ELSEVIER

Labour Economics 5 (1998) 231–242

LABOUR ECONOMICS

Employee labor market information: comparing direct world of work measures of workers' knowledge to stochastic frontier estimates

Solomon W. Polachek [a,*], John Robst [b]

[a] *Department of Economics and Dean of Harpur College, State University of New York at Binghamton, Binghamton, NY 13902, USA*

[b] *Department of Economics, State University of New York at Binghamton, Binghamton, NY 13902, USA*

Received 3 December 1996; accepted 5 August 1997

Abstract

A number of papers use stochastic frontier estimation to measure a worker's incomplete information about available wages. These papers define incomplete information as the difference between a worker's wage and his or her maximum potential wage. Many question this approach since it essentially measures incomplete information as a residual, without independent evidence relating this residual to incomplete information. This paper introduces independent direct measures of workers' knowledge of the world of work obtained from the National Longitudinal Survey of Young Men (NLSYM). Frontier estimates of incomplete information are compared to the direct measures of workers' knowledge. The results verify that stochastic frontier estimates provide a reasonable measure of a worker's incomplete wage information. © 1998 Elsevier Science B.V. All rights reserved.

JEL classification: J0; J3

Keywords: Stochastic frontier estimation; Wage; National Longitudinal Survey of Young Men (NLSYM)

* Corresponding author. Harpur College Dean's Office, SUNY-Binghamton, Binghamton, NY 13902. Tel.: +1-607-777-2144; fax: +1-607-777-4900; e-mail: bg2776@bingvmb.cc.binghamton.edu.

PII S0927-5371(97)00030-4

1. Introduction

To check the validity of search theory's implications one ideally should have a measure of a worker's information about prospective wages. Despite the importance of such a measure, current studies devote little attention to developing and testing measures of worker information. Only one set of studies obtains measures of such incomplete worker information by using a stochastic frontier methodology to estimate the maximum attainable wage for a worker (Hofler and Polachek, 1982, 1985; Polachek and Yoon, 1987, 1996; Hofler and Murphy, 1992, 1994; and Groot and Oosterbeek, 1994). Because costless information implies workers are able to locate the job paying the highest wage commensurate with his or her human capital, these studies attribute the difference between a worker's actual and maximum attainable wage to imperfect information. These studies find that workers with more education, job tenure, who are male, white, union members, and who received unemployment benefits have more wage information and hence suffer less from incomplete labor market information.

While the results obtained using this measure are consistent with search theory, there has been no testing of whether this frontier measure accurately depicts information. One possible problem with the stochastic frontier approach is that incomplete worker information is essentially measured as a residual. As such, one might be skeptical whether the estimates derived from stochastic frontier estimation really do represent a measure of incomplete information. To rectify this deficiency one should ideally have direct measures of a worker's knowledge of labor market opportunities so they can be compared to estimates based on residuals. This paper utilizes such measures of worker information based on the knowledge of the world of work test administered to respondents in the National Longitudinal Survey of Young Men (NLSYM). These measures are then compared to incomplete information estimates based on the stochastic frontier framework. Support for the stochastic frontier method is provided if the stochastic frontier measures of information are positively correlated with the direct information measures in the NLSYM.

Section 2 describes the frontier approach to estimating worker information. Section 3 describes data from the world of work test as well as the data to implement frontier estimation. The results are provided in Section 4. Section 5 contains concluding remarks.

2. Stochastic estimation of incomplete information

Individuals who earn less than they could given their observed and unobserved human capital have incomplete information. Estimating the effect of incomplete information, is based on econometric techniques developed by Aigner et al.

(1977).[1] To get at what an individual could earn, denote W_i^{p} to be individual i's potential wage. One's potential wage is dependent upon human capital and other characteristics, and thus can be denoted as:

$$W_i^{\mathrm{p}} = \boldsymbol{X}_i \beta + \nu_i, \tag{1}$$

where the dependent variable represents the individual's maximum potential wage offer; $\boldsymbol{X}$ is a vector of independent variables; and ν is a disturbance distributed as $N(0,\sigma_\nu^2)$. With perfect information about what each firm pays, a worker would be able to locate a job paying his or her potential wage at no cost.[2] However, in a world of incomplete information search is expensive, implying that the typical worker usually accepts an offer (W_i^{o}) that is below his or her potential wage. The literature using a stochastic frontier framework to estimate incomplete worker information defines the difference between the potential wage and the actual accepted offer to be an index of a worker's incomplete information. As such, the offer wage one accepts can be specified as:

$$W_i^{\mathrm{o}} = W_i^{\mathrm{p}} - u_i, \tag{2}$$

where $u_i \geq 0$. The 'one-sided' residual u_i represents an estimate of incomplete information. Combining Eqs. (1) and (2) leads to:

$$W_i^{\mathrm{o}} = X_i \beta + \epsilon_i, \tag{3}$$

where $\epsilon_i = \nu_i - u_i$.

Typically the maximum likelihood approach proposed by Aigner et al. (1977) is used to estimate the earnings frontier. The techniques of Jondrow et al. (1982) can be used to obtain two individual specific estimates of incomplete information:

$$E(u_i/\epsilon_i) = \frac{\sigma_u^2 \sigma_\nu^2}{\sigma_2}\left[\frac{f(\epsilon_i \lambda \sigma)}{1 - F(\epsilon_i \lambda \sigma)} - \left(\frac{\epsilon_i \lambda}{\sigma}\right)\right], \tag{4}$$

and

$$\begin{aligned} M(u_i/\epsilon_i) &= -\epsilon_i\left(\sigma_u^2/\sigma^2\right) \quad &&\text{if } \epsilon_i \leq 0 \\ &= 0 &&\text{if } \epsilon_i > 0, \end{aligned} \tag{5}$$

where $E(u_i/\epsilon_i)$ is the mean and $M(u_i/\epsilon_i)$ is the mode for each individual.

Estimates of incomplete information are converted to measures of information by computing the percentage of potential wages received by each worker. Workers

---

[1] With panel data, one could perform this estimation taking into account unobserved characteristics (Polachek and Yoon, 1996). However, this approach is not necessary here since we are merely interested in testing the validity of currently used residual estimates of worker incomplete information by comparing them to direct measures.

[2] Were data available, one could generalize and talk not just about wages, but the whole amenity package offered by firms.

with more information will have a figure closer to one, while workers with less information will be closer to zero. The relationships between individual information and various worker characteristics are then considered by comparing average levels of information for various subgroups of workers. This type of analysis was performed by Hofler and Polachek (1982, 1985), Polachek and Yoon (1987), and Hofler and Murphy (1992, 1994). [3]

This approach may be problematic when used to examine differences in information between workers since the two-stage procedure is inconsistent in its assumption that the inefficiency effects are independent in the two stages (Coelli, 1994). Groot and Oosterbeek (1994) also note that this technique ignores possible correlations between different groups of workers. They simultaneously estimate the frontier, and model incomplete information as a function of individual characteristics. We follow a similar approach using a method developed by Battese and Coelli (1993, 1995) and Coelli (1994).

The approach is similar to the Aigner, Lovell, and Schmidt model except the stochastic frontier and inefficiency effects are modelled as a function of individual specific factors in a one-step procedure. Incomplete information, u_i, is a function of a vector of variables that accounts for incomplete information and is specified as:

$$U_i = z_i \boldsymbol{\delta} + \alpha_i \tag{6}$$

where, for the purposes of our test, z_i are the independent measures of information in the NLSYM, and α_i is a random variable defined by the truncation of the normal distribution with zero mean and variance σ^2. If the vector $\boldsymbol{\delta}$ equals zero the model reduces to the Aigner, Lovell, and Schmidt model. Maximum likelihood techniques are used to estimate the frontier and inefficiency parameters. [4] A negative relationship between u_i and z_i indicates support for the use of the frontier measure of information.

Individual specific information is calculated as:

$$\text{Info}_i = 1 - (z_i \boldsymbol{\delta} + \alpha_i) \tag{7}$$

The individual specific estimates can be used to compare information between subgroups of workers.

[3] Polachek and Yoon (1987, 1996) and Groot and Oosterbeek (1994) examine firm ignorance in addition to worker ignorance about the labor market. We concentrate on worker incomplete information since we do not have independent measures of firm information.

[4] Estimation is performed using Frontier Version 4.1. See Coelli (1992) for a description of Version 2.0 of the Frontier program, and Coelli (1994) details Version 4.1. The likelihood function is given in Battese and Coelli (1993).

3. Data

The 1966 wave of the NLSYM contains an independent measure of worker information. Each respondent was asked three sets of multiple choice questions about the labor market. The first ten questions gave various job titles and determined if the respondent could choose the correct job description. The ten occupations were hospital orderly, machinist, acetylene welder, stationary engineer, statistical clerk, fork lift operator, economist, medical illustrator, draftsman, and social worker. The next ten questions asked how much education was required for the same ten jobs. The last eight questions asked the respondent to select the highest paying occupation from among two choices. The eight comparisons were auto mechanic/electrician, medical doctor/lawyer, aeronautical engineer/medical doctor, truck driver/grocery store clerk, unskilled laborer in steel mill/in shoe

Table 1
Variable means

	Mean	Standard deviation
Knowledge	36.33	7.38
Knowledge–occupations	15.09	3.61
Knowledge–education	8.23	3.65
Knowledge–earnings	13.01	2.23
Hourly wage	2.42	3.23
Education	12.02	1.65
Experience	3.01	2.37
White	0.82	0.39
SMSA	1.94	0.80
Health	0.11	0.31
Hours	45.14	12.14
Test score	95.98	14.11
Professional	0.10	0.30
Managerial	0.03	0.18
Clerical	0.11	0.31
Sales	0.04	0.19
Crafts	0.20	0.40
Operatives	0.35	0.48
Service	0.07	0.25
Laborers	0.12	0.32
Observations	991	

Data: NLS 1966; Variable definitions: Knowledge = the total score on the world of work test; knowledge–occupations = the score on the occupational descriptions test; knowledge–education = the score on the education requirements test; knowledge–earnings = the score on the earnings comparison test; hourly wage = hourly wage; education = years of completed schooling; experience = (age–education–6); white = dummy variable, 1 = white; SMSA: 1 = SMSA, central city, 2 = SMSA, not in central city, 3 = not in SMSA; health = dummy variable, 1 = health limits work; hours = hours worked per week; test score = score on IQ test; the occupational dummy variables denote the individual's occupation.

Table 2
Regression results

	OLS	Stochastic frontier estimation			
	(1)	(2)	(3)	(4)	(5)
Stochastic frontier					
Intercept	−0.4581** (0.137)	−0.2266 (0.217)	−0.3225** (0.143)	−0.2384** (0.030)	−0.4532** (0.172)
Education	0.0757** (0.010)	0.0686** (0.009)	0.0699** (0.010)	0.0678** (0.009)	0.0758** (0.009)
Experience	0.1075** (0.015)	0.1120** (0.027)	0.1000** (0.015)	0.1137** (0.011)	0.1076** (0.015)
Experience-squared	−0.0068** (0.002)	−0.0076** (0.004)	−0.0058** (0.002)	−0.0074** (0.002)	−0.0068** (0.002)
SMSA	−0.0849** (0.015)	−0.0865** (0.023)	−0.0766** (0.015)	−0.0895** (0.008)	−0.0850** (0.015)
White	0.1525** (0.035)	0.1546** (0.070)	0.1329** (0.035)	0.1161** (0.034)	0.1528** (0.035)
Test score	0.0033** (0.0010)	0.0027** (0.001)	0.0031** (0.001)	0.0032** (0.001)	0.0034** (0.001)
Health	−0.0595 (0.038)	−0.0588 (0.038)	−0.0577 (0.038)	−0.0553 (0.037)	−0.0594 (0.038)
Hours	−0.0075** (0.001)	−0.0076** (0.001)	−0.0073** (0.001)	−0.0079** (0.001)	−0.0075** (0.001)
Professional	0.1968** (0.058)	0.1883** (0.050)	0.1663** (0.056)	0.2189** (0.016)	0.1968** (0.058)
Managerial	0.1767** (0.075)	0.1688** (0.074)	0.1778** (0.074)	0.1553** (0.072)	0.1768** (0.076)
Clerical	0.0628 (0.051)	0.0604 (0.051)	0.0615 (0.051)	0.0536 (0.050)	0.0635 (0.052)
Sales	0.0776 (0.071)	0.0778 (0.070)	0.0842 (0.070)	0.0777 (0.068)	0.0777 (0.072)
Crafts	0.2155** (0.045)	0.2159** (0.051)	0.2077** (0.045)	0.1895** (0.044)	0.2152** (0.046)
Operatives	0.2387** (0.042)	0.2426** (0.055)	0.2258** (0.041)	0.2519** (0.031)	0.2392** (0.041)
Service	−0.0607 (0.058)	−0.0799 (0.057)	−0.0797 (0.057)	−0.0814 (0.056)	−0.0600 (0.059)

Incomplete information model					
Constant	–	1.023** (0.187)	0.4099* (0.213)	0.1770 (0.160)	−0.1129 (0.347)
Knowledge	–	−0.0679** (0.006)	–	–	–
			–	–	–
Knowledge–occupation	–	–	−0.1951** (0.022)	–	–
Knowledge–education	–	–	–	−0.0983** (0.033)	–
Knowledge–earnings	–	–	–	–	0.0029 (0.019)
Variance parameters					
$\tau = \sigma_u / \sigma_\nu$	–	0.4790** (0.217)	0.5752** (0.030)	0.2634** (0.026)	0.0056 (0.083)
$\sigma^2 = \sigma_\nu^2 + \sigma_u^2$	–	0.2055** (0.082)	0.3134** (0.012)	0.1604** (0.003)	0.1339** (0.011)
N	991	991	991	991	991
R^2	0.2838				
Log-likelihood	−406.7	−377.2	−370.4	−386.5	−406.7

Data: NLS 1966, Dependent variable: log (hourly wage), Standard errors are in parentheses.

** Indicates significance at the 5% level.

* Indicates significance at the 10% level.

Variable definitions: Knowledge = the total score on the World of Work Test; knowledge–occupations = the score on the occupational descriptions test; knowledge–education = the score on the education requirements test; knowledge–earnings = the score on the earnings comparison test; log (hourly wage) = natural logarithm of hourly wages; education = years of completed schooling; experience = (age–education–6); white = dummy variable, 1 = white; SMSA: 1 = SMSA, central city, 2 = SMSA, not in central city, 3 = not in SMSA; health = dummy variable, 1 = health limits work; hours = hours worked per week; test score = score on IQ test, the occupational dummy variables denote the individual's occupation.

factory, lawyer/high school teacher, high school teacher/janitor, and janitor/policeman. A total score was determined that measures the person's knowledge of the world of work.

The validity of the world of work test has been addressed by Parnes and Kohen (1973) and Kohen and Breinich (1975). Parnes and Kohen (1973) found the measure to be a significant predictor of future earnings and occupational status. Also, individuals expected to perform better on the test appear to have scored higher on the test. Individuals with more schooling, a higher score on an IQ test, more work experience, and with better socioeconomic backgrounds tended to have higher world of work test scores. Kohen and Breinich (1975) perform correlation, factor, and discriminant analyses on the measure of knowledge. They conclude that overall, the measure performs well and 'leads us to commend its use, with some modification, to those involved in assessing and counselling the occupational choice process.' The primary modification they address is to either reform or drop the earnings comparison questions. They concluded that these questions were too easy to provide any substantive information.

Individuals interviewed in the NLSYM range from ages 14 to 24 in 1966. The sample is limited to the initial year of the survey and includes 991 workers who are not enrolled in school. We exclude workers enrolled in school, since they are unlikely to be fully utilizing their information about the labor market. Sample means are presented in Table 1. Given the young age of the sample, the level of potential experience (age–education–6) is small at 3.01 years. The average level of education is 12.01 years, average hourly wages are US$2.42, and the mean test score is 95.98. The average knowledge of the world of work score is 36.32 out of a maximum of 56. Workers appear to have scored the best on the earnings comparisons as the average score was 13 out of a possible 16, although this may simply indicate, as discussed above, that these questions have limited value. A large majority of the sample (82%) answered at least six of the eight questions correctly. The score on the questions about job descriptions for occupations was approximately 15 out of a maximum of 20. The education score was 8.23 out of 20.

4. Results

OLS and maximum likelihood results are provided in Table 2. For the purposes of this paper we estimate several different frontier and inefficiency specifications to compare residual based incomplete information measures (u) to knowledge of work measures in the NLS. It should be noted that the frontier will vary depending on the choice of NLS measure of information (z) as the frontier and the model of incomplete information are estimated simultaneously. Column (1) reports the OLS results. As is typical, wages are positively related to education, experience, living in an SMSA, being white, and ability as measured by the test score. However, because the data consist of workers early in their careers, the returns to experience are relatively large.

We find a significant positive relationship between the knowledge of the world of work measures and the information measure based on stochastic frontier estimation. The first frontier specification compares the total knowledge of the world of work score with the stochastic frontier estimates of incomplete information. These results are reported in column (2). Workers with more knowledge of the world of work have less incomplete information as they earn a greater percentage of their potential wages.

We also find an inverse relationship between the frontier measure of incomplete information and the independent measures of information for two of the three types of questions asked of workers. The first set of questions asked workers to select the correct job description for ten occupations. The results in column (3) show that workers with a higher score on these questions have significantly less incomplete information. The second set of questions asked workers the education required for the same ten occupations. Again, the results in column (4) show a negative relationship between the frontier measure of incomplete information and the independent world of work measure of information. The last set of questions asked workers to select the higher paying occupation from among two choices. The results in column (5) do not find a significant relationship between incomplete information and the NLS measure of information. There is very little variation in this measure of information and as a result, this specification predicts that workers have complete information. As discussed earlier, the earnings comparison has been questioned as a useful measure of information due to the ease of the questions. Thus it may be a positive result that there is no statistically significant relationship between the frontier measure of information and the score on the earnings comparisons.

Similar to previous studies, we also tested for differences in information between subgroups of workers. Workers are sorted based on the NLSYM measures of information and comparisons of the frontier estimates of information are made between subgroups. Again, significant positive relationships are found between the frontier measures of information and the total score on the World of Work test, the occupational descriptions test, and the educational requirements test. However, a negative relationship is found between the frontier measure of information and the earnings comparison test. These results are provided in Appendix A.

5. Conclusion

Measuring a worker's knowledge of the labor market is receiving increased attention in the literature. Early studies rely strictly on Stiglerian notion of wage (or price) variance (e.g., Stigler and Kindahl, 1970; Von Hoomissen, 1988; and

Lach and Tsiddon, 1992), but are defficient because wages vary for a wide variety of reasons, including differences in worker characteristics such as education and experience. Recent studies provide a measure of labor market information held by workers based on stochastic frontier estimation. Both measures have been criticized because they are residuals and thus not directly linked to information. While conclusions using these measures of information are consistent with search theory, to this point no direct test has been made to determine if the measure is a true measure of information. This paper provides evidence of a positive relationship between the stochastic frontier estimates of information and independent measures of information available in the NLSYM. Thus, this paper provides new and different evidence that the stochastic frontier estimates represent a reasonable measure of information about the labor market.

Acknowledgements

We wish to thank Tim Coelli for helpful answers to questions regarding the Frontier 4.1 computer program.

Appendix A

A.1. Comparison of measures of information

	N	Average information	Standard deviation
Knowledge— number correct on occupational description questions			
Maximum score = 20			
(1) 0–14	387	0.9126	0.0394
(2) 16	238	0.9414	0.0075
(3) 18–20	366	0.9487	0.0079
Total	991	0.9328	0.0302
t-statistics (differences in means)			
(1) vs. (2)	13.97 * *		
(2) vs. (3)	11.44 * *		
(1) vs. (3)	17.64 * *		
Knowledge— number correct on educational requirements questions			
Maximum score = 20			
(1) 0–6	337	0.9006	0.0429

(2) 7–9	293	0.9439	0.0093
(3) 10–20	361	0.9619	0.0072
Total	991	0.9357	0.0369
t-statistics (differences in means)			
(1) vs. (2)	18.03**		
(2) vs. (3)	27.01**		
(1) vs. (3)	25.84**		
Number correct on earnings comparison questions			
Maximum score = 16			
1. 0–10	176	0.9925	0.0002
2. 12	295	0.9920	0.0001
3. 14–16	520	0.9913	0.0003
Total	991	0.9917	0.0005
t-statistics (differences in means)			
(1) vs. (2)	−23.92**		
(2) vs. (3)	−41.12**		
(1) vs. (3)	−48.16**		
Knowledge of the world of work score—total number correct			
Maximum score = 56			
1. 0–33	310	0.9056	0.0453
2. 34–39	323	0.9406	0.0117
3. 40–56	358	0.9537	0.0084
Total	991	0.9344	0.0334
t-statistics (differences in means)			
(1) vs. (2)	13.20**		
(2) vs. (3)	16.52**		
(1) vs. (3)	18.40**		

Data: NLS 1966, Mean is the average worker information for the specified group.

** Indicates significance at the 5% level.

* Indicates significance at the 10% level.

References

Aigner, D.C., Lovell, C.A.K., Schmidt, P., 1977. Formulation and estimation of stochastic frontier production function models. J. Econometrics 5, 21–38.

Battese, G.E., Coelli, T.J., 1993. A Stochastic Frontier Production Function Incorporating a Model for Technical Inefficiency Effects. Working Papers in Econometrics and Applied Statistics No. 69. Department of Econometrics, University of New England, Armidale, Australia.

Battese, G.E., Coelli, T.J., 1995. A model for technical inefficiency effects in a stochastic frontier production function for panel data. Empir. Econ. 20, 325–332.

Coelli, T., 1992. A computer program for frontier production function estimation: Frontier Version 2.0. Econ. Lett. 39, 29–32.

Coelli, T., 1994. A Guide to Frontier Version 4.1: A Computer Program for Stochastic Frontier Production and Cost Function Estimation. Department of Econometrics, University of New England, Armidale, Australia.

Groot, W., Oosterbeek, H., 1994. Stochastic reservation and offer wages. Labour Econ. 1, 383–390.

Hofler, R.A., Murphy, K.J., 1992. Underpaid and overworked: measuring the effects of imperfect information on wages. Econ. Inq. 30, 511–529.

Hofler, R.A., Murphy, K.J., 1994. Estimating reservation wages of workers using a stochastic frontier. South. Econ. J. 60, 961–976.

Hofler, R.A., Polachek, S.W., 1982. Ignorance in the Labor Market: A New Approach for Measuring Information Content. Proceedings of the 1982 Business and Statistics Section (American Statistical Association, Washington DC) pp. 422–425.

Hofler, R.A., Polachek, S.W., 1985. A new approach for measuring wage ignorance in the labor market. J. Econ. Bus. 37, 267–276.

Von Hoomissen, T., 1988. Price dispersion and inflation: evidence from Israel. J. Pol. Econ. 96, 1303–1314.

Jondrow, J., Lovell, C.A.K., Materov, I., Schmidt, P., 1982. On the estimation of technical inefficiency in the stochastic frontier production function model. J. Economet. 19, 233–238.

Kohen, A.I., Breinich, S.C., 1975. Knowledge of the world of work: a test of occupational information for young men. J. Vocat. Behav. 6, 133–144.

Lach, S., Tsiddon, D., 1992. The behavior of prices and inflation: an empirical analysis of disaggregated price data. J. Pol. Econ. 100, 349–389.

Parnes, H.S., Kohen, A.I., 1973. Occupational Information and Labor Market Status: The Case of Young Men. Center for Human Resource Research, The Ohio State University, Columbus.

Polachek, S.W., Yoon, B.J., 1987. A two-tiered earnings frontier estimation of employer and employee information in the labor market. Rev. Econ. Stat. 69, 296–302.

Polachek, S., Yoon, B.J., 1996. Panel estimates of a two-tiered earnings frontier. J. Appl. Economet. 11, 169–178.

Stigler, G., Kindahl, J., 1970. The Behavior of Industrial Prices. Columbia Univ. Press, New York.