Skills and Tasks in the Labor Market

James J. Heckman
University of Chicago

Econ 312, Winter 2021

A Simple Model of Task Selection

N possible tasks, $T_1, \ldots, T_N$. Tasks can be occupations or some category of input valuable in final production.
Final output Y^j in J sectors:

$$Y^j = f^j(T_1^j, \ldots, T_N^J), j = 1, \ldots, J.$$

Assume constant returns to scale. T_l^j is the amount of task l used in firm (sector) j. Assume task production functions are uniform across sectors:

$$T_l^j = g^l(S_1^l, \ldots, S_K^l) \text{ for all } j = 1, \ldots, J$$

where $\{S_1^l, \ldots, S_K^l\}$ are skills used in producing l (technology mapping skills to tasks is the same across sectors).

Let skill prices be uniform across sectors. π_r is the price of task r. How determined? Cost minimization. Consider task l.
$\min \sum_{r=1}^{N} \pi_r S_r^l$ subject to $T_l = g^l(S_1^l, \ldots, S_K^l) = T_l$: Let price per unit skill be arranged in a vector $(W_l, \ldots, W_k) = W$.

$$\mathcal{L} = W^1 S^l + \lambda_l(\bar{T}_l - g^l(S_1^l, \ldots, S_K^l)).$$

$\lambda_l = mc$ of producing task l. As the maps $S^l \to T_l$ change, so do prices of the tasks. Total demand for skill by sector j is determined from

$$\max P_j Y^j - \sum_{l=1}^{N} \pi_l T_l^j$$

FOC for S_l^j:

$$P_j \frac{\partial Y^j}{\partial T_l^j} \frac{\partial T^j}{\partial S_l^j} - \pi_l \frac{\partial T_l^j}{\partial S_l^j} \geq 0.$$

Interior solution:

$$P_j \frac{\partial Y^j}{\partial T_l^j} = \pi_l$$

determines S_l^j. Total input of S_l by sector j is $\sum_{l=1}^{N} S_l^j$.

Total demand for skill l:

$$D(l) = \sum_{j=1}^{J} S_l^j.$$

Comments:

1. Output Y^j is positive even if some tasks not used. Skills used in $j = \{j | P_j \frac{\partial f^j}{\partial T^j} \geq \pi_j\}$.
2. New tasks may be used as $\pi = (\pi, \ldots, \pi_K)$ changes.
3. Analogous to Becker household production model (see your 301 notes).

THE UNIVERSITY OF CHICAGO

- T_i is total amount of task employed in sector i.
- A_i is the vector of non-labor inputs in sector i.
- $F^{(i)}(T_i, A_i)$ is the aggregate output of sector i
- $F^{(i)}$ is twice continuously differentiable, increasing and concave.
- Furthermore, $F^{(i)}(0, A_i) = F^{(i)}(T_i, 0) = F^{(i)}(0, 0) = 0$. (This is not essential.)
- P_i is the price of sector i output.
- π_i is the price of one unit of sector i specific task.

Application: Heckman & Sedlacek

- Two market sectors, $i = 1, 2$.
- Each agent is endowed with skills $s \in \mathbb{R}_+^J$.
- The population distribution of s is $g(s \mid \Theta)$ where Θ is a vector of parameters.
- Agents do not invest in order to change skills s. (Will be relaxed.)

Task Function

- $t_i(s)$ is a function that expresses the amount of sector i specific tasks a worker with endowment of skills s can perform.
- Determined technologically. (But will change this when we consider Hedonic models.)
- Convenient representation, widely used in subsequent literature.

Firm optimization implies:

$$\pi_i = P_i \frac{\partial F^{(i)}}{\partial T_i} \tag{1}$$

In a two-sector economy, an agent with endowment s works in sector i if:

$$\pi_i t_i(s) \geq \pi_j t_j(s) \, i, j \in \{1, 2\}. \tag{2}$$

Let $\mathcal{L}_i$ denote the set of agents working in sector i :

$$\mathcal{L}_i = \{s : \pi_i t_i(s) \geq \pi_j t_j(s), i \neq j\}.$$

The log wage in sector i of an individual with endowment s is:

$$\ln w_i(s) = \ln \pi_i + \ln t_i(s) \tag{3}$$

The proportion of the population working in sector i is:

$$pr(i) = \int_{\mathcal{L}_i} g(s|\Theta)\, ds, \quad i = 1, 2$$

Roy model assumes that $g(s|\Theta)$ and $t_i(s)$ are such that:

$$\begin{bmatrix} \ln t_1(s) \\ \ln t_2(s) \end{bmatrix} \sim N\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Sigma\right]$$

- In the Roy model agents choose between two possible sectorial wages:

$$\ln w_1 = \ln \pi_1 + \mu_1 + u_1$$

or

$$\ln w_2 = \ln \pi_2 + \mu_2 + u_2$$

- Workers enter sector 1 if $\ln w_1 \geq \ln w_2$.
- Otherwise they enter sector 2.

- Thus,

$$\underbrace{\ln w_1 - \ln w_2}_{\text{selection index } I_1} \geq 0$$

  plays a key role
- Index:

$$I_1 = \ln(\pi_1/\pi_2) + \mu_1 - \mu_2 + u_1 + u_2$$

- Let $\mathrm{Var}(u_1 - u_2) = (\sigma^*)^2$

$$\frac{I_1}{\sigma^*} = \frac{\ln \pi_1/\pi_2 + \mu_1 - \mu_2 + u_1 + u_2}{\sigma^*}$$

Normal Version

Let

$$\sigma^* = \sqrt{var\left(u_1 - u_2\right)}$$

$$E\left(\frac{u_1 - u_2}{\sigma^*} \mid \frac{I_1}{\sigma^*} > 0\right)$$

$$= E\left(\frac{u_1 - u_2}{\sigma^*} \mid \frac{u_1 - u_2}{\sigma^*} > -\left[\ln \pi_1/\pi_2 + \mu_1 - \mu_2\right]\right)$$

$$= E\left(\frac{u_1 - u_2}{\sigma^*} \mid \frac{u_1 - u_2}{\sigma^*} > -c_1\right)$$

- $\frac{u_1 - u_2}{\sigma^*}$ is standard normal, as is $\frac{u_2 - u_1}{\sigma^*}$

$$= E\left(\frac{u_1 - u_2}{\sigma^*} \mid c_1 \geq \left(\frac{u_2 - u_1}{\sigma^*}\right)\right)$$

(Symmetry of normal)

The mean of log wages observed in sector i is (more generally):

$$E(\ln w_i \mid \ln w_i \geq \ln w_2) = \ln \pi_i + \mu_i + E(U_i \mid \ln w_i \geq \ln w_2) \qquad (4)$$

For normal:

$$E\left(\ln w_i \mid \ln w_1 \geq \ln w_2\right) = \ln \pi_i + \mu_i + \left(\frac{\sigma_{ii} - \sigma_{ij}}{\sigma^*}\right) \lambda\left(c_i\right) \qquad (5)$$

The variance of log wages observed in sector i is:

$$var\left(\ln w_i \mid \ln w_1 \geq \ln w_2\right) = \sigma_{ii} \underbrace{\left[\begin{array}{r} \rho_i^2\left(1 - c_i \lambda\left(c_i\right) - \lambda^2\left(c_i\right)\right) \\ +\left(1 - \rho_i^2\right) \end{array}\right]}_{\leq 1} \qquad (6)$$

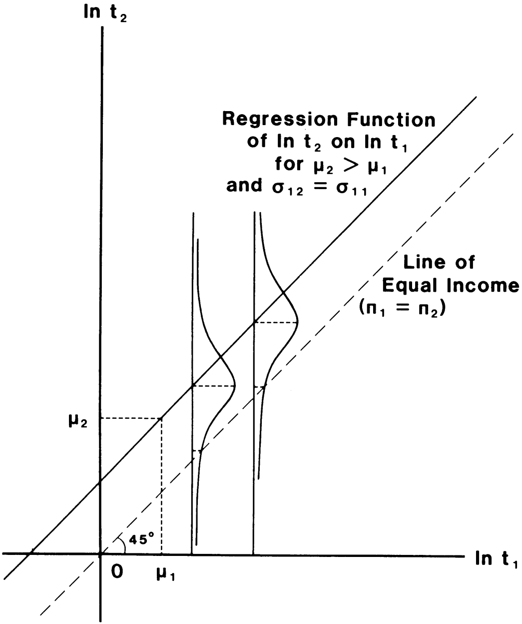
(True for all log concave distributions, e.g., normal.)

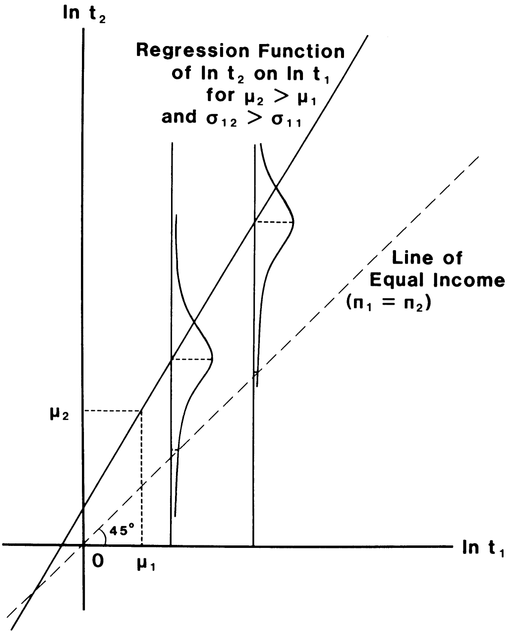
The linear projection (regression) of $\ln t_2$ conditional on $\ln t_1$ is:

$$\ln t_2 = \mu_2 + \frac{\sigma_{12}}{\sigma_{11}} (\ln t_1 - \mu_1) + \varepsilon_2 \qquad (7)$$

$E(\varepsilon_2) = 0,\ var(\varepsilon_2) = \sigma_{22}\left[1 - \frac{\sigma_{12}^2}{\sigma_{11}\sigma_{22}}\right].$

$\varepsilon_2 = u_2 - E(u_2 \mid w_1 > w_2)$


ln t_2
Regression Function
of ln t_2 on ln t_1
for $\mu_2 > \mu_1$
and $\sigma_{12} = \sigma_{11}$
Line of
Equal Income
$(\pi_1 = \pi_2)$
μ_2
45°
0
μ_1
ln t_1


ln t_2
Regression Function
of ln t_2 on ln t_1
for $\mu_2 > \mu_1$
and $\sigma_{12} > \sigma_{11}$
Line of
Equal Income
($\pi_1 = \pi_2$)
μ_2
45°
0
μ_1
ln t_1

Estimating the Model

Estimating the Model

a) need to identify the parameters of the distribution of tasks g and functions t_i.

b) parameters of the sectoral demand functions.

The data available are:

(i) time-series data on the aggregate amount of **compensation** paid to workers in each sector.

(ii) microeconomic repeated cross-section data on the wages of workers by sector and their associated demographic and productivity characteristics

(iii) time-series data on sectoral determinants of the demand for tasks.

The task function is assumed to be:

$$\ln t_i = \beta_i X + u_i, \;\; i = 1, 2 \tag{8}$$

The log real wages are:

$$\ln w_i = \ln \pi_i + \beta_i X + u_i, \;\; i = 1, 2 \tag{9}$$

In normal case, unless $\sigma_{ii} - \sigma_{ij} = 0$, OLS estimators are inconsistent because of selection bias.

The intercept of equation (9) combines two parameters: the log of the real price of task i, $\ln \pi_i$, and the intercept of the task function, denoted β_{0i}.

- Call intercept $\ln \tilde{\pi}_i = \ln \pi_i + \beta_{0i}$

To obtain the quantities of log task employed in each sector in each period, subtract the estimated intercept from the log real wage bill in each sector i, $\ln WB_i$.

- I.e., $\ln T_i = \ln WB_i - \ln \pi_i$

This produces an estimated of labor aggregate $\ln T_i$ up to a known additive constant β_{0i}.

Let l denote a year subscript, assuming that the aggregate derived demand for tasks is loglinear in aggregate tasks and real task prices, write:

$$\ln T_{il} = \delta_{0i} + \delta_{1i} \ln \left(\frac{\pi_{il}}{P_{il}} \right) + \delta_{2i} \ln \left(\frac{P_{Al}}{P_{il}} \right) + e_{il} \tag{10}$$

where:

- e_{il} is mean zero stationary stochastic process
- P_{Al} is a vector of real prices for other inputs
- P_{il} is the real price of output of sector i at time l.

- Set $\pi_{i1} = 1$ for $l = 1$ for both $i = 1$ and $i = 2$ defines the units of tasks T_{il}.
- $WB_{il} = \pi_{il} T_{il}$.
- Write (10) as:

$$\ln\left(\frac{WB_{il}}{\pi_{il}}\right) = \left[\delta_{0i} - \beta_{0i}\left(1 + \delta_{1i}\right)\right] + \left(1 + \delta_{1i}\right)\left(\ln \tilde{\pi}_{il}\right)\delta_{il}\left(\ln P_{il}\right) + \delta_{2i}\ln\left(\frac{P_{Al}}{P_{il}}\right) + e_{il} \tag{11}$$

where $\ln \tilde{\pi}_{il}$ is the estimator $\ln \pi_i$ from the intercepts of wage equations.

Because aggregate shocks e_{il} affect P_{il} and π_{il}, OLS is inconsistent in (11).

When the Roy model is fit on CPS earnings data disaggregated into manufacturing and nonmanufacturing sectors, it is rejected:

1. **The proportionality hypothesis: assumes invariance of wage functions, except for intercepts.**
2. Tested and rejected.
3. χ^2 goodness-of-fit strongly rejects distributional assumptions.

An Extended Roy Model

Utility Maximizing Version

1. Assume workers maximize utility.
2. Decompose earnings into hourly wages rates and hours of work.
3. General nonnormal model for (u_1, u_2) that nests Roy's model as a special case.
4. Incorporates nonmarket sector as an alternative market.

In place of task function (8), consider, Box-Cox transformation.

$$\frac{t_i^{\lambda_i} - 1}{\lambda_i} = \beta_i X + u_i \tag{12}$$

Random variable u_i is equated to an underlying mean zero normal random variable u_i^* for values of that variable that produce positive values of t_i, that is, $u_i = u_i^*$ if

$$1 + \lambda_i \left(\beta_i X + u_i^*\right) \geq 0 \tag{13}$$

When $\lambda_i = 0$ equation (12) specializes to the Roy model (8). By estimating λ one can determine whether or not the lognormal Roy model fits the data.

Let V_i denote the utility of participating in sector i, where $i = 1, 2, 3$, where $i = 3$ designates the nonmarket sector. An agent chooses to participate in sector i if, and only if:

$$V_i > V_j, \ i \neq j, \ i = 1, 2, 3. \tag{14}$$

Let Z_i denote a vector of measured sector-specific consumption attributes and household characteristics variables.

Let $\mathbf{f} = (Z, X, \ln \pi_i)$. The reduced form linearized index function:

$$\ln V_i = \gamma_i \mathbf{f} + \upsilon_i, \;\; i = 1, 2, 3 \tag{15}$$

Assume that $\mathbf{f}$ is distributed independently of all the υ_i and that $(\upsilon_1, \upsilon_2, \upsilon_3)$ is a mean zero multivariate normal random variable:

$$(\upsilon_1, \upsilon_2, \upsilon_3) \sim N(0, \Sigma_\upsilon) \tag{16}$$

This specification produces a multivariate probit model.

Since only sectoral choices and not the V_i are directly measured, it is possible to identify only parameters of the contrasts of utility evaluations among sectors. Without any loss of generality we normalize $V_3 = 0$ so $\gamma_3 = 0$ and $\upsilon_3 = 0$. Using this convention, sector i is chosen if

$$\ln V_i - \ln V_j > 0 \Rightarrow$$

$$(\gamma_i - \gamma_j)\,\mathbf{f} + (\upsilon_i - \upsilon_j) > 0 \text{ for all } i \neq j \tag{17}$$

If there is at least one nondegenerate regressor in $\mathbf{f}$, it is possible to identify $\gamma_1, \gamma_2, var\,(\upsilon_2)$, and $cov\,(\upsilon_1, \upsilon_2)$.

Empirical Estimates:
Estimates of the Extended Roy Model

Figure 1: Estimates of the Model Parameters

	Estimated Coefficient	Standard Error*	Normal Statistic†
Utility function in the nonmanufacturing sector ($\boldsymbol{\gamma}_1$):			
Intercept	4.238367	.469394	9.029442
Education	.338785	.042739	7.926800
Experience	.224682	.028620	7.850411
Experience squared/100	−.333751	.071232	−4.685396
South dummy	.282627	.136377	2.072390
Predicted nonlabor income/100	.242310	.033105	7.319353
1980 intercept (γ_{01t} for 1980)	.113196	.094107	1.202837
Utility function in the manufacturing sector ($\boldsymbol{\gamma}_2$):			
Intercept	3.103701	.565689	5.486586
Education	.285896	.053022	5.392017
Experience	.163867	.036530	4.485828
Experience squared/100	−.257929	.072256	−3.569655
South dummy	.019389	.106355	.182301
Predicted nonlabor income/100	.172409	.036337	4.744774
1980 intercept (γ_{02t} for 1980)	.017729	.074623	.237583
Correlation coefficient between υ_1 and υ_2:			
correl(υ_1, υ_2)	.296560	.147650	2.008529
Standard deviation of υ_2:			
$[\text{var}(\upsilon_2)]^{1/2}$	.850640	.117044	7.267723
Parameters of the mapping of the observed skills to the nonmanufacturing task ($\boldsymbol{\beta}_1$):			
Intercept	−.112678	.101883	−1.105953
Education	.040472	.007908	5.117798
Experience	.005979	.008301	.720287

Figure 1: Estimates of the Model Parameters (cont.)

	Estimated Coefficient	Standard Error*	Normal Statistic†
Experience squared/100	.019015	.018805	1.011173
South dummy	.016770	.042527	.394325
1980 intercept (β_{01l} for 1980)	−.312877	.356679	−.877195
Parameters of the mapping of the observed skills to the manufacturing sector task ($\boldsymbol{\beta}_2$):			
Intercept	−.331493	.299324	−1.107471
Education	.082424	.010596	7.778808
Experience	.027506	.012970	2.120790
Experience squared/100	−.027446	.028786	−.953469
South dummy	−.102184	.060104	−1.700135
1980 intercept (β_{02l} for 1980)	.038270	1.152317	.033212
Covariance structure of the latent task distribution:			
$(\sigma_{11}^{1/2}) = [\text{var}(u_1^*)]^{1/2}$	.574169	.006098	94.159852
$(\sigma_{22}^{1/2}) = [\text{var}(u_2^*)]^{1/2}$	.486769	.081631	5.963048
$\rho_{12}^* = \text{correl}(u_1^*, \upsilon_2 - \upsilon_1)$	.241512	.029820	8.351013
$\rho_{11}^* = \text{correl}(u_1^*, \upsilon_1)$	.454436	.029116	15.607939
$\rho_{21}^* = \text{correl}(u_2^*, \upsilon_2 - \upsilon_1)$	.235583	.009276	25.397051
$\rho_{22}^* = \text{correl}(u_2^*, \upsilon_2)$	.159303	.004145	38.435299
1980 estimated log task price change where $\pi_1(1976) = \pi_2(1976) = 1$:			
Nonmanufacturing sector ($\ln \widehat{\pi}_{1l}$ for 1980)	.216560	.003588	60.358733
Manufacturing sector ($\ln \widehat{\pi}_{2l}$ for 1980)	−.225510	.005036	−44.777223

Figure 1: Estimates of the Model Parameters (cont.)

	Estimated Coefficient	Standard Error*	Normal Statistic†
Nonmanufacturing sector (λ_1)	−.060494	.032839	−1.842148
Manufacturing sector (λ_2)	.082650	.008020	10.305595
Log-likelihood for the model −2,099.01			
Number of individuals in the sample 3,262			

	χ^2 Statistic for the Hypothesis	Number of Degrees of Freedom	Values of χ^2 Random Variables at 5 Percent Significance Level for Stated Number of Degrees of Freedom
Likelihood ratio test for restricted model, $\lambda_1 = \lambda_2 = 0$:			
1976 data	8.18	2	5.99
1980 data	7.46	2	5.99
Goodness-of-fit test‡ for the extended Roy model:			
Manufacturing	34.1	50	67.51
Nonmanufacturing	64.7	50	67.51
Goodness-of-fit‡ for the lognormal three-sector model with $\lambda_1 = \lambda_2 = 0$:			
Manufacturing	42.7	50	67.51
Nonmanufacturing	71.9	50	67.51
Strong proportionality hypothesis	15.7	26	38.89

* Standard errors are computed from the square root of the diagonal elements of minus the inverse of the Hessian of the log likelihood.

† The ratio of the estimated coefficient to the estimated standard error. This ratio, when multiplied by the square root of the sample size, is asymptotically normal under the null hypothesis that the corresponding population parameter is zero.

‡ The χ^2 goodness-of-fit statistics were computed for the conditional (on sectoral choice) log wage distributions in each sector using 51 equispaced log wage intervals starting from ln 0.75 in intervals of length 0.07535 and terminating at ln 35.0. The statistics compare predicted and actual log wage distributions in each interval, integrating out the regressor variables. In computing the χ^2 statistics we account for parameter estimation error following Moore (1977). We pool 1976 and 1980 data to perform the test.

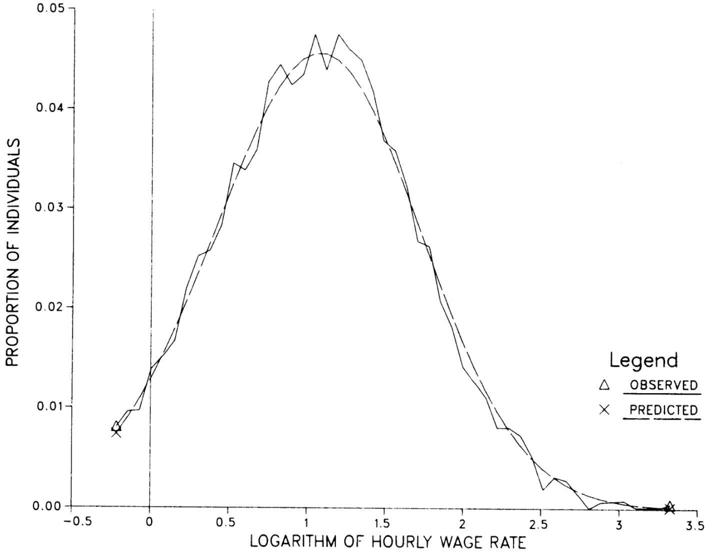


FIG. 3.—Nonmanufacturing sector: predicted versus observed log wage distribution

Source: Heckman and Sedlacek, 1985

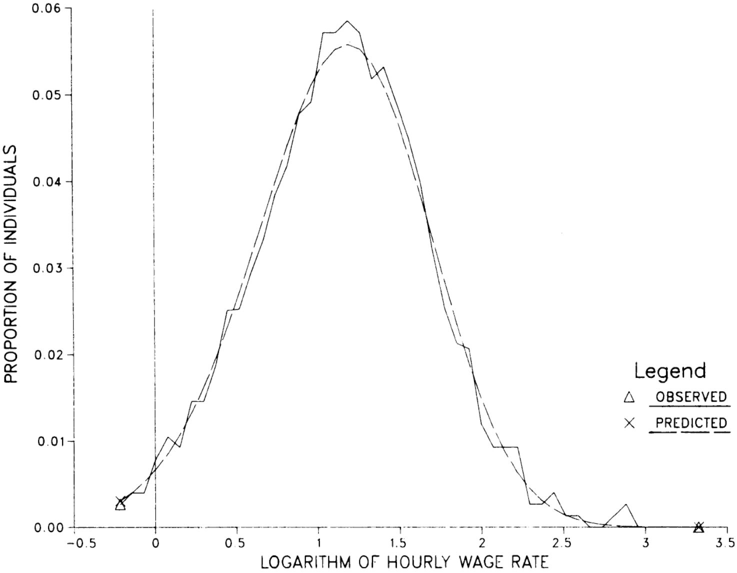


FIG. 4.—Manufacturing sector: predicted versus observed log wage distribution

Source: Heckman and Sedlacek, 1985

Link to Heckman and Sedlacek Tables

Estimating the Demand for Aggregate Sector-Specific Tasks

Figure 2: Demand functions for aggregate tasks (Eq. 19)

	Ordinary Least Squares Estimates: Estimated Coefficient	Ordinary Least Squares Estimates: Standard Error	Instrumental Variable Estimates*: Estimated Coefficient	Instrumental Variable Estimates*: Standard Error
	Nonmanufacturing Sector			
Constant (δ_{01})	12.119010	.11277737	11.900640	1.6234258
Log task price (δ_{11})†	−.951021	.02161820	−.934039	.3674947
Log energy price index (δ_{21})	.394647	.07575938	1.120513	1.0200879
Log intermediate goods price (δ_{31})	−.488665	.49281826	−.116150	6.4352536
Log user cost of capital (δ_{41})	−.099360	.05669152	.7744651	.7744651
R^2	.9958		...	
Number of observations (1968–81)	14		14	
Durbin-Watson statistic‡	1.447		1.462	
	Manufacturing Sector			
Constant (δ_{02})	11.057958	.11730702	10.797219	1.8079848
Log task price (δ_{12})	−.977697	.02421021	−.974127	.4916065
Log energy price index (δ_{23})	.162611	.09507995	.925919	1.2423610
Log intermediate goods price (δ_{32})	−.706052	.51737473	−.345029	6.8214180
Log user cost of capital (δ_{42})	−.045386	.06814409	.099210	1.1129916
R^2	.9905		...	
Number of observations (1968–81)	14		14	
Durbin-Watson statistic	1.966		2.200	

Note.—For the definitions of these variables see App. C.

* The instruments are: log energy price index, log intermediate goods price index, log user cost of capital, total population, total population squared, average weekly hours worked in the nonmanufacturing sector, unemployment rate in the United States. For further discussion see App. C. The regression results are unaffected when the hours worked variable is not used as an instrument.

† The reported coefficients are the estimated coefficients on log task prices from regression equations of the form (19) minus one.

‡ The lower limit for the Durbin-Watson test for a 5 percent significance level with five regressors (including an intercept) and 15 observations is 0.69. The upper limit is 1.97. The limits for 14 observations are wider.

Exploring the Importance of Aggregation Bias in Aggregate Wages

Figure 3: Simulation of a 1% increase in the energy price index

	Manufacturing Sector	Nonmanufacturing Sector	U.S. Aggregate
Year: 1972:			
1. Percentage change in persons employed	−1.854	1.320	...
2. Percentage change in mean task or quality level for the employed population	.919	−1.496	...
3. Percentage change in task price	−1.480	.471	−.062*
4. Percentage change in observed average wage (2 + 3)	−.561	−1.025	−.950
Year: 1976:			
1. Percentage change in persons employed	−2.007	1.371	...
2. Percentage change in mean task or quality level for the employed population	.886	−1.461	...
3. Percentage change in task price	−1.480	.471	−.063*
4. Percentage change in observed average wage (2 + 3)	−.594	−.990	−.939
Year: 1980:			
1. Percentage change in persons employed	−1.993	1.244	...
2. Percentage change in mean task or quality level for the employed population	.953	−1.568	...
3. Percentage change in task price	−1.480	.471	−.034*
4. Percentage change in observed average wage (2 + 3)	−.527	−.997	−.949

NOTE.—The data sets on which the simulations are performed are defined in App. C.
* This is a weighted average of the task price change in each sector using the relative proportions employed in the sector in the year.

Assessing the Impact of Self Selection on Inequality in log Wages

Figure 4: Assessing the impact of self-selection on the means and variances of log wage rates for white males, 1980

	Prediction of Extended Roy Model	Actual 1980 Value	Random Assignment Economy Using 1980 Equilibrium Task Prices
	Nonmanufacturing Sector		
Mean of log wages (M_1)	1.054	1.040	.651
Variance of log wages (σ_1)	.319	.323	.344
Proportion of population in sector (P_1)	.619*	.630	.619*
	Manufacturing Sector		
Mean of log wages (M_2)	1.199	1.202	.968
Variance of log wages (σ_2)	.192	.201	.211
Proportion of population in sector (P_2)	.200*	.206	.200*
	Economywide		
Mean of log wages $\left(\frac{P_1M_1 + P_2M_2}{P_1 + P_2}\right)$	1.089	1.079	.728
Sum of within-sector variance $\left(\frac{P_1\sigma_1 + P_2\sigma_2}{P_1 + P_2}\right)$	.288	.293	.311
Between-sector variance $\left[\frac{P_1P_2(M_1 - M_2)^2}{(P_1 + P_2)^2}\right]$	.003	.004	.018
Total variance†	.291	.297	.329

* The random assignment economy is restricted to have the proportion of people in each of the three sectors predicted by our model using 1980 equilibrium values.

† Total variance = within-variance + between-variance

$$= \left(\frac{P_1\sigma_1 + P_2\sigma_2}{P_1 + P_2}\right) + \left[\frac{P_1P_2(M_1 - M_2)^2}{(P_1 + P_2)^2}\right].$$

Further Tests of the Model

Extract from:

Self-Selection and the Distribution of Hourly Wages

Heckman and Sedlacek

Journal of Labor Economics, Vol. 8, No. 1

Part 2: Essays in Honor of Albert Rees (Jan. 1990), pp. S329–S363.

Empirical Estimates

Parameter Estimates of the Lognormal 2-Sector Model

A. Estimates

	Estimated Coefficient	Standard Error*
Utility function (relative utility: nonmanufacturing-manufacturing contrast) (γ_1):		
Intercept	.047978	.005696
Education	.049867	.006102
Experience	−.031556	.013043
Experience squared/100	.096166	.012868
South dummy	.245822	.087785
Predicted nonlabor income/100	−.036966	.005778
1980 intercept ($\gamma_{01\ell}$ for 1980)	.014025	.003535
Parameters of the mapping of the observed skills to the nonmanufacturing task (β_1):		
Intercept	−.155252	.180722
Education	.043946	.007833
Experience	.038892	.000681
Experience squared/100	−.065439	.015376
South dummy	.008170	.007082

Parameters of the mapping of the observed skills to the manufacturing task (β_2):		
Intercept	$-.016670$	.007414
Education	.077241	.010083
Experience	.028578	.003478
Experience squared/100	$-.039019$	.004573
South dummy	$-.124738$	.050928
Covariance structure of the latent task distribution:		
$(\sigma_{11})^{1/2} = [\text{var}(u_1)]^{1/2}$	.651607	.157299
$(\sigma_{22})^{1/2} = [\text{var}(u_2)]^{1/2}$	.498593	.161700
$\text{corr}(u_1, \nu_1)$	.110562	.037013
$\text{corr}(u_2, \nu_1)$	.707105	.102844
1980 Estimated log task price change where π_1 (1976) = π_2 (1976) = 1:		
Nonmanufacturing sector ($\ln \pi_{1\ell}$ for 1980)	.795274	.058032
Manufacturing sector ($\ln \pi_{2\ell}$ for 1980)	.641054	.013046
Log likelihood for the model	$-1{,}825.46$	
No. of individuals in the sample	2,654	

THE UNIVERSITY OF CHICAGO

B. Tests of the Model

	χ^2 Statistic for the Hypothesis	No. of Degrees of Freedom	Significance Level for the Test-Statistic
Strong proportionality hypothesis	206.4	17	.0001
Goodness-of-fit for the lognormal 2-sector model:†			
Manufacturing	102.6	50	.0001
Nonmanufacturing	384.2	50	.0001

* SEs are computed from the square root of the diagonal elements of minus the inverse of the Hessian of the log likelihood.

† The χ^2 goodness-of-fit statistics were computed for the conditional (on sectoral choice) log-wage distributions in each sector using 51 equispaced log-wage intervals starting from log .75 in the intervals of length .07535 and terminating at log 35.0. The statistics compare predicted and actual log-wage distributions in each interval, integrating out the regressor variables. In computing the χ^2 statistics we account for parameter estimation error following Heckman (1984). We pooled 1976 and 1980 data to perform the test.

- Reject

Parameter Estimates of the Box-Cox 2-Sector Model

A. Estimates

	Estimated Coefficient	Standard Error*
Utility function (relative utility: nonmanufacturing-manufacturing contrast) (γ_1):		
Intercept	.020968	.018254
Education	.051012	.007803
Experience	−.020606	.010245
Experience squared/100	.029398	.006339
South dummy	.285383	.174511
Predicted nonlabor income/100	−.010772	.002340
1980 intercept ($\gamma_{01\ell}$ for 1980)	.044794	.009603
Parameters of the mapping of the observed skills to the nonmanufacturing task (β_1):		
Intercept	.001909	.031294
Education	.048707	.007052
Experience	.034994	.004957
Experience squared/100	−.056451	.029420
South dummy	.038845	.042857
1980 constant	−.039643	.030782

Parameters of the mapping of the observed skills to the manufacturing task (β_2):		
Intercept	−.002646	.014413
Education	.067668	.012139
Experience	.032094	.008802
Experience squared/100	−.047250	.020862
South dummy	−.076224	.023341
1980 constant	.169560	.092070
Covariance structure of the latent task distribution:		
$(\sigma_{11})^{1/2} = [\text{var}(u_1)]^{1/2}$	.434907	.146281
$(\sigma_{22})^{1/2} = [\text{var}(u_2)]^{1/2}$	.351704	.105359
corr(u_1, ν_1)	−.182659	.102566
corr(u_2, ν_1)	−.458622	.195652
1980 estimated log task price change where π_1 (1976) = π_2 (1976) = 1:		
Nonmanufacturing sector (ln $\pi_{1\ell}$ for 1980)	.317606	.036841
Manufacturing sector (ln $\pi_{2\ell}$ for 1980)	.188139	.026500
Task transformation parameter		
Nonmanufacturing sector (λ_1)	−.186308	.048648
Manufacturing sector (λ_2)	−.100325	.035101
Log likelihood for the model	−1,762.70	
No. of individuals in the sample	2,654	

B. Tests of the Model

	χ^2 Statistic for the Hypothesis	No. of Degrees of Freedom	Significance Level for the Test-Statistic
Strong proportionality hypothesis	327.6	17	.0001
Goodness-of-fit for the lognormal 3-sector model:†			
Manufacturing	96.1	50	.0001
Nonmanufacturing	277.1	50	.0001
Likelihood-ratio test for restricted: model $\lambda_1 = \lambda_2 = 0$ (log-likelihood value -1825.46)	62.7	2	.0001

* SEs are computed from the square root of the diagonal elements of minus the inverse of the Hessian of the log likelihood.

† The χ^2 goodness-of-fit statistics were computed for the conditional (on sectoral choice) log-wage distributions in each sector using 51 equispaced log-wage intervals starting from log .75 in the intervals of length .07535 and terminating at log 35.0. The statistics compare predicted and actual log-wage distributions in each interval, integrating out the regressor variables. In computing the χ^2 statistics we account for parameter estimation error following Heckman (1984). We pooled 1976 and 1980 data to perform the test.

- Reject

The 3-Sector Extended Lognormal Roy Model

Empirical Estimates

Parameter Estimates of the Lognormal 3-Sector Model

A. Estimates

	Estimated Coefficient	Standard Error*
Utility function in the nonmanufacturing sector (γ_1):		
Intercept	4.851587	.443123
Education	.302087	.033746
Experience	.224618	.024399
Experience squared/100	−.297119	.065548
South dummy	.311855	.164999
Predicted nonlabor income/100	.247142	.029371
1980 intercept ($\gamma_{01\ell}$ for 1980)	.121791	.083692
Utility function in the manufacturing sector (γ_2):		
Intercept	3.321221	.599132
Education	.265219	.069635
Experience	.148654	.045367
Experience squared/100	−.340256	.085867
South dummy	.021460	.022615
Predicted nonlabor income/100	.146096	.091641
1980 intercept ($\gamma_{02\ell}$ for 1980)	.023336	.074984
Correlation coefficient between ν_1 and ν_2, corr(ν_1, ν_2)	.313598	.126617
SD of ν_2 $[\text{var}(\nu_2)]^{1/2}$	.960441	.194522

Parameters of the mapping of the observed skills to the nonmanufacturing task (β_1):		
Intercept	−.216854	.193443
Education	.052312	.008900
Experience	.013192	.012199
Experience squared/100	−.034313	.039652
South dummy	.182352	.154673
Parameters of the mapping of the observed skills to the manufacturing task (β):		
Intercept	−.320217	.200301
Education	.086914	.009140
Experience	.046019	.036089
Experience squared/100	−.134138	.019652
South dummy	.014589	.044673
Covariance structure of the latent task distribution:		
$(\sigma_{11})^{1/2} = [\mathrm{var}(u_1)]^{1/2}$	.526106	.008341
$(\sigma_{22})^{1/2} = [\mathrm{var}(u_2)]^{1/2}$	.510861	.070528
$\rho_{12} = \mathrm{corr}(u_1, \nu_2 - \nu_1)$	.221705	.338960
$\rho_{11} = \mathrm{corr}(u_1, \nu_1)$	.423159	.366167
$\rho_{21} = \mathrm{corr}(u_2, \nu_2 - \nu_1)$	.202134	.011768
$\rho_{22} = \mathrm{corr}(u_2, \nu_2)$	.144388	.005359
1980 estimated log task price change where $\pi_1\,(1976) = \pi_2\,(1976) = 1$:		
Nonmanufacturing sector ($\ln \pi_{1\ell}$ for 1980)	−.102087	.058032
Manufacturing sector ($\ln \pi_{2\ell}$ for 1980)	−.195462	.013046
Log likelihood for the model	−2,105.71	
No. of individuals in the sample	3,262	

B. Tests of the Model

	χ^2 Statistic for the Hypothesis	No. of Degrees of Freedom	Significance Level for the Test-Statistic
Strong proportionality hypothesis	21.3	22	.5023
Goodness-of-fit for the lognormal 3-sector model:†			
Manufacturing	42.7	50	.7585
Nonmanufacturing	71.9	50	.0229

* SEs are computed from the square root of the diagonal elements of minus the inverse of the Hessian of the log likelihood.

† The χ^2 goodness-of-fit statistic were computed for the conditional (on sectoral choice) log-wage distributions in each sector using 51 equispaced log-wage intervals starting from log .75 in intervals of length .07535 and terminating at log 35.0. The statistics compare predicted and actual log-wage distributions in each interval, integrating out the regressor variables. In computing the χ^2 statistics we account for parameter estimation error following Heckman (1984). We pooled 1976 and 1980 data to perform the test.

- Close in

Single Skill Box-Cox Model

Empirical Estimates

Parameter Estimates of the Participation-Nonparticipation Box-Cox 2-Sector Model

A. Estimates

	Estimated Coefficient	Standard Error*
Utility function in the market sector (γ_1) (nonmarket utility normalized to zero):		
Intercept	−3.222193	.078399
Education	.326871	.009711
Experience	.131634	.037425
Experience squared/100	−.063985	.005093
South dummy	.110597	.038274
Predicted nonlabor income/100	−.016929	.014889
1980 intercept ($\gamma_{0\ell}$ for 1980)	.956770	.402147

Parameters of the wage equation (β):		
Intercept	−.225465	.038934
Education	.084014	.022800
Experience	.068398	.001503
Experience squared/100	−.114323	.033552
South dummy	−.169693	.039298
1980 constant	1.654688	.075344
Covariance structure of the latent task distribution:		
$\sigma^{1/2} = [\text{var}(u)]^{1/2}$	.784779	.279160
$\rho = \text{corr}(u, \nu_1)$	.145953	.056628
1980 estimated log task price change where π (1976) = 1 (ln π for 1980):	−.217606	.098217
Task Transformation Parameter (λ)	.225259	.071395
Log likelihood for the model	−3,249.20	
No. of individuals in the sample	3,262	

B. Tests of the Model

	χ^2 Statistic for the Hypothesis	No. of Degrees of Freedom	Significance Level for the Test-Statistic
Likelihood-ratio test for restricted model ($\lambda = 0$) for wage equation	45.3	1	.0001
Goodness-of-fit for the single-skill model†	349.5	50	.0001
Strong proportionality hypothesis	153.1	10	.0001

* SEs are computed from the square root of the diagonal elements of minus the inverse of the Hessian of the log likelihood.

† The χ^2 goodness-of-fit statistics were computed for the conditional (on sectoral choice) log-wage distributions in each sector using 51 equispaced log-wage intervals starting from log .75 in intervals of length .07535 and terminating at log 35.0. The statistics compare predicted and actual log-wage distributions in each interval, integrating out the regressor variables. In computing the χ^2 statistics, we account for parameter estimation error following Heckman (1984). We pooled 1976 and 1980 data to perform the test.

- Reject

Return to main text