International Labour Underpayment: A Stochastic Frontier Comparison of Canada and the United States

Kenneth G. McClure and Paul B. Girma
Department of Business
Skidmore College
Saratoga Springs, NY 12866-1632

Richard A. Hofler
College of Business Administration
University of Central Florida
Orlando, Florida 32816

The search theory of unemployment suggests that workers, after expending time and resources in the search process, may accept a position that provides compensation at a level below that which is appropriate to the worker's true capability. In this regard the worker is "underpaid" in the sense that there is unused worker capacity which could be productively employed if there was a perfect match between the worker and the actual job accepted.[1] This underpayment occurs because information about jobs is costly to acquire, because searchers possess positive discount rates, and because searchers have time horizons of finite length. As a result, searchers stop their job hunting before discovering the highest-paying job for which they are qualified.

From an overall United States economic perspective, the degree of underpayment represents lost output (reduced GDP). Since this loss is a result of inefficient matches in the labour market, the issue of federal policies to reduce such inefficiencies and improve national output may be important. Whether this issue is nationally important depends on the answers to two principal questions:

- Does the amount of underpayment represent a major loss in U.S. GDP?
- Does the amount of underpayment in the United States compare favourably or unfavourably with our principal international competitors for the Global Market?

If the answer to either of these questions suggests policy action by the U.S. government may be productive, appropriate policy alternatives should be considered. Such alternatives might be a national/regional "Job Bank" of readily available up-to-date employment information (to reduce the information collection time) and/or increased unemployment benefits (to reduce the marginal search costs to the worker and make the job seeker more selective in final employment choice).

Hofler and Murphy (1992) [hereafter H&M (1992)] apply the stochastic frontier technique originated by Aigner et al (1977) to the measurement of underpayment in the United States. H&M (1992) use the 1983 Current Population Survey data (6439 observations) to address the first of the two questions posed above. Their results indicate the average U.S. worker in 1983 had 9.57 percent underpayment.

The primary purpose of this paper is to extend the research of H&M (1992) to the second question above regarding the comparative performance of the U. S. in the international labour market, specifically Canada. A secondary purpose is to compare the performance of the U.S. at different times (1983 vs. 1986).

Search Theory and Underpayment

We assume that employment seekers use a sequential stopping method in their search process.[2] This approach presumes that each job seeker understands the lower order parameters of the wage distribution but does not possess job-specific information. To obtain the missing information the person searches for a job. Each job offer received during the search process must be accepted or rejected at the time of offer and the decision cannot be deferred or reconsidered at a latter date (the worker cannot "recall" an offer). The worker's problem is to decide when to stop the sampling process by accepting an offer. The degree to which the wage of the accepted job is different from the wage of the job which perfectly matches the skills of the worker is underpayment. This represents lost economic output to the nation.

McCall (1970), Mortensen (1970) and Lippman and McCall (1976) propose that under these conditions the optimal tactic for the job seeker is to adopt a reservation wage strategy. That is, the job seeker sets a reservation wage $w(r)$ so that the first offer equal to or greater than $w(r)$ will be accepted at wage $w(a)$ and all lower offers will be rejected. The worker determines the reservation wage $w(r)$ by equating the marginal benefits and the marginal costs of incremental changes in $w(r)$.

The wage which an employer will pay is assumed to be a positive function of the required worker skills (q) for proper job performance. The higher the required skills, the higher the wage that an employer will pay to a worker with those required skills. Therefore, the potential wage for any worker is based on required skills is $w(q)$. Since required job skills will differ by employer the potential wage [$w(q)$] offered by employers (and for specific jobs) will also differ.

Each worker possesses a unique set of skills (*qo*) which represents the worker's accumulated human capital and predetermines the highest possible compensation available (such as, the potential wage *w*(*qo*)). Given the worker's human capital (*qo*) and the availability of appropriate employment paying *w*(*qo*), the optimisation process for the worker is to match his or her skills with various jobs until the potential wage *w*(qo) is found. In a perfect world with costless information this employment in the worker's most productive use could be found. However, information is not costless. Therefore, the worker will follow the reservation wage strategy until an acceptable position paying $w(a) \geq w(r)$ is found. Underpayment for the individual worker is represented by the difference between the potential wage and actual wage for the position accepted:

$$UNDERPAYMENT = W_{(QO)} - W_{(A)} \geq 0 \qquad (1)$$

From Figure 1 above, the area under the density function *f*(*w*) between *w*(*r*) and *w*(*qo*) represents the job seeker's probability of finding acceptable employment. Above *w*(*qo*) the worker does not possess the job skills required by the employer (for example, is not qualified for the position) and will theoretically not receive an offer of employment in a competitive market. This probability (*Pr*) of finding an acceptable job (*Ja*) is:

$$PR_{(JA)} = \int_{W(R)}^{W(QO)} F(W)DW \qquad (2)$$

and given a reservation wage *w*(*r*), the conditional expected job seeker's wage will be:

$$E\ [W|\,W \geq W_{(R)}] = \frac{\int_{W(R)}^{W(QO)} WF(W)DW}{\int_{W(R)}^{W(QO)} F(W)DW} \qquad (3)$$

From the above it is clear that, given a distribution of available jobs and wages, the expected wage will be dependent on the costly information obtained by the worker to set the reservation wage *w*(*r*). This expected wage will fall between $w(a) \geq w(r)$. It also follows that a larger reservation wage leads to both a larger expected actual wage and a smaller degree of underpayment. (Appendix A presents the proof of this statement). Therefore, factors that systematically determine reservation wage rates will, in turn, determine the degree of underpayment.

As noted above, the reservation wage is determined by equating, at the margin, the benefits and costs of a further increment to the reservation wage, as in Figure 2.

The benefits of a further increment to the reservation wage are greater lifetime earnings once employment is secured. The height of the marginal benefit curve in the graph depends on the usual factors such as the searcher's discount rate and the amount of time the searcher expects to remain employed.

The cost of a further increment to the reservation wage consists of higher out of pocket costs and additional foregone earnings resulting from a longer duration of unemployment. The height of the marginal cost curve depends on the worker's search efficiency, the worker's skill level (which determines the opportunity cost), the worker's wealth level, and the availability and amount of social welfare payments (such as unemployment benefits) that can be collected during a spell of unemployment.

FIGURE 1 Theoretical Wage Distribution for Job Seeker

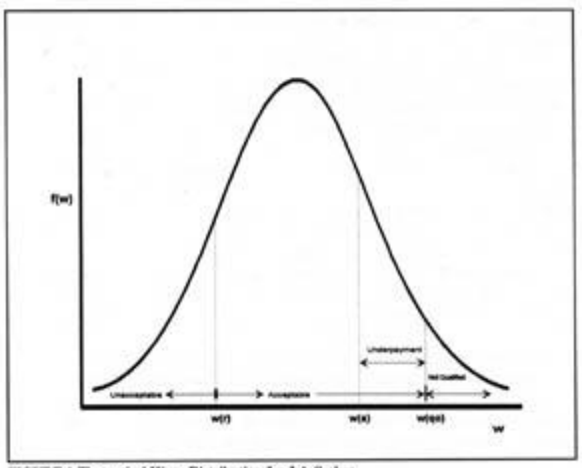

FIGURE 1 Theoretical Wage Distribution for Job Seeker

FIGURE 2 Reservation Wage Determination for Job Seeker

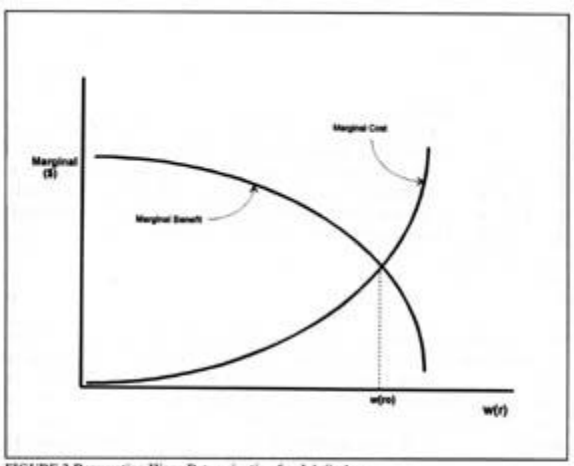

FIGURE 2 Reservation Wage Determination for Job Seeker

Hypotheses

Workers differ in terms of the various factors that determine the marginal benefits and marginal costs of increments to the reservation wage. Because of this, the search model discussed earlier implies many testable hypotheses about differences in the degree of underpayment found across workers. In the empirical portion of this paper, we test four hypotheses. [3]

Hypothesis 1

Demographic groups with high rates of time preference will exhibit more underpayment than groups with low rates of time preference. This hypothesis intimates that prime-age males will suffer less underpayment than other groups such as females and young workers. Such differences could arise because the former group is less hard pressed to work sooner than later, expects to work longer in the next job, and faces a more extensive geographic labour market.

Hypothesis 2

More-educated workers will exhibit less underpayment. This prediction follows for several reasons, two of which follow. First, higher education reveals patience in investing in schooling, which is likely to spill over into patience (and greater selectivity) in seeking employment. Second, direct search costs faced by the more-educated job seekers are likely to be lower than for the less-educated persons for two reasons. They are likely to be more efficient searchers and they are likely to have access to better information networks.

Hypothesis 3

Workers in urban areas will exhibit less underpayment. Search costs, and therefore underpayment, are lower in urban locales than in rural area for two reasons. Firms are less dispersed geographically in urban locations and urban transport networks are generally better.

Hypothesis 4

Workers with greater wealth will exhibit less underpayment. Wealthier seekers can consume part of their wealth while job hunting and, thus, be more selective.

This study estimates and then uses the underpayment levels for the United States (in 1986) and Canada (in 1987) to empirically test these hypotheses.

Methodology

Stochastic Frontier Estimation Technique

The stochastic frontier model was originally presented in Aigner et al (1977). It has been expanded by Jondrow et al (1982) and many others (Bauer 1990; Battese 1992) for application to production maximisation and cost minimisation cases. Further developments of the technique for application with nonproduction and noncost applications have been made by Polachek and Hofler (1985), Hofler and Murphy (1989), and others (see, Bauer 1990). This study will estimate a frontier giving a potential (maximum possible) wage, *w*(*qo*), for each worker. Based upon this frontier, each individual's unique level of underpayment will be calculated.

Following H&M (1992), the relationship between *y* and *x* can be expressed as:

$$Y_i = \alpha + \beta X_i + \varepsilon_i \qquad (4)$$

Where:

ε_i is $N(0, \sigma^2)$

The above model assumes that it is the mean of *y* that is of interest to the researcher. However, the challenge in some economic research is to identify either the maximum possible value of *y* (such as, production or income) or its minimum possible value (such as, cost). In measuring underpayment, the goal is to calculate the maximum value that a job seeker's wage could take given that individual's accumulated human capital. We call this wage the "potential wage."

To estimate the "potential wage" through the use of the stochastic frontier technique requires two steps. The first step is to run a regression on all the data to obtain the normal parameters (α and βs) and to determine the individual values of ϵ_i for all observations. The second step is to rerun the regression on a restricted data set of all observations with $\varepsilon_i \geq 0$ (the top half of the data distribution). This new "stochastic frontier" represents an estimate of an empirical maximum for each observation of "potential wage." This frontier, $\hat{Y}_{i,STO}$, is the maximum value of y for one specific observation whose $x = x_i$. The stochastic frontier is:

$$\hat{Y}_{i,STO} = \alpha' + \beta X_i + \gamma_i \qquad (5)$$

Where:

γ_i = the two-sided error component where γ_i is $N(0,\sigma^2)$

Combining equations (4) and (5) with the entire data set allows the decomposition of the original error term into two constituent parts, a randomly distributed error term (Y_i) and one-sided component constrained to be equal to or less than zero $(\phi_i \leq 0)$

$$Y_I = \alpha + \beta X_I + \gamma_I + \phi_I \qquad (6)$$

Where:

ε_i = $\gamma_i + \phi_i$

γ_i = two-sided component where γ_i is $N(0, \sigma^2)$

ϕ_i = one-sided component where $\phi_i \leq 0$ and $E(\phi_i) = \mu < 0$

Equation (6) contains both a "deterministic frontier" and a "stochastic frontier." The deterministic frontier for observation *I* is

$$\hat{Y}_{I,DET} = \alpha' + \beta X_I = (\alpha - M) + \beta X_I \qquad (7)$$

Where:

$E(\phi_i) = \mu$

This represents the "average maximum" *y* for all observations at a specific level of *x*, say x_i.

The frontier in equation (6) is stochastic because, even if two observations have the same level of *x*, they likely will differ in the unmeasured factors captured in the two-sided error, *Y*. In such cases each observation will have its own individual frontier.

Using the stochastic frontier approach permits us to estimate $\hat{Y}_{i,STO}$ for each observation. This, in turn, permits us to estimate each individual's unique level of underpayment, equation (1), which is estimated by ϕ_i .

Based on the above, a two-step process is necessary to learn the individual levels of underpayment:

- The model's parameters are estimated by maximum likelihood.
- These parameters are used to estimate the individual levels of underpayment from ϕ_i , the one-sided error term.

The software LIMDEP will be used for this purpose.

Model Specification

The proper estimation model must be selected. Murphy and Welch (1990) [hereafter M&W (1990)], expand on the widely accepted Mincer (1974) earnings function to develop a model expressing worker income (defined as the logarithm of wage/period) as a quartic (a quadratic in a quadratic) function of accumulated human capital. The general model developed by M&W (1990) is:

$$Y_{IXT} = \alpha_{IT} + \beta_{1IT} Z_{IXT} + \beta_{2IT} Z_{IXT}^2 \qquad (8)$$

Where:

$$Z_{IXT} = \alpha' + \beta_1' X + \beta_2' X^2 \qquad (9)$$

And:

y = Income (ln wage/week)
I = Education (category)
x = Years of Experience
t = Time (years of observations)

After some manipulation, equation (8) becomes the following model for estimation (M&W 1990):

$$Y_{IXT} = \alpha + \beta_1 X + (\beta_2 + \beta_1 \Gamma) X^2 + 2\Gamma\beta_2 X^3 + \Gamma^2 \beta_2 X^4 \qquad (10)$$

Where:

$\Gamma = 1/60$ in M&W (1990).

M&W (1990) present empirical evidence on 1964-87 data which suggests that this restricted quartic function would perform better than commonly-used earnings functions. Equation (10) will be the deterministic part of the estimation model. In other words, equation (10) will be the $\alpha + \beta x$ part of equation (5) for estimating underpayment ϕ_i and for subsequent testing and analysis.

Estimation and Testing Procedure

As stated above, the model must first be estimated by maximum likelihood. According to Aigner et al (1977), the distribution of can be parameterised as:

$$f(\varepsilon) = \left(\frac{2}{\sigma}\right) f\left(\frac{\varepsilon}{\sigma}\right)[1 - F(\varepsilon \Gamma \sigma^{-1})] \qquad (11)$$

Where:

$\sigma^2 = \sigma^2_\phi + \sigma^2_\gamma$

$\Gamma = \sigma_\phi / \sigma_\gamma$

The log-likelihood function is

$$\text{LN } L = CONSTANT + N \text{ LN} \sigma^{-1} + \sum_{i=1}^{N} \text{LN}[1 - F(\varepsilon_i \Gamma \sigma^{-1})] - (1/2)\sigma^2 \sum_{i=1}^{N} \varepsilon_i^2 \qquad (12)$$

From (12) estimates for all coefficients, Γ, and σ^2 are obtainable. The parameters Γ and σ^2 can be solved for σ_ϕ^2 and σ_γ^2, which are useful in the next step discussed below.

This study's approach is to estimate the model once for each country and then compute person-specific ϕ_i estimates. The ϕ_i error component plays a crucial role in this analysis. ϕ_i will equal zero and underpayment will be zero, if the worker finds the perfect job-skills match; the actual wage will equal the potential wage for that worker. ϕ_i will be negative and underpayment will exist, if the worker finds an imperfect job-skills match; the actual wage will fall short of the potential wage for that worker. The work of Jondrow et al (1982) provides a technique for obtaining individual worker estimates of ϕ_i. Equation (13) below presents the conditional estimate of ϕ_i given ε_i:

$$E(\phi_i | \varepsilon_i) = \frac{\sigma_\phi \sigma_\gamma}{\sigma} \left[\frac{f\left(\frac{\varepsilon_i \lambda}{\sigma}\right)}{1 - F\left(\frac{\varepsilon_i \lambda}{\sigma}\right)} - \frac{\varepsilon_i \lambda}{\sigma} \right] \qquad (13)$$

Where:

σ^2	=	$\sigma^2_\phi + \sigma^2_\gamma$
λ	=	$\sigma_\phi / \sigma_\gamma$
f	=	the standard normal density function
F	=	the standard normal distribution function

With an estimate of ϕ_i in hand, the estimate for each individual's unique potential wage, $w(qo)_I$ is simply the individual's actual wage, $w(a)_I$ plus the individual's underpayment estimate, ϕ_i. From this point, testing various hypotheses for differing worker characteristics becomes straightforward.

Given the above potential wage, the actual wage divided by the potential wage for each individual can be calculated. This variable will be termed RATIO. RATIO measures each person's success in attaining his or her potential wage. Therefore, it should take values between 0% - 100%.[4] Mean values of RATIO can be calculated in total and for various subsets of workers so that Hypotheses 1 - 4 may be tested.

Standard pairwise t-tests (when only two subsets of data within a group are compared) and ANOVA F-tests (when more than two subsets of data are compared) will be used to test for significant differences (H_o: Mean RATIOs are equal) among sample subsets within the countries.[5] When there are more than two subsets of data (categories) within a group (for example, age and education each have more than two categories) an additional Waller-Duncan K-Ratio t-test for subset pairs and a Tukey Studentized t-test will be performed to identify significant differences between pairs of categories within the group. A level of significance of 5% will be employed in all tests for statistical significance.

The Data

The Luxembourg Income Study (hereafter LIS), originated in 1983 and based in Walferdange, Luxembourg represents the database. The LIS is a Division of CEPS/INSTEAD and represents an ongoing, cooperative consortium of researchers in the social sciences from eleven member nations.

From the LIS database, random samples for the U.S. (sample size = 2000 observations) and Canada (sample size = 2000 observations) were obtained for use in this paper (see Data Addendum A). The total random samples for each nation were further reduced for missing data and extreme outliers.

The above data were converted to a consistent currency (US$) using the average annual exchange rate from the International Financial Statistics Yearbook published by the International Monetary Fund for the year of the reported data. The data were further adjusted to minimise any inconsistencies among reporting categories (for example, education, occupations) and new analysis variables were created from the LIS data for use in this research. These new analysis variables are presented and defined in Data Addendum B: Analysis Variables.

Presentation Of Results

Overall

The estimated models for both the United States in 1986 and Canada in 1987 are presented in Results Table A3. The overall quality of these models is suggested by two measures: a chi-square statistic and a pseudo-R^2. The chi-square statistic, which tests whether there is an overall relationship between the dependent variable and the set of independent variables, shows that there is such a significant relationship in each model. The pseudo-R^2 reveals a value of 0.178381 for the United States model in 1986 and a value of 0.203988 for the Canada model in 1987.

These two estimated models are similar to many earnings functions found in the literature. Like most earnings functions, these models confirm that age, education, and experience influence earnings. We cannot extract the magnitude of those influences because of the "quadratic within a quadratic" nature of these variables. See equations (8)-(10). Some of the earnings adjustment variables play their expected roles. For instance, being female significantly lowers earnings per hour in both countries and earnings differences across occupations do exist in Canada. Greater wealth (as proxied by home ownership: the variable OWN) increases earnings as expected in the United States. The incidental parameter lambda is significantly greater than zero in both countries, foreshadowing the underpayment results which are discussed below.

TABLE 1 National Comparisons: U.S. in 1983 & 1986 and Canada in 1987

TABLE 1 National Comparisons: U.S. in 1983 & 1986 and Canada in 1987

Variable	US86	CN87	US83 (H&M)
RATIO	83.69%	83.14%	90.43%
Underpayment	16.32%	15.86%	9.57%
Wage/hour	$11.69	US $10.39	$9.09
UNDWAGE	$1.5759	US $1.6811	$0.96
Worker Age	40.15	38.78	38.26
Worker Education (years)	13.16	11.04	14.23

The results for the United States (1986) and Canada (1987) are presented and also are compared to the H&M (1992) results based on 1983 data for the United States. In this discussion, the term underpayment refers to (1 - RATIO) where RATIO is the actual wage divided by the potential wage. That is, the underpayment results show the percentage shortfall of actual wages below potential wages. Results Tables A7-A11 show findings for RATIO only. Underpayment can be calculated as described above.

Table 1 presents selected information from Results Table A7 in the Summary of Results section. The results from the H&M (1992) paper for the U.S. in 1983 are also presented for comparison.

The results show very similar labour market characteristics for the United States in 1986 and Canada in 1987, with an increase in underpayment in the United States from 9.57% in 1983 to 16.31% in 1986 (statistically significant at = .001). The highest level of underpayment is in Canada - although the U.S. in 1986 is very close. The highest actual wages are in the United States in 1986.

Hypothesis Tests

The specific results of the hypotheses' tests are presented in Tables A7-A11 in the Summary of Results section. In the following sections the results of the significance testing for each hypothesis for both nations and for H&M (1992) for the U.S. in 1983 are summarised.

Hypothesis 1: Underpayment Differs by Demographic Category

Gender, marital status, and age are all related to significant differences in underpayment in the United States (1986) and Canada. In both the United States (1986) and Canada lower underpayment was experienced by workers who were males, married, and in the category prime-age married males. In both the United States (1986) and Canada the lowest underpayment was in the prime-age married males category. These results support the findings of H&M (1992) for the United States in 1983.

The specific results for each nation sampled are:

<u>United States (1986):</u>

- Male underpayment (15.08%) is less than female underpayment (21.25%).
- Married workers have less underpayment than single workers: 14.48% vs. 19.24%.
- Underpayment varies across age categories: prime age workers have the lowest underpayment: 14.50%; young workers have the highest underpayment: 24.17%; old workers have underpayment between those others: 18.88%. Middle-aged workers (45-55) have underpayment of 16.01% that is also different from young

and old workers but is not different from prime age workers.

Both the Waller-Duncan and Tukey tests reveal that prime-age and middle-age workers differ statistically from both the young and old workers' groups in their underpayment.

- Prime-age married males have less underpayment than young males: 13.15% vs. 22.11%.
- Prime-age married males (13.15%) are less underpaid than females (21.25%).

<u>Canada (1987):</u>

- Males (15.94%) are less underpaid than female workers (21.08%).
- Single workers suffer greater underpayment than married workers: 19.75% vs. 15.41%.
- Underpayment differs over age categories. Middle-aged workers have the lowest underpayment: 15.48%.; young workers have the highest underpayment: 24.57%; and old (19.52%) and prime age (15.81%) workers have underpayment between those others.

The formal tests show that the strongest differences from other groups occur in the young workers. They are significantly different from all three other groups. Furthermore, the prime-age workers suffer similarly to middle-aged workers.

- Prime-age married males suffer less underpayment than young males: 15.18% vs. 23.33%.
- Females (21.09%) experience greater underpayment than prime-age married males (15.18%).

<u>United States (1983):</u>

- Prime age married males (7.93%) endure less underpayment than young workers (12.45%).
- Prime age married males (7.93%) experience less underpayment than female workers (11.08%).

Hypothesis 2: Workers With More Education Will Have Less Underpayment

For the United States (1986) and Canada, educational differences were associated with underpayment differences. This result supports H&M (1992) for the United States in 1983. The results for each country for the sample years are:

United States (1986):

- Underpayment decreases progressively with higher education: 0-8 years: 22.24%; 9-11 years: 19.21%; 12 years: 17.25%; 13-15 years: 15.85%; 15+ years: 13.35%.

The Waller-Duncan and Tukey tests reinforce each other in revealing patterns in the levels of underpayment by education. The highest educational level has lower underpayment than any other group. The middle groups (9-11,12, 13-15) endure similar underpayment.

Canada (1987):

- As in the U.S., underpayment, in general, decreases progressively with higher education: 0-8 years: 19.31%; 9-11 years: 17.15%; 12 years: 13.78%; 13-15 years: 16.94%; 15+ years: 12.95%.

The formal tests show that the three middle educational groups suffer roughly equal underpayment. Only the extreme groups (0-8 and 15+) exhibit significantly different levels. The most-educated workers endure the lowest underpayment whereas the least-educated workers suffer the most.

United States (1983):

- Workers with college education (9.29%) do better than high school graduates (10.85%).
- High school graduates (10.85%) suffer less than those who never enter high school (12.78%).

Hypothesis 3: Workers in Urban Areas Are Less Underpaid

For the United States (1986) the location (urban vs. rural) of the worker, when measured by population (MSA vs. non-MSA) or land use (farm vs. nonfarm), is not significantly connected to underpayment. For Canada the location of the worker is not significant when based on population but is significant when measured by land use (nonfarm workers have less underpayment). These results are weaker than the findings of H&M (1992) for the United States in 1983. The specific results for each country sampled are:

United States (1986):

- Significant differences in underpayment by location are not found. This is true for both population (MSA vs. non-MSA) and land use (farm vs. nonfarm) distinctions.

Canada (1987):

- Urban-rural differences in underpayment, when measured by population, are not significant.
- Urban-rural differences in underpayment, when measured by land use, are significant. Non-farm workers experience less underpayment: 15.92% vs. 20.33%.

United States (1983):

- Workers in urban areas (whether measured by population or land use) endure less underpayment than rural workers: population 9.10% vs. 10.58%; land use 9.49% vs. 12.84%.

Hypothesis 4: Workers with Greater Wealth Will Have Less Underpayment

Wealth is proxied by home ownership under the assumption that homeowners have greater wealth than nonhomeowners. In both countries for the sample periods, home ownership was significantly correlated with lower underpayment. This result supports the findings of H&M (1992). The specific sample results for each nation are:

United States (1986):

- Homeowners have less underpayment than nonhomeowners: 14.67% vs. 19.00%.

Canada (1987):

- Homeowners have less underpayment: 15.62% vs. 19.18%.

United States (1983):

- Homeowners have less underpayment: 9.05% vs. 10.84%.

Conclusions

The general conclusion from the data is that the United States in 1986 and Canada in 1987 have similar underpayment performance and determinants. Furthermore, the findings are consistent and logical and confirm the research performed by Hofler and Murphy (1992) on 1983 data for the United States.

Viewing national underpayment as a measure of a country's lost GDP, the 16.315% underpayment for the United States for 1986 represents a potential gain of $826.9 Billion, if effective policies could be carried out to eliminate the underpayment.[6] In Canada, the 16.857% underpayment represents a gross potential gain of $81.5 Billion.[7]

The increase in underemployment between the Hofler & Murphy paper (9.57%) using 1983 data and this paper (16.31%) using 1986 data is statistically significant.[8] This suggests that efficiency in processing job market information in the United States dropped during this period and resulted in lower GDP growth than might have been achieved. No attempt has been made in this paper to identify the reasons for the estimated decline in United States national performance or to identify policies that might lead to absolute domestic improvement or relative international improvement.

Possibilities for further research are many. First, the reasons for the decline in the underpayment performance of the United States could be investigated. Second, as more recent information becomes available from LIS, other nations could be added and trends in performance among countries could be studied. Finally, the stochastic frontier technique could be used in new domestic and international applications.

References

Aigner, D., C.A. Knox Lovell and P. Schmidt. 1977. "Formulation and Estimation of Stochastic Frontier Production Function Models". *Journal of Econometrics,* 6: 21-38.

Battese, G. 1992. "Frontier Production Functions and Technical Efficiency: A Survey of Empirical Applications in Agricultural Economics". *Agricultural Economics*, forthcoming.

Hofler, R. and K. Murphy .1989. "Using a Composed Error Model to Estimate the Frictional and Excess Supply Components of Unemployment". *Journal of Regional Science,* 29: 213-228.

________. 1992. "Underpaid and Overworked: Measuring the Effect of Imperfect Information on Wages". *Economic Inquiry*, 30: 511-529.

Jondrow, J., C.A. Knox Lovell, I.S. Materov and P. Schmidt. 1982. "On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model". *Journal of Econometrics*, 19: 233-238.

Lippman, S. and J. McCall. 1976. "The Economics of Job Search: A Survey". *Economic Inquiry*, 14: 155-189.

McCall, J. 1970. "Economics of Information and Job Search". *Quarterly Journal of Economics*, 84: 113-126.

Mincer, J. 1974. *Schooling, Experience, and Earnings*. New York: National Bureau of Economic Research.

Mortensen, D.T. 1970. "Job Search, the Duration of Unemployment, and the Phillips Curve". *American Economic Review*, 60: 847-862.

Murphy, K.M. and F. Welch. 1990. "Empirical Age-Earnings Profiles". *Journal of Labor Economics,* 8: 202-229.

Appendix A

Appendix A

Hofler and Murphy (1991) Proof:

$$\left(E\left[W \mid W \geq W(R)\right] \cdot W(R)\right)$$

Prove: The higher the individual reservation wage ($w(r)$) selected the higher the conditional expected job seeker's wage ($E[w \mid w \geq w(r)]$) and the lower the level of underemployment (μ). Since the potential wage ($w(qo)$) is fixed based on the value of the job seekers accumulated human capital, the higher the individual's expected wage, the lower the level of underemployment [$\mu = w(qo) - w(a)$]. The challenge is to prove that:

Prove:

$$E\left[W \mid W(R)\right] \geq W(R)$$

and

$$\frac{\partial E\left[W \mid W \geq W(R)\right]}{\partial W(R)} \cdot 0$$

Given:

$$E\left[W \mid W \geq W(R)\right] = \frac{\int_{W(R)}^{W(QO)} WF(W)DW}{\int_{W(R)}^{W(QO)} F(W)DW}$$

Then:

$$\frac{\partial E\left[W \mid W \geq W(R)\right]}{\partial W(R)} = \frac{\partial}{\partial W(R)} \left(\frac{\int_{W(R)}^{W(QO)} WF(W)DW}{\int_{W(R)}^{W(QO)} F(W)DW} \right)$$

$$\frac{W \geq W(R)]}{W(R)} = \frac{F(W(R))\left[\int_{W(R)}^{W(QO)} WF(W)DW\right] - W(R)\,F(W(R))\left[\int_{W(R)}^{W(QO)} F(W)\right.}{\left[\int_{W(R)}^{W(QO)} F(W)DW\right]^2}$$

$$\frac{[W \mid W \geq W(R)]}{\partial W(R)} = \frac{F(W(R))\left[\int_{W(R)}^{W(QO)} WF(W)DW - W(R) \int_{W(R)}^{W(QO)} F(W)DW\right.}{\left[\int_{W(R)}^{W(QO)} F(W)DW\right]^2}$$

Because:

$$E\left[W \mid W \geq W(R)\right] = \frac{\int_{W(R)}^{W(QO)} WF(W)DW}{\int_{W(R)}^{W(QO)} F(W)DW}$$

Therefore: (Proof)

$$E\left[W \mid W \geq W(R)\right] \cdot W(R)$$

TABLE A1 variables obtained from the Luxembourg Income Study (LIS) for the United States ('86) and Canada ('87)

TABLE A1 variables obtained from the Luxembourg Income Study (LIS) for the United States ('86) and Canada ('87)

LIS Variable Name	Definition
D1	Age: Head of Family
D3	Sex: Head of Family
D4	Number of Persons in Family
D5	Family Structure (Single/Multiple/Economic/Tax/etc.)
D7	Geographic Location Indicator A (Farm/NonFarm)
D10	Education Level: Head of Family
D12	Occupational Training: Head of Family
D14	Occupational Classification: Head of Family
D16	Industry Classification: Head of Family
D18	Type of Worker Group: Head of Family (Agric/NonAgric/Govt/etc.)
D20	Geographic Location Indicator B (Urban/Rural)
D21	Marital Status: Head of Family
D22	Tenure (Owned or Rented Housing)
D27	Children under age 18
COUNTRY	Country
HWEIGHT	Family Unit Sample Weight
GI	Total Gross Annual Income: Family
SOCI	Social Insurance Income: Family
LFSHD	Labor Force Status: Head of Family
HRSHD	Hours worked per week: Head of Family
SOCRHD	Social Retirement Income: Head of Family
UNEMPHD	Unemployment Income: Head of Family
PRVPENHD	Private Pension Income: Head of Family
PUBPENHD	Public Pension Income: Head of Family
WEEKHDFT	Full-time weeks worked per year: Head of Family
WEEKHDPT	Part-time weeks worked per year: Head of Family
WEEKHDUP	Unemployed weeks per year: Head of Family
V39	Gross Annual Wage/Salary: Head of Family

TABLE A2 Analysis Variables

TABLE A2 Analysis Variables

Input Variables:	Variable Name	Variable Definition
Y	LNWAGEHR	Ln of Wage per hour (Dependent Variable)

X1	AGEQT	Murphy-Welch Age Variable (Gamma = 0.0166667 = 1/60)
X2	AGEQTSQ	Murphy-Welch Age Variable Squared
X3	EDQT	Murphy-Welch Education Var. (Gamma = 0.0166667 = 1/60)
X4	EDQTSQ	Murphy-Welch Education Variable Squared
X5	EXPQT	Murphy-Welch Experience Var. (Gamma = 0.0166667 = 1/60)
X6	EXPQTSQ	Murphy-Welch Experience Variable Squared
X7	NRKIDS	Number of Children under age 18 in Family
X8	FAMINC	Other Family Income (Not Head of Family)
X9	MSA	Dummy Variable (Population > or < 100,000)
X10	LANDUSE	Dummy Variable (Farm/NonFarm)
X11	MARSTAT	Dummy Variable (Married/NonMarried)
X12	SEX	Dummy Variable (Male/Female)
X13	OCCMP	Dummy Variable (Occupation: Mgmt/Professional)
X14	OCCSTCH	Dummy Variable (Occupation: Sales/Tech/Serv/Admin)
X15	OCCCRFT	Dummy Variable (Occupation: Crafts/Skills)
X16	OCCOPER	Dummy Variable (Occupation: Operator/ Assembler/Test)
X17	OCCUNSK	Dummy Variable (Occupation Unskilled)
X18	INDMIN	Dummy Variable (Industry: Mining)
X19	INDCON	Dummy Variable (Industry: Construction)
X20	INDDUR	Dummy Variable (Industry: Durable Goods)
X21	INDNON	Dummy Variable (Industry: NonDurable Goods)
X22	INDTPU	Dummy Variable (Industry: Transport/Comm/Utilities)
X23	INDTRD	Dummy Variable (Industry: Wholesale & Retail Trade)
X24	INDFIRE	Dummy Variable (Industry: Fin/Ins/RE)
X25	INDSERV	Dummy Variable (Industry: Service)
X26	OWN	Dummy Variable (Homeowner)
Output Variables:		
	RATIO	Ratio of Actual Wage to Potential Wage
	WAGEHR	Wage per Hour (Currency)
	UNDWAGE	Measure of Underemployment (Currency)
	UTMNF	Natural Log of Measure of Underemployment

TABLE A3 Summary of Results

TABLE A3 Summary of Results

Variable	US86 Coefficient	t-statistic	CN87 Coefficient	t-statistic
Constant	0.45735	1.123	1.3272	7.856
AGEQT	0.94311E-01	2.703	0.98533E-01	3.345
AGEQTSQ	-0.24446E-03	-3.695	-0.17961E-03	-3.185
EDQT	-0.26150E-01	-0.385	-0.16838	-3.256
EDQTSQ	-.067790E-02	-0.639	0.28351E-02	3.804
EXPQT	-0.96906E-01	-2.114	-0.11801	-3.120
EXPQTSQ	0.32383E-03	2.285	0.33210E-03	2.790
NRKIDS	-0.11298E01	-0.494	-0.32059E-01	-1.501
FAMINC	0.11540E-05	1.881	-0.22811E-05	-1.394
MSA	0.21375	4.116	-0.18385E-01	-0.400
LANDUSE	-0.20424	-1.163	-0.10751	-2.096
MARSTAT	0.56633E-01	1.008	0.14031	2.080
SEX	-0.18623	-3.140	-0.13723	-2.174
OCCMP	0.22304	1.745	0.88046	9.679
OCCSTCH	-0.11487E-01	-0.091	0.74498	8.780
OCCCRFT	0.79607E-01	0.639	0.71700	7.352
OCCOPER	-0.49412E-01	-0.349	0.77336	8.197
OCCUNSK	-0.93720E-01	-0.698	N/A	N/A
INDMIN	0.38015	1.026	0.31162	3.187
INDCON	0.20195	1.836	0.37173E-05	0.000
INDDUR	0.14466	1.331	0.42743E-01	0.449
INDNON	0.13427	1.137	0.51555E-01	0.490
INDDTPU	0.11073	0.937	0.13972	1.570
INDTRD	-0.23752E-01	-0.229	-0.77021E-01	-1.080
INDFIRE	-0.35816E-01	-0.304	-0.49676E-01	-0.445
INDSERV	-0.18414	-1.898	-0.10803	-1.501
OWN	0.16347	3.232	0.79221E-01	1.584
LAMBDA ()	1.8334	14.326	2.6491	14.034
SIGMA ()	0.87193	53.859	0.93278	55.661

Note: 1. Frontier Regression Results: Dependent Variable is Ln Wage/Hr

2. United States in 1986 (US86) & Canada in 1987 (CN87)

3. Observations: US86 = 1122; CN87 = 1087

TABLE A4 Results Continued

TABLE A4 Results Continued

Measure	US86	CN87
Log-Likelihood (Unrestricted)	-0.1039649E+04[a]	-0.9819075E+03[a]
Log-Likelihood (Restricted)	-0.1265367E+04[a]	-0.1233533E+04[a]
Pseudo-R^2	0.178381[b]	0.203988[b]
Chi-Square Statistic	451.436[c]	503.251[c]
Number of Observations	1122	1087

- The Log-Likelihood (unrestricted), denoted by $\ln L_{UR}$, is the value of the logarithm of the likelihood function at the optimum (maximum). The Log-Likelihood (restricted), denoted by $\ln L_R$, is the value of the logarithm of the likelihood function when all slope coefficients equal zero.
- This is calculated as $1-(\ln L_{UR}/\ln L_R)$. It is interpreted roughly the same as the usual coefficient of determination.
- This is calculated as $2(\ln L_{UR} - \ln L_R)$. It tests the null hypothesis that all slope coefficients equal zero. It also implicitly tests a second null that the Pseudo-R^2 equals zero. Both null hypotheses are rejected in both the US86 and CN87 models.

TABLE A5 Results, United States in 1986

TABLE A5 Results, United States in 1986

Variable	Mean	Std. Dev.	Min	Max
LOG-LIKELIHOOD	-1039.6490			
RATIO	.83685	.14133	.4145E-03	.9956
WAGEHR	11.687	11.296	.3077E-01	240.4
UNDWAGE	1.5759	2.5821	1.061	74.20
LNWAGEHR	2.2049	.74773	-3.481	5.482
UTMNF	.34793	.29476	.5953E-01	4.307

TABLE A6 Results, Canada in 1987

TABLE A6 Results, Canada in 1987

Variable	Mean	Std. Dev.	Min	Max
LOG-LIKELIHOOD	-981.9075			
RATIO	.83143	.16363	.2142E-03	.9838
WAGEHR	10.386	6.3444	.2357E-01	63.98
UNDWAGE	1.6811	4.0790	1.036	110.0
LNWAGEHR	2.1324	.75301	-3.748	4.159

UTMNF	.32274	.38751	.3525E-01	4.700

TABLE A7 Results, Overall

TABLE A7 Results, Overall

Item	US '86	CANADA '87 (US$)
Number of Observations	1122	1087
Analysis Variable		
RATIO	83.685%	83.143%
WAGEHR	$11.687	US$10.386
UNDWAGE	$1.5759	US$1.6811
LNWAGEHR	2.2049	2.1324
UTMNF(ln undwage)	0.34793	0.32274
Indiv.Gross Wage	$23,419	US$20,806
Other Family Income	$11,244	US$10,535
Total Income	$34,663	US$31,341
Exchange Rate		1.326
Age	40.153	38.777
Average Education	13.157	11.044
H.S. (Attended)	11.77%	29.81%
H.S. (Graduate)	36.28%	1.66%
College (Attended)	20.41%	25.58%
College (Graduate)	27.72%	17.39%
Other Educ.	3.83%	25.58%
All Education %	100.00%	100.00%

TABLE A8 Results, Hypothesis 1: Demographic Group

TABLE A8 Results, Hypothesis 1: Demographic Group

Country Variable = RATIO

	Male	Female	T-Statistic ()	Prob>T ()
US86	.8492	.7875	5.5960	.0001
CN87	.8406	.7891	3.5503	.0005

	Single	Married	T-Statistic ()	Prob>T ()
US86	.8076	.8552	-5.4888	.0001
CN87	.8025	.8459	-3.9163	.0001

	Age < 25	Age 25-44	Age 45-55	Age >55	F-Stat. ()	Prob>F ()
US86	.7583	.8550	.8400	.8112	16.4261	.0001
CN87	.7554	.8424	.8455	.8041	9.46	.0001

	Prime Age Married Males (Age 25-44)	Young Males (Age < 25)	T-Statistic ()	Prob>T ()
US86	.8685	.7789	-5.0232	.0001
CN87	.8482	.7667	-2.8160	.0068

	Prime Age Married Males (Age 25-44)	Females (All Ages)	T-Statistic ()	Prob>T ()
US86	.8685	.7875	-7.0235	.0001
CN87	.8482	.7891	-3.8679	.0001

TABLE A9 Results, Hypothesis 2: Level of Education

TABLE A9 Results, Hypothesis 2: Level of Education

Country Variable = RATIO

	0-8 Years	9-11 Years	12 Years	13-15 Years	15 + Years	F-Statistic ()	Prob>F ()
US86	.7747	.8080	.8275	.8415	.8664	7.5462	.0001
CN87	.8069	.8285	.8622	.8307	.8705	4.4988	.0013

TABLE A10 Results, Hypothesis 3: Urban vs. Rural Location

TABLE A10 Results, Hypothesis 3: Urban vs. Rural Location

Country	Variable = RATIO			
	Non-MSA	MSA	T-Statistic ()	Prob>T ()
US86	.8238	.8416	-1.8645	.0628
CN87	.8246	.8407	-1.6457	.1001

	Non-Farm	Farm	T-Statistic ()	Prob>T ()
US86	.8376	.7806	1.2649	.2249
CN87	.8409	.7967	3.0476	.0025

TABLE A11 Results, Hypothesis 4: Wealth

TABLE A11 Results, Hypothesis 4: Wealth

Country Variable = RATIO

	Non-Homeowner	Homeowner	T-Statistic ()	Prob>T ()
US86	.8100	.8533	-5.1093	.0001
CN87	.8082	.8438	-3.4481	.0006