. By the same reasoning

$$\int_{-\infty}^{b}\Pr(Z\leqslant a)\,da \quad\text{and}\quad \int_{b}^{\infty}\Pr(Z\geqslant a)\,da$$

are log concave in b provided the integrals exist. The argument may be repeated for integrals of these integrals and so forth.

For the case of J log convex with the support of $Z\geqslant 0$, we may write

(B-2)″ $$M^*(a,y)=N^*(a+y)=\int_{a+y}^{\infty} J(t)\,dt$$

and use the Polya-Szego theorem to prove that $N^*(a+y)$ is TP_2 and hence log convex in a. Thus $\Pr(Z\geqslant a)$ is log convex in a. The analogous expression corresponding to (B-1)′,

$$M(a,y)=\int_0^a J(q+y)\,dq=\int_y^{a+y} J(t)\,dt,$$

is TP_2. However, this does not imply that $\Pr(Z\leqslant a)$ is log convex in a. Thus we cannot assert that $\Pr(Z\leqslant a)$ is log convex in a. This is so because $M(a,y)$ is a function of both a and y and not just $a+y$.

For example, if the density of Z is Pareto with

$$f(z)=(\gamma-1)z^{-\gamma},\qquad z\geqslant 1,\ \gamma>1,$$

$\Pr(Z>a>1)=a^{1-\gamma}$ which is log convex in a. But $\Pr(Z\leqslant a)=1-a^{1-\gamma}$ is log convex in a only if $a<(\gamma)^{-1/(\gamma-1)}$ which is not always true (e.g., $\gamma=2$, $a=3$).

To derive the results stated in the text, it is useful to define

$$S_{j+1}(a)=\int_a^{\infty} S_j(z)\,dz \quad\text{where}\quad S_0=S(a)\overset{\text{def}}{\equiv}P(z>a)$$

and

$$F_{j+1}(a)=\int_{-\infty}^{a} F_j(z)\,dz \quad\text{where}\quad F_0=F(a)\overset{\text{def}}{\equiv}P(z\leqslant a).$$

Then for $E|Z| < \infty$, using integration by parts,

$$\text{(B-3)} \qquad E(Z|Z>a) = a + \frac{S_1(a)}{S_0(a)},$$

$$\text{(B-4)} \qquad E(Z|Z\leqslant a) = a - \frac{F_1(a)}{F_0(a)}.$$

For $E(Z^2) < \infty$, integration by parts produces

$$\text{(B-5)} \qquad \operatorname{Var}(Z|Z>a) = \frac{2S_2(a)}{S_0(a)} - \left(\frac{S_1(a)}{S_0(a)}\right)^2,$$

$$\text{(B-6)} \qquad \operatorname{Var}(Z|Z\leqslant a) = \frac{2F_2(a)}{F_0(a)} - \left(\frac{F_1(a)}{F_0(a)}\right)^2.$$

Observe that if the density of Z is log convex (concave), then (B-3)–(B-6) are differentiable with respect to a almost everywhere with respect to Z. Taking derivatives,

$$\text{(B-7)} \qquad \frac{\partial E(Z|Z>a)}{\partial a} = \frac{S_1(a)S_1''(a)}{(S_1'(a))^2}\begin{cases} \leqslant 1 & (Z \text{ log concave}),\\ \geqslant 1 & (Z \text{ log convex}),\end{cases}$$

$$\text{(B-8)} \qquad \frac{\partial E(Z|Z\leqslant a)}{\partial a} = \frac{F_1(a)F_1''(a)}{(F_1'(a))^2} \leqslant 1 \quad (Z \text{ log concave}),$$

$$\text{(B-9)} \qquad \frac{\partial \operatorname{Var}(Z|Z>a)}{\partial a} = \frac{-(2S_0'(a))}{S_0^2(a)}\left(S_2(a) - \frac{S_1^2(a)}{S_0(a)}\right)\begin{cases} \leqslant 0 & (Z \text{ log concave}),\\ \geqslant 0 & (Z \text{ log convex}),\end{cases}$$

$$\text{(B-10)} \qquad \frac{\partial \operatorname{Var}(Z|Z\leqslant a)}{\partial a} = \frac{2F_0'(a)}{F_0^2(a)}\left(-F_2(a) + \frac{F_1^2(a)}{F_0(a)}\right) \geqslant 0 \quad (Z \text{ log concave}).$$

Note that the hypothesis that the density of Z is log concave (convex) is *stronger* than is required to produce the orderings in (B-7)–(B-10). If $S_1(a)$ is log concave (log convex) the slope of the truncated mean (B-3) is less than or equal to (greater than or equal to) one. $S_1(a)$ can be log concave or convex without the density of Z being log concave or convex. Similarly the log concavity of $F_1(a)$ is all that is required in (B-8). By parallel logic, propositions (B-9) and (B-10) hinge on the log convexity (concavity) of $S_2(a)$ and log concavity of $F_2(a)$. The inequalities are strict if convexity (concavity) is strict. Note that if Z does not have a log concave density, then $S_1(a)$ may be log concave while $F_1(a)$ need not be and vice versa.

Note that if Z were defined to be log convex only on the support $[0, m]$, $m < \infty$, (B-7) and (B-9) or (20) and (21) in the text need not be true. For example if Z is $U(0,1)$, it is log convex (but not strictly) on the support $[0,1]$ but for $0 \leqslant a \leqslant 1$

$$\frac{\partial E(Z|Z>a)}{\partial a} = \frac{1}{2} < 1$$

and

$$\frac{\partial \operatorname{Var}(Z|Z>a)}{\partial a} = \frac{a-1}{6} \leqslant 0.$$

Partial versions of the results given here for the log concave case are presented by Flinn and Heckman (1983) and Goldberger (1983). The papers by Prekopa (1971, 1973), Borell (1975), and Brascamp and Lieb (1975, 1976) give results that can be used in the log concave case. Karlin (1968) is the basic reference. Karlin (1982) provides a different proof of our result on variances for the log concave case.

APPENDIX C

PROOFS OF THEOREMS 2 AND 6

PROOF OF THEOREM 2: Let P_1 be the proportion of the population in sector one in the Roy economy. $P_2 = 1 - P_1$ is the proportion in sector two. The overall variance of log earnings in the economy is

$$V = P_1 \operatorname{Var}[\ln W_1 | \ln W_1 > \ln W_2] + P_2 \operatorname{Var}[\ln W_2 | \ln W_1 \leqslant \ln W_2]$$
$$+ P_1 P_2 \big(E[\ln W_1 | \ln W_1 > \ln W_2] - E[\ln W_2 | \ln W_1 \leqslant \ln W_2] \big)^2.$$

Assume that the skill prices in the random assignment economy are $\tilde{\pi}_1$ and $\tilde{\pi}_2$. The skill prices in the Roy economy are π_1 and π_2. The variance of log earnings in the random assignment economy is

$$\tilde{V} = P_1 \sigma_{11} + P_2 \sigma_{22} + P_1 P_2 (\ln \tilde{\pi}_1 + \mu_1 - \ln \tilde{\pi}_2 - \mu_2)^2.$$

We establish that $\tilde{V} > V$. Define $c_* = (\ln \pi_1 + \mu_1 - \ln \pi_2 - \mu_2)/\sigma$. Also define $\tilde{c}_* = (\ln \tilde{\pi}_1 + \mu_1 - \ln \tilde{\pi}_2 - \mu_2)/\sigma$, $\rho_1 = a_1 \sigma / \sqrt{\sigma_{11}}$ and $\rho_2 = a_2 \sigma / \sqrt{\sigma_{22}}$.

In the Roy economy, $P_1 = \Phi(c_*)$ and $P_2 = \Phi(-c_*)$ where Φ is the cdf of the unit normal. Using (10), (11), (R-1), and (R-4), we obtain

$$V = \sigma_{11} \Phi(c_*) \big[1 + \rho_1^2 (-c_* \lambda(-c_*) - \lambda^2(-c_*)) \big]$$
$$+ \sigma_{22} \Phi(-c_*) \big[1 + \rho_2^2 (c_* \lambda(c_*) - \lambda^2(c_*)) \big]$$
$$+ \sigma^2 \Phi(c_*) \Phi(-c_*) \left(c_* + \frac{\rho_1 \sqrt{\sigma_{11}}}{\sigma} \lambda(-c_*) - \frac{\rho_2 \sqrt{\sigma_{22}}}{\sigma} \lambda(c_*) \right)^2$$

and

$$\tilde{V} = \Phi(c_*) \sigma_{11} + \Phi(-c_*) \sigma_{22} + \sigma^2 \Phi(c_*) \Phi(-c_*) \tilde{c}_*^2.$$

We will now prove that $\tilde{V} - \sigma^2 \Phi(c_*) \Phi(-c_*) \tilde{c}_*^2 \geqslant V$ which will imply the theorem. In other words we will prove for $-\infty < c_* < \infty$ that

$$\text{(C-1)} \qquad \sigma_{11} \Phi(c_*) \big[\rho_1^2 (-c_* \lambda(-c_*) - \lambda^2(-c_*)) \big] + \sigma_{22} \Phi(-c_*) \big[\rho_2^2 (c_* \lambda(c_*) - \lambda^2(c_*)) \big]$$
$$+ \sigma^2 \Phi(c_*) \Phi(-c_*) \left(c_* + \frac{\rho_1 \sqrt{\sigma_{11}}}{\sigma} \lambda(-c_*) - \frac{\rho_2 \sqrt{\sigma_{22}}}{\sigma} \lambda(c_*) \right)^2 \leqslant 0.$$

Notice that with a_1 defined as in the text,

$$a_1 = \frac{\rho_1 \sqrt{\sigma_{11}}}{\sigma} \quad \text{and} \quad 1 - a_1 = \frac{\rho_2 \sqrt{\sigma_{22}}}{\sigma},$$

so we may express the left hand side of (C-1) as

$$\sigma^2 \big(a_1^2 \Phi(c_*) (-c_* \lambda(-c_*) - \lambda^2(-c_*)) + (1 - a_1)^2 \Phi(-c_*) (c_* \lambda(c_*) - \lambda^2(c_*))$$
$$+ \Phi(c_*) \Phi(-c_*) [a_1 (c_* + \lambda_1(-c_*)) + (1 - a_1)(c_* - \lambda_1(c_*))]^2 \big).$$

This expression may be written as

$$\sigma^2 \big(a_1^2 \eta_1 + (1 - a_1)^2 \eta_2 + 2 a_1 (1 - a_1) \eta_3 \big)$$

where

$$\begin{aligned}\eta_1 &= \Phi(c_*)(-c_*\lambda(-c_*)-\lambda^2(-c_*)) + \Phi(c_*)\Phi(-c_*)(c_*+\lambda(-c_*))^2\\ &= \Phi(c_*)(c_*+\lambda(-c_*))(-\lambda(-c_*)+\Phi(-c_*)(c_*+\lambda(-c_*))),\\ \eta_2 &= \Phi(-c_*)(c_*\lambda(c_*)-\lambda^2(c_*)) + \Phi(-c_*)\Phi(c_*)(c_*-\lambda(c_*))^2\\ &= \Phi(-c_*)(c_*-\lambda(c_*))(\lambda(c_*)+\Phi(c_*)(c_*-\lambda(c_*))),\\ \eta_3 &= \Phi(c_*)\Phi(-c_*)(c_*-\lambda(c_*))(c_*+\lambda(-c_*)).\end{aligned}$$

Now

$$\begin{aligned}\eta_1-\eta_3 &= \Phi(c_*)(c_*+\lambda(-c_*))\\ &\qquad \times(-\lambda(-c_*)+\Phi(-c_*)(c_*+\lambda(-c_*))-\Phi(-c_*)(c_*-\lambda(c_*)))\\ &= \Phi(c_*)(c_*+\lambda(-c_*))(-\lambda(-c_*)+\Phi(-c_*)(\lambda(c_*)+\lambda(-c_*)))\\ &= \Phi(c_*)(c_*+\lambda(-c_*))(-\lambda(-c_*)+\lambda(-c_*))\\ &= 0\end{aligned}$$

where we have used $\Phi(-c_*)(\lambda(c_*)+\lambda(-c_*))=\lambda(-c_*)$. So $\eta_1=\eta_3$, and by a similar argument $\eta_2=\eta_3$.

We can therefore express the left hand side of (C-1) as

$$\sigma^2\eta_3\left(a_1^2+(1-a_1)^2+2a_1(1-a_1)\right)=\sigma^2\eta_3$$

but since $\lambda(c_*)-c_*>0$ and $\lambda(-c_*)+c_*>0$, η_3 must be negative, which establishes (C-1). This completes the proof. *Q.E.D.*

Proof of Theorem 6: Using (7a) and (7b), $f(\ln w_1|\ln W_1>\ln W_2)$ and $\Pr(\ln W_1>\ln W_2)$ depend only on $\sigma_{11}(1-\rho^2)$, μ_1, $c_*=(\mu_1-\mu_2)/\sigma$, and $\sigma_{11}^{1/2}\rho(=a_1\sigma)$, so the best we can hope for is to identify μ_1, σ_{11}, ρ, and c_*. To see that these are actually identified first notice that

$$\Pr(\ln W_1>\ln W_2)=\Phi(c_*).$$

This identifies c_*. Knowledge of c_* allows us to calculate $k_1=E[Z|Z>-c_*]$, $k_2=\mathrm{Var}[Z|Z>-c_*]$, and $k_3=E[(Z-E[Z|Z>-c_*])^3|Z>-c_*]$. From (10)–(12), the first three moments of $\ln W_1$ given $\ln W_1>\ln W_2$ are

(C-2) $$E[\ln W_1|\ln W_1>\ln W_2]=\mu_1+\rho\sigma_{11}^{1/2}k_1,$$

(C-3) $$\mathrm{Var}[\ln W_1|\ln W_1>\ln W_2]=\sigma_{11}\left(\rho^2k_2+(1-\rho^2)\right)=\sigma_{11}+(k_2-1)\rho^2\sigma_{11},$$

and

(C-4) $$E\left[(\ln W_1-E[\ln W_1|\ln W_1>\ln W_2])^3\middle|\ln W_1>\ln W_2\right]=\rho^3\sigma_{11}^{3/2}k_3.$$

(C-4) identifies $\rho\sigma_{11}^{1/2}$. (C-2) then identifies μ_1. (C-3) identifies σ_{11}. This also identifies ρ as $\rho\sigma_{11}^{1/2}$ is already identified. *Q.E.D.*

REFERENCES

Atkinson, A. (1970): "On the Measurement of Inequality," *Journal of Economic Theory*, 2, 244–263.

Basu, A. P., and J. K. Ghosh (1978): "Identifiability of the Multinormal and Other Distributions under Competing Risks Model," *Journal of Multivariate Analysis*, 8, 413–429.

Borell, C. (1975): "Convex Functions in d-Space," *Periodica Mathematica Hungarica*, 6, 111–136.

Brascamp, H., and E. Lieb (1975): "Some Inequalities For Gaussian Measures," in *Functional Integral and Its Applications*, ed. by A. Arthurs. Oxford: Clarendon Press.

——— (1976): "On Extensions of the Brunn-Minkowski and Prekopa-Leindler Theorems Including Inequalities for Log Concave Functions and with Applications to the Diffusion Equation," *Journal of Functional Analysis*, 22, 336–389.

Caplin, A., and B. Nalebuff (1989): "Aggregation and Social Choice: A Mean Voter Theorem," Princeton University, unpublished paper.

COX, D. R. (1962): *Renewal Theory*. Muthuen's Monographs on Applied Probability and Statistics. London: Methuen and Co.

COX, D. R., AND W. OAKES (1984): *Analysis of Survival Data*. London: Chapman and Hall.

CREEDY, J. (1985): *Dynamics of Income Distribution*. Oxford: Basil Blackwell.

FLINN, C., AND J. HECKMAN (1983): "Are Unemployment and Out of The Labor Force Behaviorally Distinct Labor Force States?" *Journal of Labor Economics*, 1, 28–42.

GOLDBERGER, A. (1983): "Abnormal Selection Bias," in *Studies in Econometrics*, *Time Series and Multivariate Statistics*, ed. by S. Karlin, T. Amemiya, and L. Goodman. New York: Academic Press.

GRONAU, R. (1974): "Wage Comparisons—A Selectivity Bias," *Journal of Political Economy*, 82, 1119–1144.

HECKMAN, J. J. (1974): "Shadow Prices, Market Wages and Labor Supply," *Econometrica*, 42, 679–694.

——— (1976): "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Model," *Annals of Economic and Social Measurement*, 15, 475–492.

——— (1990): "Varieties of Selection Bias," *American Economic Review*, 80, 313–318.

HECKMAN, J. J., AND B. E. HONORÉ (1989): "The Identifiability of the Competing Risks Models," *Biometrika*, 76, 325–330.

HECKMAN, J. J., AND G. SEDLACEK (1981): "The Impact of the Minimum Wage on the Employment and Earnings of Workers in South Carolina," *Report of Minimum Wage Study Commission*, Vol. 4. Washington, D.C.: U.S. Government Printing Office.

——— (1985): "Heterogeneity, Aggregation and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market," *Journal of Political Economy*, 93, 1077–1125.

——— (1990): "Self-Selection and the Distribution of Hourly Wages," *Journal of Labor Economics*, 8, S329–363.

KALBFLEISCH, J. D., AND R. L. PRENTICE (1980): *The Statistical Analysis of Failure Time Data*. New York: Wiley.

KARLIN, S. (1968): *Total Positivity*. Stanford: Stanford University Press.

——— (1982): "Some Results on Optimal Partitioning of Variance and Monotonicity with Truncation Level," in *Statistics and Probability*: *Essays in Honor of C.R. Rao*, ed. by G. Kallianpur, P.R. Krishnaiah, and J.K. Ghosh. Amsterdam: North Holland.

KARLIN, S., AND Y. RINOTT (1983): "Comparison of Measures, Multivariate Majorization and Applications to Statistics," in *Studies in Econometrics*, *Time Series and Multivariate Statistics*, ed. by S. Karlin, T. Amemiya, and L. Goodman. Orlando: Academic Press.

LEE, L.F. (1978): "Unionism and Wage Rates: A Simultaneous Equation Model with Qualitative and Limited Dependent Variables," *International Economic Review*, 19, 415–433.

——— (1982): "Some Approaches to the Correction of Selectivity Bias," *Review of Economic Studies*, 49, 355–372.

MCELROY, M., AND M. HORNEY (1981): "Nash Bargained Household Decisions: Toward a Generalization of the Theory of Demand," *International Economic Review*, 22, 333–349.

MILLER, R. (1984): "Job Matching and Occupational Choice," *Journal of Political Economy*, 92, 1086–1120.

POLYA, G., AND G. SZEGO (1972): *Problems and Theorems in Analysis I*. Berlin and New York: Springer Verlag.

PRATT, J. (1981): "Concavity of the Log Likelihood," *Journal of the American Statistical Association*, 76, 103–106.

PREKOPA, A. (1971): "Logarithmic Concave Measures with Application to Stochastic Programming," *Acta Scientiarum Mathematicarum*, 32, 301–315.

——— (1973): "On Logarithmic Concave Measures and Functions," *Acta Scientiarum Mathematicarum*, 34, 335–343.

ROBINSON, C., AND N. TOMES (1980): "Self Selection and Interprovincial Migration in Canada," *Canadian Journal of Economics*, 15, 474–502.

ROY, A. D. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers* (*New Series*), 3, 135–146.

TSIASTIS, A. (1975): "A Nonidentifiability Aspect of the Problem of Competing Risks," *Proceedings of the National Academy of Sciences*, 72, 20–22.

WILLIS, R., AND S. ROSEN (1979): "Education and Self-Selection," *Journal of Political Economy*, 87, S7–S36.