Decomposing Changes in Wage Distributions: A Unified Approach

Thomas Lemieux
Department of Economics
University of British Columbia
997-1873 East Mall
Vancouver, British Columbia, V6T 1Z1

tel: 604-822-6763
email: tlemieux@interchange.ubc.ca

Abstract

Over the last fifteen years, a large number of studies have attempted to explain the determinants and changes of wage inequality. This paper proposes a simple procedure to decompose changes in the distribution of wages or in other distributions into three factors: changes in regression coefficients, changes in the distribution of covariates, and residuals changes. The procedure requires only estimating standard OLS regressions augmented by a logit or probit model. The procedure can be extended by modelling residuals as a function of unmeasured skills and skill prices. Two empirical examples showing how the procedure works in practice are considered. The first example looks at sources of differences in the wage distribution in Alberta and British Columbia. The second example re-examines the sources of changes in overall wage inequality in the United States from 1973 to 1999. Finally, the proposed procedure is compared to other existing procedures.

Résumé

Au cours des quinzes dernières années, nombre d'études se sont penchées sur les déterminants et les changements de la distribution des salaires. Ce mémoire propose une procédure pour décomposer les changements de la distribution des salaires en trois éléments: les changements dans les coefficients de régression, la distribution des regresseurs et les changements résiduels. Cette procédure ne nécessite que l'estimation de regressions par moindre carrés ordinaires et d'un modèle probit ou logit. L'auteur montre aussi comment modéliser les résidus en fonction de compétences non mesurées. La procédure proposée est mise en application dans le contexte de deux exemples: la distribution des salaires en Alberta et en Colombie-Britannique et les changements dans la distribution des salaires de 1973 à 1999 aux Etats-Unis. Le mémoire examine aussi comment cette procédure se compare aux méthodes proposées par d'autres chercheurs.

JEL classification code: J3

1. Introduction

Studying changes in the distribution of wages has been an active area of research over the last fifteen years. One reason for the resurgence of interest for this topic is that wage inequality increased steeply in several countries, and in particular in the United States, since the early 1980s. Another reason is that it is now much easier to conduct in-depth empirical studies thanks to the increased power of computers and the easy access to large micro data sets for a large number of countries and time periods.

Most studies have focused on one particular aspect of wage inequality, namely the wage gap between more and less-educated workers. For example, Katz and Murphy (1992) and Bound and Johnson (1992) show that after declining during the 1970s, the wage gap between college- and high school-educated workers increased steeply in the 1980s in the United States. Both of these studies propose a simple supply and demand explanation to this phenomena. They suggest that while the relative demand for more educated workers increased steadily during the 1970s and 1980s, the growth in the relative supply did slowdown in the 1980s relative to the 1970s. As a result, the growth in relative demand due to factors such as skilled-biased technological change outstripped the growth in relative supply during the 1980s, causing the relative wage of college-educated workers to increase.

Ten years later, this supply-demand-technology paradigm remains to a large extent the accepted theoretical framework for understanding the evolution of the college-high school wage gap. While recent research has considered richer versions of this basic approach, supply, demand and technology remain the key factors used to understand this dimension of wage inequality.[1] There is also a consensus in the literature on the basic facts to be explained, namely that the U.S. college-high school wage gap declined in the 1970s, increased rapidly in the 1980s, and increased more slowly in the 1990s.[2] Part of this consensus stems from the fact that the college-high school wage gap, or returns to education more generally, is simply a difference in conditional means that can be readily estimated using regression methods.

By contrast, there is much less of a consensus in the literature when it comes to understanding the evolution of overall measures of wage inequality over the last 20 or 30 years. One possible reason for the lack of consensus is that, unlike the case of mean wage differences by education groups, there are no unique and well-accepted measures of overall wage inequality. As is well known in the income distribution literature, different summary measures of inequality like the coefficient of variation or the Gini coefficient can yield different rankings of inequality as they put different weights on different parts of the distribution.[3] As a result, recent research has increasingly focused on more global, though less parsimonious, methods for describing changes in the whole distribution of wages such as kernel density estimation.

There is no consensus either on the underlying economic explanations for the changes in overall wage inequality. For example, Juhn, Murphy and Pierce (1993) conclude that much of the increase in wage inequality for U.S. men in the 1980s is due to increased returns to skills. By contrast, DiNardo, Fortin, and Lemieux (1996) find that the decline in the real value of the minimum wage played an important role in the increase in wage inequality for both men and women during the same period. Lee (1999) and Teulings (2002) push this idea further and conclude that *most* of the increase in inequality during this period is due to changes in the minimum wage. These studies all propose innovative procedures for estimating changes in overall wage inequality, but it is difficult to compare the results of these various approaches because of differences in methodologies.

The main goal of this paper is to propose a unified and simple approach to analyzing changes in the distribution of wages that is economically interpretable using the standard tools of human capital theory. One related goal is to compare the proposed approach to other techniques that have been used in the literature. One last goal is to show how the proposed approach can also be used to analyze changes or differences in distribution in other contexts. The approach can be viewed as a generalization of well-known Oaxaca-Blinder decomposition of means to the full distributional case.

I illustrate how the method works using two empirical examples. I first analyze the sources of difference in the wage distribution of women in Alberta and British Columbia in year 2000. I then use the proposed approach to re-assess the sources of changes in the distribution of wages in the United States between 1973 and 1999. Like most other studies, I find that increases in the returns to measured skills like experience and education play a major role in secular increases in wage inequality in the United States. I also find, however, that this explanation does not account well for the changes at the bottom end of the wage distribution. More importantly, I find that much of the increase in residual wage inequality is due to changes in the composition of the workforce. This suggests that increases in the price of unmeasured skills does not play much of a role in the overall growth in wage inequality since 1973.

2. Economic model: human capital

Before turning to econometric issues, it is useful to first provide some background on models of wage determination across individuals. Almost all empirical studies of wage determination use as their point of departure Mincer (1974)'s famous human capital earnings function

$$\ln w = c + rS + b_1E + b_2E^2 + e,$$

where w is earnings (or the hourly wage when available), c is a constant, S is years of schooling, E is years of labor market experience, and e is an error term. Hundreds, if not thousands of studies have estimated this equation for a large number of countries and time periods.[4] This earnings function is probably the closest thing to an "empirical law" in labor economics, since almost all studies show that schooling has a positive and significant effect on earnings ($r>0$) and that earnings are a concave function of labour market experience ($b_1 > 0$ and $b_2 < 0$). Mincer shows how this earnings equation can be obtained as the outcome of a process by which individuals optimally invest in two types of human capital, education and on-the-job training (OJT).[5] Years of schooling is a fairly direct measure of education human capital, while years of labour market experience is viewed as a proxy for on-the-job training.

In this context, S and E can be though as the "quantity" of human capital, while r, b_1, and b_2 are the "prices" or returns to human capital. In general, the distribution of wages depends both on the distribution of human capital and its price. For instance, ignoring all wage determinants but schooling means that $\ln w = c + rS$ and that the variance of log wages is the product of the squared price (return) of schooling times the variance of schooling ($\mathrm{Var}(\ln W) = r^2 \times \mathrm{Var}(S)$).

Mincer also discusses the implications of the human capital approach for the conditional distribution of the error term e. In particular, he considers the case in which different individuals with the same level of formal schooling S invest with differential OJT intensity. As in other aspects of human capital investments, individuals who devote more time and efforts in OJT implicitly pay for it by accepting lower earnings initially. But these investments eventually pay off in terms of higher earnings in the future. This also means that the slope of the experience-earnings profile is steeper for those who invest more.

This is illustrated for several types of individuals in Figure 1. Individuals of type H invest more than average (M), while individuals of type L invest less than average. Since a wage regression only captures conditional means, the parameter estimates of b_1 and b_2 capture the average experience earnings profile M. However, one additional empirical prediction of the model in Figure 1 is that residual wage dispersion decreases as a function of experience until the "overtaking" point (10 years of experience in this example) where the earnings of type H catch up with those of others. Residual dispersion starts expanding past this point.

In terms of the conditional variance of the error term e, this model implies that the variance should first decline and then expand as a function of experience. In econometric terms, this means that e is heteroskedastic. In economic terms, it suggests that it is legitimate to interpret differences in the residual variance across experience levels as evidence that the dispersion in unmeasured human capital systematically varies across those experience levels. For instance, Juhn, Murphy and Pierce (1993) argue that the increase in the residual variance of wages in the United States during the 1970s and 1980s should be interpreted as evidence that the "price" of unmeasured human capital increased during this period. The early work by Mincer and others provide the conceptual and empirical basis for this type of statement.

Related arguments can be used to show that residual wage dispersion may also depend on the level of schooling of individuals. The simplest argument is that years of schooling is an imperfect proxy for true educational inputs and that the error term includes unmeasured aspects of educational inputs such as school quality. Systematic differences in the residual variance across education groups arise if the residual dispersion in school quality is different for different levels of schooling. A more sophisticated argument is that if individuals who invest more in education have higher marginal returns to education than others, the log wage-schooling relationship will generally be convex, i.e. the labour market "price" of schooling will be higher at higher levels of education.[6] This results in more residual wage dispersion at higher than lower levels of education even if the dispersion in unmeasured educational inputs is the same at different levels of education.

3. Decomposing the wage distribution

This section proposes a decomposition procedure that unifies several existing approaches by combining elements of Juhn, Murphy, and Pierce (1993) and DiNardo, Fortin, and Lemieux (1996). Consider a general regression model

$$y_{it} = x_{it}\beta_t + e_{it} , \qquad (1)$$

where i indicates the observation (individual) and t the time period; x_{it} is a 1×k vector of covariates (including a constant), β_t is a k×1 vector of parameters, and the error term e_{it} is assumed to have a zero conditional mean ($E(e_{it} \mid x_{it}) = 0$). In terms of the earnings model of Section 2, y_{it} corresponds to log wages while x_{it} is a vector of human capital and other variables such as schooling, experience, and other socio-economic characteristics.

For simplicity, I focus on the case where there is a series of cross-sectional samples for different time periods t. However, this setting can readily be applied to cases where t represent various groups (men vs. women, immigrants vs. non-immigrants, etc.) or regions instead of time periods.

While it is customary to assume random sampling, the sampling structure of most micro data sets collected by Statistics Canada and other survey agencies is considerably more complex. I ignore all these issues and assume random sampling except for the fact that the probability of sampling of a given observation i at time t, $1/\omega_{it}$, depends on a set of exogenous factors such as geographic location, urban/rural status, etc. This probability is the inverse of the sample weight, ω_{it}. To simplify the notation, I also normalize the sample weights so that they sum up to one ($\sum_i \omega_{it}=1$).

Consider the OLS estimate b_t of β_t. The estimated regression equation is:

$$y_{it} = x_{it}b_t + u_{it}, \qquad (2)$$

where u_{it} is the regression residual that has, by construction, a zero average and is uncorrelated with the covariates. The sample average of y in period t is

$$\bar{y}_t = \bar{x}_t b_t , \qquad (3)$$

where $\bar{y}_t = \sum_i \omega_{it}y_{it}$ and $\bar{x}_t = \sum_i \omega_{it}x_{it}$.

3.1. Decomposing changes in the mean: *the Oaxaca-Blinder decomposition*.

Consider an alternative time period s. The sample average of y in period s is

$$\bar{y}_s = \bar{x}_s\, b_s. \tag{4}$$

The difference between the average value of y in periods t and s can be decomposed as

$$\bar{y}_t - \bar{y}_s = \bar{x}_t\,(b_t - b_s) + (\bar{x}_t - \bar{x}_s) b_s\,, \tag{5}$$

where the first term on the right hand side captures differences in the estimated parameters, while the second term captures differences in the average values of the covariates between the two samples. This kind of decomposition, first suggested by Oaxaca (1973) and Blinder (1973), is now a standard practice in empirical economics. One useful interpretation of this decomposition is that $\bar{x}_t b_s$ is a counterfactual average value of y that would be obtained if the parameters in period t were replaced by those in period s. In terms of the wage model, this represents the average wage that would prevail in period t if the “price” of human capital were the same as in period s. Denote this counterfactual as $\bar{y}_t^a$ where

$$\bar{y}_t^a = \bar{x}_t b_s. \tag{6}$$

This counterfactual can be used to rewrite equation (5) as:

$$\bar{y}_t - \bar{y}_s = (\bar{x}_t b_t - \bar{y}_t^a) + (\bar{y}_t^a - \bar{x}_s b_s) = (\bar{y}_t - \bar{y}_t^a) + (\bar{y}_t^a - \bar{y}_s).$$

Consider y_{it}^a, the individual-specific counterfactual wage:

$$y_{it}^a = x_{it} b_s + u_{it} = y_{it} + x_{it}(b_s - b_t).$$

In terms of computation, $\bar{y}_t^a$ can be either obtained by computing the sample average of x_{it} and applying equation (6), or by computing directly the sample average of y_{it}^a:

$$\bar{y}_t^a = \sum_i \omega_{it} y_{it}^a. \tag{7}$$

3.2. Decomposing changes in the variance

Before trying to go from the simple mean decomposition to a full decomposition of the whole distribution, consider the case of the variance which is one among many summary measures of dispersion. The variance is the most natural measure of dispersion to use in a regression context since, by construction, the residuals u_{it} are uncorrelated with the covariates x_{it}. The variance of y_{it} is, therefore, the sum of the variance of the predicted ($x_{it}b_t$) and residual (u_{it}) parts of the regression:

$$V_t = b_t'\Omega_{x,t} b_t + \sigma_t^2, \qquad (8)$$

where V_t is the variance of y_{it}, $\Omega_{x,t}$ is the variance-covariance matrix of x_{it}, and σ_t^2 is the variance of the residuals. This suggests three potential sources of change in the variance of y: changes in the estimated parameters, b_t, changes in the variance-covariance matrix of the covariates, $\Omega_{x,t}$, and changes in the residual variance σ_t^2.

It is useful to link these three sources of changes back to the economic model of wage determination of Section 2. In this setting, changes in b_t represent the role of "price effects" in changes in the variance of wages. For example, if b_t doubles between two periods, the predicted part of the variance, $b_t'\Omega_{x,t} b_t$, will quadruple, holding the distribution of human capital constant.

At first glance, equation (8) suggests that the impact of changes in the distribution of human capital (x) on the variance of wages can be captured by the variance-covariance matrix $\Omega_{x,t}$. The human capital approach suggests, however, that this is only one possible impact of changes in the distribution of human capital. For example, if the workforce becomes more experienced over time, this may change the wage distribution for two separate reasons. The first reason is that if the variance of experience also happens to change over time, then the predicted variance $b_t'\Omega_{x,t} b_t$ should change too. The second reason is that if the residual variance is higher for more experienced workers (as in Mincer's OJT model), the overall residual variance σ_t^2 should expand because of composition effects.

To see this point more clearly, consider the case where $x_{it} = [x_{i1t}, ..., x_{ijt}, ..., x_{iJt}]$ is an exhaustive set of dummy variables that divide the sample in a set of J cells. For example, if the two basic covariates are age and education, J indicates the number of age-education cells (or skill groups) in which the sample can be divided. In the first empirical example below, I use public use files of the Labour Force Survey (LFS) in which education is coded in 7 categories and age is coded in 5-year age bands. J is thus equal to 70 for workers age 15 to 65 (10 age categories).

The sample average of the dummy variable x_{ijt} is the proportion of the sample in cell j, θ_{jt}:

$$\bar{x}_{jt} = \sum_i \omega_{it}\, x_{ijt} = \sum_{x_{ijt}=1} \omega_{it} = \theta_{jt}.$$

Furthermore, the OLS estimates b_t (with no intercept) are just the sample means of y_{it} for each cell j:

$$b_t' = [b_{1t},..., b_{jt}, ..., b_{Jt}] = [\bar{y}_{1t} ,..., \bar{y}_{jt} ,..., \bar{y}_{Jt}],$$

where

$$\bar{y}_{jt} = (1/\theta_{jt}) \sum_{x_{ijt}=1} \omega_{it} y_{it}.$$

The sample mean of y is:

$$\bar{y}_t = \sum_i \omega_{it}\, y_{it} = \sum_j \theta_{jt}\, \bar{y}_{jt} = \sum_j \bar{x}_{jt}\, b_j = \bar{x}_t\, b_t\ .$$

Similarly, the variance of y is:

$$\begin{aligned} V_t &= \sum_i \omega_{it}(y_{it}- \bar{y}_t)^2 = \sum_i \omega_{it}(x_{it}\, b_t - \bar{x}_t\, b_t)^2 + \sum_i \omega_{it} u_{it}^2 \\ &= \sum_j \theta_{jt}\, (\bar{y}_{jt} - \bar{y}_t)^2 + \sum_j \theta_{jt}\, \sigma_{jt}^2, \end{aligned} \qquad (9)$$

where $\sigma_{jt}^2 = (1/\theta_{jt}) \sum_{x_{ijt}=1} \omega_{it}\, u_{it}^2$.

Equation (9) is the standard between/within decomposition of the variance. The between-group variance is the weighted sum of squared deviations across skill groups (first term of the right hand side of the equation) while the within-group variance is the weighted sum of the residual variance over skill groups (second term).

It is now clear how changes in the distribution of x affect both the predicted (between) and residual (within) components of the variance of wages. The sample proportions θ_{jt}, which are equal to $\bar{x}_{jt}$, completely describe the distribution of the x's in this dummy variables model.[7] Therefore, the change in the between group-variance induced by changes in sample composition is just an alternative way of describing the effect of changes in the variance of x on the predicted part of the variance of wages.

The second effect of changes in the distribution of human capital on the variance of wages is captured by the residual variance term $\sum_j \theta_{jt}\, \sigma_{jt}^2$. Since the overall residual variance is a weighted sum of cell-specific variances, it should systematically depend on the distribution of the x's through the sample proportions θ_{jt}.

These various effects can be captured in a variance decomposition by first considering the counterfactual variance obtained by replacing b_t (i.e. the $\bar{y}_{jt}$'s) in equation (9) by its value in period s:

$$V_t^a = \sum_j \theta_{jt}\, (\bar{y}_{js} - \bar{y}_t^a)^2 + \sum_j \theta_{jt}\, \sigma_{jt}^2. \qquad (10)$$

Just like the difference between $\bar{y}_t$ and the counterfactual mean $\bar{y}_t^a$, the difference between V_t and V_t^a represents the effect of changes in the price of human capital, b_t, between periods t and s. The variance that would prevail in period t if the distribution of x were also as in period s is obtained by replacing the sample proportions θ_{jt} by θ_{js}:

$$V_t^b = \sum_j \theta_{js} (\bar{y}_{js} - \bar{y}_t^b)^2 + \sum_j \theta_{js} \sigma_{jt}^2, \qquad (11)$$

where $\bar{y}_t^b = \sum_j \theta_{js} \bar{y}_{js} = \bar{y}_s$.

The counterfactual mean $\bar{y}_t^b$ represents the mean that would prevail if both the price of human capital (as in the case of $\bar{y}_t^a$) and the distribution of human capital were as in period s. This turns out to be the same as the actual mean in period s, $\bar{y}_s$. By contrast, the counterfactual variance V_t^b is not the same as V_s because of differences in the residual variance term σ_{jt}^2 and σ_{js}^2. Using the two counterfactual variances V_t^a and V_t^b, the change in the variance can be decomposed as:

$$V_t - V_s = (V_t - V_t^a) + (V_t^a - V_t^b) + (V_t^b - V_s). \qquad (12)$$

The first term on the right hand side is the contribution of changes in b_t to changes in the variance:

$$V_t - V_t^a = \sum_j \theta_{jt} [(\bar{y}_{jt} - \bar{y}_t)^2 - (\bar{y}_{js} - \bar{y}_t^a)^2]$$

The second term in equation (12) represents the contribution of changes in the distribution of x's (i.e sample proportions θ_{jt}) to both the between- and within-group variance:

$$V_t^a - V_t^b = [\sum_j \theta_{jt}(\bar{y}_{js} - \bar{y}_t^a)^2 - \sum_j \theta_{js}(\bar{y}_{js} - \bar{y}_t^b)^2] + \sum_j (\theta_{jt} - \theta_{js})\sigma_{jt}^2.$$

The last term captures changes in the residual variance within each cell:

$$V_t^b - V_s = \sum_j \theta_{js} (\sigma_{jt}^2 - \sigma_{js}^2).$$

These decompositions can be implemented in practice by computing the various elements of equations (10) and (11). An alternative is to compute the counterfactual variance V_t^a as the sample variance of the counterfactual wage y_{it}^a, and then use a re-weighting procedure to compute V_t^b. The main advantage of this latter approach is that it can also be used to compute *any other* distributional statistic.

3.3. Decomposing changes in the whole distribution: a re-weighting procedure

A comparison of equations (10) and (11) indicates that the counterfactual variance V_t^b is simply "reweighted" version of V_t^a in which the sample proportions θ_{jt} are replaced by θ_{js}. In other words, V_t^b can be written as:

$$V_t^b = \sum_j \theta_{jt}\, \psi_j (\bar{y}_{js} - \bar{y}_t^b)^2 + \sum_j \psi_j\, \theta_{jt}\, \sigma_{jt}^2,$$

where ψ_j is a re-weighting factor defined as

$$\psi_j = \theta_{js} / \theta_{jt}\ .$$

The counterfactual variance can also be written as a function of the counterfactual wage y_{it}^a using an individual-level reweighting factor ψ_i :

$$V_t^b = \sum_i \omega_{it}\, \psi_i\, (y_{it}^a - \bar{y}_t^b)^2,$$

where

$$\psi_i = \sum_j x_{ijt}\ \theta_{js} / \theta_{jt.} \qquad (13)$$

Since $\{x_{ijt}$, j=1,...,J$\}$ is a set of indicator variables for the various cells, equation (13) simply assigns to each observation the period s to period t ratio of the sample proportions of the cell to which it belongs.

The computational advantages of this re-weighting procedure are limited in the case of the variance since one could as well directly use the closed form formula in equation (11). It not generally possible, however, to derive similar closed form formula for most other popular distribution statistics such as the Gini coefficient or the interquartile range, or for point-by-point estimates of the distribution like the Lorentz curve or kernel density estimates.

By contrast, once the counterfactual wage y_{it}^a and the counterfactual weight $\omega_{it}^a = \omega_{it}\, \psi_i$ have been computed, it is straightforward to compute *any* counterfactual statistics. Since sample weights ω_{it} need to be used in the first place to compute sample statistics that are representative of the whole population, using the counterfactual weight poses no additional computing requirements relative to computing simple descriptive statistics.

Irrespective of the distributional statistic being computed, using y_{it}^a instead of y_{it} always shows what this statistic would have been if b_t had remained at its period s level. Similarly, using the weights $\omega_{it}^a = \omega_{it}\, \psi_i$ instead of ω_{it} yields the distributional statistic that would have prevailed if the distribution of x had remained as in period s.

The proposed procedure unifies existing procedures in the cell-by-cell case by combining elements from both Juhn, Murphy and Pierce (1993) and DiNardo, Fortin and Lemieux (1996). The idea of replacing y_{it} by y_{it}^a to compute any distributional statistic comes from Juhn, Murphy and Pierce (JMP thereafter), while the re-weighting procedure was suggested by DiNardo, Fortin and Lemieux (DFL thereafter). However, JMP did not suggest an explicit way of accounting for changes in the distribution of the x's, while DFL only partially addressed the issue of changes in the b's. Both of these papers also deal with the case where it is not feasible or practical to divide the observations in a limited number of cells. The proposed procedure is easily modified to account for this more general case. This is not much of an issue for computing $y_{it}^a = x_{it}b_s + u_{it}$ since usual regression and model specification methods can be used to choose an appropriate specification for the regression function. Similarly, DFL suggest using a standard logit or probit model to compute the reweighting factor ψ_i in the case where the cell-based approach cannot be used. The idea is to pool the period s and t samples and estimate a probit or logit model for the probability of being in year t. The estimated model yields a predicted probability of being in period t conditional on the x's. Call this predicted probability

$$P_{it} = \text{Prob}(\text{period}=t \mid x_{it}).$$

DFL then define the reweighting factor as:[8]

$$\psi_i = [(1- P_{it}) / P_{it}] \times [P_t / (1-P_t)],$$

where P_t the unconditional probability that an observation is in period t (the weighted share of the pooled sample which is in period t). In the case where a full set of dummies for the J cells are included in a probit or logit model, P_{it}/P_t is simply the sample proportion θ_{jt} and the reweighting factor is $\psi_i = \theta_{js} / \theta_{jt}$ as before. As in the case of the regression model, standard model specification procedures can be used to choose a sufficiently accurate specification for the logit or probit model when a fully unrestricted cell-by-cell approach is not feasible or appropriate.

Table 1 provides a summary of how various counterfactual values of y and ω can be combined to generate a variety of counterfactual distributions.

3.4 Reweighting vs Oaxaca-Blinder regression-based decompositions and propensity score estimation.

In the Oaxaca-type decomposition of the mean considered above, it was sufficient to compute a single counterfactual mean $\bar{y}_t^a = \bar{x}_t b_s$ to carry over the decomposition

$$\bar{y}_t - \bar{y}_s = (\bar{y}_t - \bar{y}_t^a) + (\bar{y}_t^a - \bar{y}_s).$$

The usual way of performing the decomposition involves estimating regressions for both time periods, though one could compute $\bar{y}_t^a$ by just estimating the regression model in period s. In the original case considered by Oaxaca (1973) where "group t" are women and "group s" are men, $\bar{y}_t^a = \bar{x}_t b_s$ is the average wage women would earn if they were paid according to the wage equation of men.

An alternative way of computing $\bar{y}_t^a$ is to reweight the male sample so that the distribution of x for men becomes the same as for women. Interestingly, in the "cell" model these two approaches yield *identical* results. This follows from the fact that in this model, $\theta_{jt} = \bar{x}_{jt}$ and $\bar{y}_{jt} = b_{jt}$, so that:

$$\bar{y}_t^a = \bar{x}_t b_s = \sum_j \bar{x}_{jt} b_{js} = \sum_j \theta_{jt} \bar{y}_{js} = \sum_j \psi_j' \theta_{st} \bar{y}_{js},$$

where $\psi_j' = (1/\psi_j) = \theta_{jt} / \theta_{js}$ is the reweighting factor used to transform the distribution of x for men (group s) into the distribution of x for women.

In the general case where it is not feasible or desirable to divide observations into a limited number of cells, the regression approach and the re-weighting approach do not yield numerically identical results. The results should nevertheless be very close if the regression model and the probability model like the one used by DFL are "well-specified".[9] In the case where the data can be divided in J cells, θ_{jt} and $\bar{y}_{jt}$ are non-parametric estimates of the conditional probabilities and conditional means, respectively. In the general case where observations cannot be divided in cells, it is nonetheless possible to estimate non-parametrically the conditional probabilities and conditional means using available methods. It could be shown in this case that both the (non-parametric) reweighting and the (non-parametric) regression method yield consistent estimates of the counterfactual $\bar{y}_t^a$.

Some of the issues that arise when decomposing the means also arise when looking at the whole distribution. For example, the results of the decomposition are sensitive to the order in which each factor is analyzed because of interactions between those factors. One pragmatic solution to this problem used in Figure 2 is to carry out the decomposition in different orders to verify the robustness of the conclusions.[10] It is also possible to look separately at the impact of each individual covariate and its coefficient, just like in the usual Oaxaca-Blinder decomposition.[11]

The reweighting approach is also closely linked to propensity score estimation that has recently been used by several researchers to estimate the impact of training programs or other "causal" effects.[12] For example, if group t represent the "treatment" group of individual receiving training while group s represents a "control group", the re-weighting factor ψ_i can be interpreted as a "propensity score" that represents the relative probability of being in the treatment group given the covariates x. Most papers simply use the propensity score as an additional regressor or "control function" on the right hand side of the outcome equation (e.g. a wage equation with a training dummy and other covariates on the right-hand side). However, in a recent paper, Hirano, Imbens, and Ridder (2000) suggest an alternative estimator that uses the propensity score to reweight the treatment group. This particular type of propensity score estimation turns out to be identical to the re-weighting approach suggested by DFL.

4. Empirical application no. 1: Wage distributions in Alberta and British Columbia

I now illustrate how the decomposition procedure of Section 3 works in practice using data for female workers in Alberta and British Columbia from the Labour Force Surveys (LFS) of January to October 2000. (Data from November and December are not used because the minimum wage was increased in British Columbia in November 2000, see below). Since January 1997, the LFS has collected detailed information on the wages and earnings of all wage and salary workers. Since each monthly sample of the LFS contains information on over 100,000 individuals throughout Canada, the resulting wage samples are very large, even at the provincial level.

Another feature of the LFS is that the socio-economic characteristics of individuals are only released at a relatively aggregated level in the public use files. In particular, age is only coded in 5-year intervals and education is coded as a categorical variable taking only seven possible values.[13] All individuals age 15 to 64 can thus be assigned to a total of 70 age-education cells. Since age and education are the two key human capital variables in Mincer-type wage regressions, the LFS lends itself naturally to the cell-based decomposition method discussed in Section 3.[14]

After removing few observations from very small cells or with very high or very low hourly wages, the final sample contains 19,319 observations for Alberta and 20,587 observations for British Columbia.[15] Each of the 57 remaining age-education cells contains about 350 observations by province, on average.

Table 2 reports several descriptive and counterfactual statistics for log wages in Alberta and British Columbia. The LFS sample weights are used throughout the table. Column 1 shows that the mean log wage is about 13 percent higher in British Columbia than in Alberta. By contrast, a comparison of rows 1 and 4 shows that the variance of log wages is substantially lower in British Columbia (0.199) than Alberta (0.218). Not surprisingly, the difference between the 90th and the 10th percentile (column 5), which is an alternative measure of wage dispersion, is also lower in British Columbia (1.196) than in Alberta (1.253). This means that the ratio of the wage at the 90th percentile over the 10th percentile is about 6 percentage points higher in Alberta than in British Columbia.

These descriptive statistics illustrate why Alberta and British Columbia are two interesting provinces to compare from a distributional point of view. Looking at the mean, the variance, and the 90-10 gap suggest that British Columbia is the "high wage/low dispersion" province while Alberta is the "low wage/high dispersion" province. A closer examination of the evidence suggests, however, that this may be an oversimplification of the facts. For example, columns 6 and 7 indicate that while the gap between 90th and the 50th percentiles is larger in Alberta than in British Columbia, as expected, the reverse is true for the 50-10 gap. This suggests that the shape of the wage distributions are quite different in the two provinces.

Figure 2a indicates that the two wage distributions look indeed very different. The figure reports kernel density estimates of the wage distributions in the two provinces.[16] While the density for Alberta looks more or less like the usual bell-shaped curve, the density for British Columbia is bimodal or "twin-peaked".[17] The figure also indicates the respective values of the minimum wage in the two provinces during this period. Throughout year 2000, the minimum wage was \$5.65 in Alberta. By contrast, the minimum wage was \$7.15 in British Columbia from January to October 2000 before increasing to \$7.65 in November 2000. This change in the minimum wage is the reason why the sample used only goes from January to October.

It is quite clear from the figure that the 25 percent difference in the minimum wage between the two provinces has a visually large impact on the wage distributions. The minimum wage does not appear to have much "bite" in Alberta. By contrast, the data strongly suggests that the higher minimum wage in British Columbia is the source of the "second peak" in the lower end of the wage distribution in this province. This large visual impact of the minimum wage on the wage distribution for women is quite consistent with the findings of DiNardo, Fortin and Lemieux (1996) for the United States, though the "twin-peak" distribution in British Columbia is more unusual.

It is nonetheless possible that factors other than the minimum wage explain why the wage distributions have such different shapes in the two provinces. For example, if education were highly polarized in British Columbia (e.g. many university graduates and high school dropouts with few people in between), this could result in a bimodal wage distribution in this province even in the absence of minimum wage effects. This can be formally tested by looking at whether the BC wage distribution would be unimodal if the distribution of age and education in the province were the same as in Alberta. Similarly, it would be interesting to see whether the wage distribution in Alberta would become bimodal if the distribution of age and education in the province were the same as in British Columbia. Both of these counterfactual distributions can readily be computed using the reweighting procedure suggested in Section 3.

The remaining panels of Figure 2 compare several counterfactual distributions to the actual distributions in Alberta and British Columbia. Figure 2b compares the actual wage density for British Columbia to the density that would prevail in Alberta if the regression coefficients b (the mean log wages by cell) were the same as in British Columbia. The counterfactual density is obtained by applying kernel density estimation on y_{it}^{a} instead of y_{it} (in this setting "t" refers to Alberta and "s" to British Columbia). On the one hand, this counterfactual densities lines up better with the BC density in terms of location parameters (mean or median). The lower and upper tails of the two distributions also look relatively similar. On the other hand, the relative shapes of the two distributions remain very different. Like the actual density for Alberta, the counterfactual density clearly has a single mode, which is very different from the actual density for British Columbia.

Figure 2c shows the density that would prevail in Alberta if both the regression coefficients b and the distribution of the covariates were as in British Columbia. In terms of the notation of Table 1, this counterfactual density is obtained by applying kernel density estimation on y_{it}^{a} using the weights ω_{it}^{a} instead of ω_{it}. The counterfactual is not visually very different from the counterfactual in Figure 2b, suggesting that differences in the distribution of the covariates in the two provinces have a modest impact on the wage distribution. Differences in the distribution of covariates clearly cannot account for the fact that the distribution of wages in bimodal is British Columbia but unimodal in Alberta.

The order of the decomposition procedure is reversed in Figure 2d. This figure compares the actual Alberta density to the density that would prevail in British Columbia if both the regression coefficients and the distribution of the covariates were as in Alberta. As in Figures 2b and 2c, the counterfactual density lines up much better in terms of location with the actual density for the other province than in Figure 2a. But once again, neither b nor the distribution of covariates can account for the twin-peaks in the BC density.

The impact of the various counterfactual exercises on summary measures of the distribution are reported in rows 6, 7 and 8 of Table 2. Consistent with the figures, most of the 13 percent gap in average wages is due to differences in the regression coefficients (row 6). Differences in the regression coefficients also account for .008 (row 6, column 2) of the .019 difference in variances, and for more than half (.036) of the .057 difference in the 90-10 gap.[18]

The effect of differences in the distribution of covariates between the two provinces is reported in row 7 of Table 2. This accounts for the remaining -.027 difference in mean log wages, suggesting that BC workers are more “skilled” (higher education and experience) than their Alberta counterparts. More interestingly, the effect of differences in the distribution of covariates has the “wrong sign” for the variance. It helps explain why the variance is *lower*, not higher, in Alberta than in British Columbia because of two offsetting effects. On the one hand, the distribution of age and education in British Columbia is more compressed than in Alberta. The variance of the predicted part of wages in Alberta is .002 higher (column 3) than it would be with the BC distribution of covariates. On the other hand, the residual variance is .005 lower (column 4) because of composition effects. Remember from Section 3 that this last effect is given by $\sum_j(\theta_{jt} - \theta_{js})\sigma_{jt}^2$, where θ_{jt} and θ_{js} represent the sample proportions in Alberta and in British Columbia in this setting. What happens is that there are systematically more workers in “high-dispersion” cells (σ_{jt}^2) in British Columbia than in Alberta. More specifically, workers are older and more educated in British Columbia and, as predicted by the human capital model of Section 2, residual wage dispersion is larger for these groups than for younger and less educated workers. These composition effects are substantial since they increase by 50 percent (from .009 in row 5 to .014 in row 8) the Alberta-British Columbia gap in the residual variance.

5. Skill prices and residual wage dispersion

In the previous example, the systematic part of the wage equation (xb) could not account for most of the Alberta-British Columbia difference in wage dispersion and in the shape of wage distributions. This is perhaps not surprising since most of the cross-sectional variance in wages remains typically unexplained in a standard Mincer-type wage regression. For instance, the R-square of the regressions estimated in Section 4 is around 35 percent, which is typical for these types of models.[19]

Under an unstructural interpretation of a regression equation, there is nothing else to be said about this last (residual) component of the decomposition. The residual is just the part of the dependent variable that cannot systematically be accounted for by the covariates. By contrast, the human capital approach presented in Section 2 has a variety of implications for the residual variation in wages. To the extent that residual wage dispersion is due to unmeasured differences in human capital investment, the residual dispersion should increase when the “price” or “return” to human capital increases. For instance, when the return to years of schooling increases, it is reasonable to expect that the return to (unmeasured) school quality would increase too.

To put this in a more explicit context, a large number of studies have documented a steep increase in both the return to education and residual dispersion in wages in the United States over the last two decades.[20] Under a strict human capital interpretation of wages, it is tempting to conclude that both of these phenomena are direct consequences of a pervasive increase in the return to human capital resulting from skill-biased changes in the demand for labour. This case that was first made forcefully by Juhn, Murphy and Pierce (1993) has become the standard explanation for the changes in wage inequality in the United States (see, for example, Acemoglu

(2002) and Katz and Autor (2000)).

JMP argue that residual wage dispersion is mostly a consequence of the fact that human capital or skills are imperfectly measured in standard data sets. Under the assumption that the distribution of the unmeasured skills is stable over time, increasing residual wage dispersion *must* be a consequence of an increase in the return to these unmeasured skills.

From an empirical point of view, however, it is not clear how this human capital-based theory of the growth in residual wage dispersion can be distinguished from other explanations. One alternative explanation is simply that the extent of measurement error in wages has increased over time. Another possible explanation suggested by DFL is the decline in the real value of the minimum wage during the 1980s. Later work by Lee (1999) and Teulings (2002) indeed suggested that most of the increase in residual wage dispersion can be linked to this factor.

Just stating that the increase in residual wage dispersion is due to an increase in skill prices is not enough to establish that this is indeed the right explanation for this phenomena. Other empirical implications of this "skill price" theory needs to be established in order to test it against alternative explanation. Otherwise, this theory is a mere tautology with no empirical content.

To fix ideas, consider the following model for the wage residual u_{it}:

$$u_{it} = p_t \eta_{it} + \epsilon_{it}, \qquad (14)$$

where η_{it} is unmeasured human capital, p_t is the price (or return) of this unmeasured human capital, and ϵ_{it} is a random error component not linked with skills and productivity (e.g. a measurement error). The variance of u_{it}, σ_t^2, is given by:

$$\sigma_t^2 = p_t^2 \sigma_{\eta,t}^2 + \sigma_{\epsilon,t}^2 ,$$

where $\sigma_{\eta,t}^2 = \text{Var}(\eta_{it})$ and $\sigma_{\epsilon,t}^2 = \text{Var}(\epsilon_{it})$. The most extreme version of the skill price story is that $\sigma_{\eta,t}^2$ is stable over time ($\sigma_{\eta,t}^2 = \sigma_\eta^2$) and $\sigma_{\epsilon,t}^2$ is either zero or stable over time. It follows that:

$$\sigma_t^2 - \sigma_s^2 = (p_t^2 - p_s^2)\sigma_\eta^2.$$

In this setting skill prices are the only source of change in residual dispersion, by assumption.

It is possible to introduce some empirical content to this model by looking at its implication for various group of workers. For instance, if residual skill dispersion σ_η^2 is larger for more experienced than less experienced workers, then the residual variance should increase more for the former group than the latter. This follows directly from

$$\sigma_{jt}^2 - \sigma_{js}^2 = (p_t^2 - p_s^2)\sigma_{j\eta}^2,$$

where $\sigma_{j\eta}^2$ is the variance of unmeasured skills for group j. Chay and Lee (2000) consider a version of this model in which they also assume that the variance of the measurement error, $\sigma_{\epsilon,t}^2$, is constant across skill groups. Using detailed data on the variance of wage by skill groups, they find evidence that the return to unmeasured skills, p_t^2-p_s^2, has indeed increased during the 1980s.

This approach can be pushed further when panel data is available. In this setting, it is natural to think of the η component of the residual as a time-invariant person-specific unmeasured skill factor. Baker (1997) and Baker and Solon (2002) both find that the return to this unmeasured skill factor, p_t, has increased over time.

One drawback of these two approaches is that they only focus on the implications of changing skill prices for the residual variance. A more flexible approach is required to go beyond a variance decomposition to see whether changes in the price of unmeasured skills can also account for more general distribution changes such as those illustrated in Section 4. The key limitation of equation (14) in this regard is that the wage residual u_{it} is assumed to be a linear function of unmeasured skills. This linear transformation strongly restricts the ways in which changes in the price of unmeasured skills can affect the wage distribution, holding the distribution of unmeasured skills constant.

Consider a more general setting in which the price p_t is replaced by a non-linear pricing scheme

$$u_{it} = p_t(\eta_{it}) + \epsilon_{it}, \qquad (15)$$

where $p_t(.)$ is a monotonic and continuous function. For simplicity, assume from now on that ϵ_{it} is equal to zero.

This model is completely flexible in the sense that any distribution of u_{it} can be generated from an arbitrary distribution of skills η_{it}. For example, assume that η_{it} follows a uniform distribution over the [0,1] interval. This choice of distribution is convenient since

$$\eta_{it} = F_t(u_{it}),$$

where $F_t(.)$ is the cumulative distribution function of u_{it}. It follows that

$u_{it} = p_t(\eta_{it}) = F_t^{-1}(\eta_{it})$.

In this setting, η_{it} can be interpreted as the rank (normalized from 0 to 1) of observation i in the distribution of residuals, while the non-linear skill price function $p_t(.)$ is the inverse of the cumulative distribution function of u_{it}. This approach is a special case of the procedure first suggested by Juhn, Murphy and Pierce (1993). I discuss their decomposition approach in more detail in Section 6 and 8.

This model has no empirical content in single cross-section because $p_t(.)$ is left completely unrestricted. However, the model does impose restrictions on the change in the entire wage distribution from a period to another. To see this, substitute equation (15) into equation (2):

$y_{it} = x_{it}b_t + u_{it} = x_{it}b_t + p_t(\eta_{it})$.

Section 3 showed how to construct a counterfactual distribution of wages that accounts for changes in b_t and in the distribution of x_{it} from period t to s. The natural next step of the decomposition is to replace the residuals at time t by the residuals that would have prevailed if the skill pricing function had been $p_s(.)$ instead of $p_t(.)$. This amounts to replacing the counterfactual wage

$y_{it}^a = x_{it}b_s + u_{it} = x_{it}b_s + p_t(\eta_{it})$.

with

$y_{it}^b = x_{it}b_s + p_s(\eta_{it}) = x_{it}b_s + u_{it}^b$,

where $u_{it}^b = p_s(\eta_{it}) = F_s^{-1}F_t(u_{it})$ is a counterfactual residual. It is now clear why the model has some empirical content. If changes in the skill pricing function can account for all the changes in the wage distribution that changes in b_t and in the distribution of x_{it} could not account for, then the distribution of y_{it}^b reweighted by the factor ψ_i should be the same as the raw distribution of wages in period s.

Figure 3 illustrates a simple example that shows why this needs not be the case. Consider two groups of workers, L and H. Type L workers earn an average of 4 dollars in both periods s and t, while type H workers earn an average of 8 dollars in each period. In both periods, each group represent half of the sample. In terms of the notation of Section 3, this means that $\theta_{Lt}=\theta_{Ls}=\theta_{Ht}=\theta_{Hs} = .5$, $\bar{y}_{Lt} = \bar{y}_{Ls} = 4$, and $\bar{y}_{Ht} = \bar{y}_{Hs} = 8$. I use this simple case where the regression coefficients (i.e. the $\bar{\gamma}$'s) and the distribution of the covariates (i.e. the θ's) play no role in changes in the overall wage distribution to illustrate more clearly the role of changes in the distribution of the residuals.

Assume that in period t, the residuals follow a uniform distribution over [-3,3] for type L workers (density of 1/6) and over [-1,1] for type H (density of ½). The resulting wage distribution, which is skewed to the left, is shown in Figure 3a. Now assume that the distribution of the residuals switches between the two groups from period t to s (uniform distributions over [-1,1] and [-3,3] for type H and L, respectively). The resulting distribution in period s, which is now skewed to the right, is illustrated in Figure 3b. However, the overall distribution of residuals remains unchanged since it just a mixture of the same uniforms distributions in both periods. Figure 3c shows that this mixture distribution has a density of 1/12 over the range [-3,-1] and [1,3], and a density of 4/12 over the range [-1,1]. Since the cumulative distribution of residuals does not change over time, $F_s^{-1}F_t(u_{it}) = u_{it}$ and changes in the pricing function $p_t(.)$ fail to explain any of the difference between the wage distributions in period t and s.

The problem is that the same transformation function $F_s^{-1}F_t(.)$ cannot explain both the narrowing in the residual distribution for group L and the widening in the residual distribution for group L. Another way to put this is that the transformation function $F_s^{-1}F_t(.)$ is a flexible monotonic function that does not, however, change the rank η_{it} of each residual in the distribution of residuals. Consider, for example, the rank in the overall distribution of residuals of the type H worker with the lowest possible residual in period t (-1). It is clear from Figure 3c that the rank of this worker is 1/6 in period t. In period s, however, the residual for the same type H worker is now -3 with a rank of 0. In other words, it is not possible to transform the wage distribution of period t to the one of period s without changing the rank of the residuals of individual workers. By assumption, the skill pricing function does not affect the rank of the residuals. Therefore, changes in the skill pricing function cannot explain the changes in the wage distribution.

Obviously, the story about changes in the price of unmeasured skills can be rescued by assuming different changes in prices for H and L type workers. One justification for this assumption is that the nature of unmeasured skills may not be the same for the two types of workers. Perhaps we are dealing with unmeasured "cognitive" skills for type H workers and unmeasured "manual" skills for type L workers. Under this scenario, changes in the wage distribution can fully be accounted for by a decline in the price of manual skills (by a factor of 3 to 1) and an increase in the price of cognitive skills (by a factor of 1 to 3).

The problem with this alternative skill price story is that it is no longer imposes any testable restrictions on the changes in the distribution of wages. By assumption, changes in the residual distribution of wages within each sub-group of the population are now explained by changes in skill prices specific to this particular sub-group. While this may be true, this is an empirically untestable explanation that does not help understand the observed distributional changes.

6. Accounting for changes in the residual pricing function in the decomposition.

This section shows how to compute empirically a counterfactual distribution that accounts for changes in the skill pricing function $p_t(.)$. The idea of constructing a counterfactual wage like $y_{it}^b = x_{it}b_s + u_{it}^b$ (see Section 5) was first suggested by JMP who suggest a simple procedure for computing the counterfactual residual u_{it}^b. Their idea is to first compute the rank $\eta_{it} = F_t(u_{it})$ from the empirical distribution of residuals and then select the residual at the same rank in the empirical distribution of residuals in period s:[21]

$$u_{it}^b = F_s^{-1}(\eta_{it}).$$

These imputed values of u_{it}^b and $y_{it}^b = x_{it}b_s + u_{it}^b$ can now be used to generate a counterfactual wage distribution that would have prevailed in period t if the price of measured human capital and the pricing function of unmeasured skills had been at their period s level (i.e. b_s and $p_s(.)$ instead of b_t and $p_t(.)$). As before, these counterfactual wages can be combined with the counterfactual weight ω_{it}^a to also control for changes in the distribution of covariates. This extends JMP's procedure that does not account explicitly for changes in the distribution of covariates. The resulting counterfactual distribution should now account for all the changes from period t to s except for those involving a change in the rank of the residual in the overall distribution of residuals. Remember from Section 5 that those are the changes that cannot be reconciled by a change in a single pricing function of unmeasured skills. This happens when the distributions of residuals for different groups change differently in a way that cannot be captured by a common transformation function.

An alternative interpretation of this decomposition exercise is that the transformation

$u_{it}^{b} = F_s^{-1}F_t(u_{it})$

yields a counterfactual distribution of residuals for period t that is the same as in period s, just like replacing b_t by b_s and reweighting by ω_{it}^{a} yields a distribution of the predicted wage $x_{it}b_s$ which is the same as in period s. However, combining the counterfactual wage residual with the counterfactual predicted wage does not generally yield the same (counterfactual) distribution as the actual distribution in period s.

The theoretical reason for this is that when a random variable y is the sum of two random variables xb and u, knowledge of the marginal distributions of xb and u is not generally sufficient to characterize the marginal distribution of y. In general, an infinite number of marginal distributions of y is compatible with given marginal distributions of xb and u. The exception is the case where xb and u are independent. A unique marginal distribution of y can then be obtained from the convolution of the distributions of xb and u. Since independence means that the distribution of residuals is the same for all values of xb, there is no scope for the rank of residuals to change over time.

By construction, xb and u are uncorrelated in the regression model. On the one hand, since zero correlation is much weaker than independence, an infinite number of wage distributions (including the period t counterfactual and the actual period s distribution) are compatible with the marginal distributions of xb and u in period s. On the other hand, zero correlation implies that the variance of the marginal distribution of wage is just the sum of the variance of xb and u (covariance term is zero). The variance of the counterfactual wage distribution obtained using y_{it}^{b} weighted by ω_{it}^{a} should thus be the same as the variance in period s.

This is another way of illustrating the empirical content of this last counterfactual exercise. Changes in the pricing function $p_t(.)$ are such that they explain all the remaining change in the variance of wages between two periods. The question is whether they can also explain more detailed aspects of distributional changes such as inter-percentile gaps or the shape of the distribution? This is an empirical question that I address in the second example in the next section.

7. Empirical application no. 2: Changes in wage inequality in the United States.

Following Juhn, Murphy, and Pierce (1993), the consensus in the literature is that residual or within-group inequality has increased steadily since the 1970s as a result of an increase in the return to unmeasured skills (Katz and Autor, 2000, Acemoglu, 2002). This case is thus well-suited to the extended decomposition method proposed in Section 5 and 6.

I re-examine this question using hourly wage data from the May 1973 Current Population Survey (CPS) and from the 1979, 1989 and 1999 outgoing rotation group files of the CPS. (These data are only available in the May supplement of the CPS from 1973 to 1978). The sample used and summary statistics are broadly similar to those of DFL, Katz and Autor (2000), and Card and DiNardo (2002). Following these studies, I weight all observations by the CPS sample weight multiplied by usual weekly hours of work. Details about data construction are provided in Appendix 1.

Unlike the Canadian LFS, the public-use files of the CPS provide detailed age and education categories. Following most of the literature, I also include marital status and race to the list of covariates. As a result, it is not longer feasible to use the cell-by-cell approach of Section 4 because there are too many cells. I use instead a flexible functional form for both the regression model and the logit model used to construct the reweighting factor. In both cases, the covariates used are a set of six education dummies fully interacted with a quartic function in experience, years of schooling, a marital status dummy and a race (white/non-white) dummy. The marital status and race dummies are also interacted with years of experience. Separate models are estimated for men and women.

Figure 4 plots kernel density estimates for men in 1973, 1979, 1989 and 1999. Consistent with previous studies, there is a clear widening in the wage distribution over this time period. Another noticeable fact is the impact of the minimum wage at the lower end of the distribution. In 1979 dollars, the minimum wage increased from 2.62 in 1973 to 2.9 in 1979, declined sharply to 1.96 in 1989 and recovered to 2.24 in 1999.[22] Consistent with DFL, panel b shows a clear spike around the relatively high minimum wage of 1979. By 1989, however, the minimum wage is so low is has no noticeable impact on the lower end of the wage distribution. The increase in the minimum wage in the 1990s also appears to have an impact on the density of wages around 2 dollars.

As in DFL, Figure 5 shows that the minimum wage has a much larger visual impact for women than men. Since women earn lower wages, on average, the minimum wage is relatively much higher for women than men. In particular, the minimum wage has a huge visual impact on the wage distribution in 1979. The minimum wage spike is in fact the mode of the distribution in that year. The dramatic change in the shape of the distribution between 1979 and 1989 appears to be driven by the steep decline of the minimum wage during the 1980s. As in the case of men, the recovery in the real value of the minimum wage during the 1990s helped move up the lower tail of the distribution.

Table 3 reports the results of the decompositions for the variance and the 90-10, 50-10, and 90-50 gaps. As documented in many other studies, there is a sharp increase in all dimensions of wage inequality during the 1980s for both men and women. Both the variance and the 90-10 increase at a much slower pace during the 1990s. This is consistent with a smaller number of studies like Card and DiNardo (2002) and Gosling and Lemieux (2002) that look at recent changes in wage inequality. For both men and women, however, the modest growth in the 90-10 gap in the 1990s masks a decline in the 50-10 gap and a substantial increase in the 90-50 gap. The recovery in the real value of the minimum wage during the 1990s is consistent with the decline of the 50-10 gap during this period. More generally, the 50-10 gap for women appears to be strongly linked to the real value of the minimum wage. The 50-10 gap decreases during both the 1970s and the 1990s when the real value of the minimum wage increases. By contrast, it expands dramatically as the real value of the minimum wage collapses during the 1980s.

For the 1970s, measures of inequality are stable for men but tend to decline for women because of the substantial decline in the 50-10 gap. This finding is more or less consistent with the limited number of studies that have looked at inequality in hourly wages.[23] By contrast, most other studies look at earnings inequality and conclude that inequality increased significantly during the 1970s.[24] Reconciling the behavior of hourly wage and earnings (wage times hours) inequality during the 1970s is beyond the scope of this paper. Since theories of wage determination including human capital theory typically focus on the hourly price of labor, I only analyze the inequality in hourly wages in this paper.

The decomposition results for men in the 1980s (column 2) are very similar to those of JMP for 1979-88. JMP find that the 90-10 gap increased by .208 compared to .169 here. As in Table 3, they find that changes in the "prices of observables" (b_t) account for more than half of the change and that the role of changes in the distribution of "observables" (distribution of x) is negligible. In their setting, residual changes in inequality account for about 40 percent of the total change, which is similar to the contribution of changes in the pricing function of unmeasured skills in Table 3. I discuss in more detail in Section 8 how the decomposition procedure of JMP compares to the one used here.

Consistent with the existing literature, the results thus suggest that the increase in inequality for men in the 1980s is due to an increase in both the return to measured characteristics and the pricing function of unmeasured skills. The results in column 6 suggest that a broadly similar conclusion could be reached for women.

Those conclusions change substantially, however, when the period of analysis is expanded beyond the 1980s. Most importantly, the role of changes in the pricing function of unmeasured skills for men becomes very small when the whole 1973-99 period is considered in column 4. For both the variance and the 90-10 gap, this source of change now accounts for less than 10 percent of the increase in inequality between 1973 and 1999. By contrast, changes in the distribution of x now account for about a third of the change in inequality, leaving the contribution of changes in b_t over 60 percent, as in the 1980s.

Panel B shows that this reversal in the role of $p_t(.)$ and the distribution of x is due to composition effects in the variance of residuals. In the 1980s, 0.021 of the 0.030 increase in the residual variance is attributable to the rise in price of unmeasured skills while composition effects only account for 0.009. However, composition effects are also important in the 1970s (0.004) and the 1990s (0.010) while the variance of the residuals is relatively stable. Over the whole 1973-99 period, composition effects now account for a 0.023 growth in the variance, leaving only 0.004 to changes in skill prices. A more detailed analysis of the variance of the residuals by groups of workers indicates that it is systematically higher for more experienced and more educated workers.[25] The composition effects are due to the fact that the working age population is getting older and more educated. Increasingly more weight is thus put on groups with higher residual variances.[26]

Results for women are qualitatively similar. Over the 1973-99 period, changes in the price of unmeasured skills account for only a small fraction of the growth in inequality. The bulk of the change is explained by changes the prices of measured skills or in their distribution. Note also that part of the change (0.037) in the 90-10 gap cannot be explained by any of these three factors. As discussed in Section 5, this suggests that changes in the skill pricing function of unmeasured skills cannot account for all the residual change in the wage distribution. Panels E and F show that a large fraction of changes in the 50-10 and 90-10 gaps cannot be explained by the three factors either. Clearly, there are some changes in the shape of the wage distribution that those three factors have a hard time explaining.

These changes in shape are examined in more detail in Figure 6. The figure shows changes in real wages at each percentile of the wage distribution. This useful way of summarizing wage changes at different points of the distribution was used extensively by JMP. One minor difference relative to JMP is that I use the "normits" of the different centiles (inverse of the cumulative standard normal distribution) on the x-axis.[27] Figure 6 shows both the overall changes by percentiles for each of the three sub-periods as well as the unexplained change that cannot be accounted for by any of the three factors.

If the three factors explained perfectly all the change in the wage distribution, then the residual change would be a horizontal line. Departures from the flat line indicate the extent to which the three factors fail to explain changes in the shape of the wage distribution. A simple characterization of Figure 6 is that the three factors are quite successful at explaining the linear trend in wage changes as a function of the normits. They are much less successful, however, at explaining departures from linearity. For example, for men in the 1980s, the three factors explain virtually all the change in wages by percentiles above the 10th percentile where the wage changes are more or less a linear function of the normits. By contrast, changes in the minimum wage induce some non-linearity at the very bottom of the distribution that ends up in the unexplained changes. Not surprisingly, the three factors have an even harder time explaining the drop in female wages around the 5-10th centile during the 1980s which is also likely linked to the decline in the real value of the minimum wage. By contrast, changes above the 20^{th} percentile are well explained by the three factors. Unexplained changes in the lower end of the distribution in the 1970s are a mirror image of what happened in the 1980s since there is an unexplained wage gain at the lower end of both the male and female distributions. This unexplained gain is consistent with wages at the low end going up because of the increase in the minimum wage.

The 1990s is another interesting case since wage changes by percentiles are U-shaped during this period, with both the lower and higher wage percentiles growing more than the median. Most of the U-shape remains in the unexplained part of the change, suggesting once again that the three factors have a hard time capturing the non-linear part of the changes by percentiles.

Figure 7 shows the changes in the wage distribution by percentiles over the whole 1973-99 period. For both men and women, there is a clear expansion in wage inequality above the 25^{th} percentile (Figures 7a and 7c). Below the 25^{th} percentile, however, the wage gains tend to be more substantial at the lower end. Except below the 5^{th} percentile, wage changes are thus a convex function of the normits. Much of this convexity remains in the unexplained part of the wage changes.

Figures 7b and 7d show the contribution of the three factors to overall changes in wages by percentile for men and women, respectively. Consistent with the results in Table 3, changes in unmeasured skill prices (the $p_t(.)$ function) explains little of the change in the male wage distribution. It accounts, however, for some of the convexity in wage changes. By contrast, change in the regression coefficients b_t account for most of the growth in inequality above the 25^{th} percentile. Wage changes due to changes in the distribution of the covariates are not as systematically linked to the overall change in wages. Most of the impact of this factor on inequality is below the 50^{th} percentile, which is not the place where most of the growth in inequality takes place. In fact, Panel E of Table 3 (column 4) shows that the effect of changes in the distribution of covariates (0.068) is larger than the overall change in the 50-10 gap (0.038). As a result, the three factors *over explain* the change in the 50-10 and the unexplained change is large and negative. By contrast, the three factors *under explain* the change in the 90-50 gap (Panel F). So while unexplained changes not accounted by the three factors are negligible for the 90-10 gap, they play a substantial role in changes in the 50-10 and 90-50 gaps. This confirms the overall message that the three factors do not explain very well changes in the shape of the wage distribution.

Figure 7d shows the impact of the three factors for women. The results are generally similar to those for men. As in the case of men, changes in the regression coefficients b_t are the most important explanation for overall changes in wages in the middle and upper part of the wage distribution. Similarly, the impact of changes in unmeasured skill prices is convex while the impact of changes in the distribution of covariates is concentrated below the 50^{th} percentile. As in the case of men too, Table 3 shows that the three factors over explain changes in the 50-10 gap but under explain changes in the 90-50 gap.

In summary, the main findings of this empirical application are:

A) Changes in the regression coefficients is the most important explanation for changes in wage inequality between 1973 and 1999.

B) Over the same period, most of the increase in the variance of the residuals is due to composition effects as opposed to an increase in the price of unmeasured skills. This is in sharp contrast with the 1979-89 period in which prices of unmeasured skills account for most of the increase in the variance of residuals.

C) The three systematic factor (changes in b_t, $p_t(.)$, and in the distribution of covariates) do not explain very well the non-linear aspects of changes in wages by percentile. In particular, they fail to account for minimum wage effects at the lower end of the distribution. This is particularly clear in the case of women in the 1980s.

The finding that most of the increase in the variance of the residuals is due to composition effects as opposed to an increase in the price of unmeasured skills is at odd with most of the existing literature summarized by Katz and Autor (2000) and Acemoglu (2002). As illustrated in Table 3, part of this discrepancy stems from the fact that I use data up to the late 1990s. Another source of difference is that, unlike most other studies, I account explicitly for composition effects in the dispersion of residuals.

8. Other decompositions methods

In this section, I contrast the proposed decomposition methods with other regression-based methods as well as more general procedures.

8.1. Juhn, Murphy and Pierce (JMP) method.

The JMP methods aims at decomposing change in the wage distribution in three components: changes in observable prices (b_t here), changes in observable quantities (changes in the distribution of covariates here), and residual changes. As mentioned earlier, there is no difference between the way JMP and Section 3 deal with b_t, the idea being simply to replace b_t by some counterfactual b. The main difference is that the JMP procedure does not account explicitly for changes in the distribution of the covariates. What JMP do instead is to account for changes in the distribution of wage residuals using their residual imputation procedure described in Section 6, and then define the effect of changes in covariates as what is left unexplained by the first two factors. By contrast, the procedure proposed in Section 3 accounts explicitly for changes in the distribution of covariates.

The outcome of this decomposition depends critically on how changes in the distribution of residuals is modelled. To see this, note that JMP formally allow for the distribution of residuals to depend on the covariates:

$$u_{it} = F_t^{-1}(\eta_{it}|x_{it}). \qquad (16)$$

From Section 5, it is easy to see what the JMP decomposition would yield if the distribution of residuals were not allowed to depend on the x's, i.e. if we had $u_{it} = F_t^{-1}(\eta_{it})$ instead of equation (16). This is a special case of the JMP decompositions that has been used in several studies such as Blau and Kahn (1997). Call this first version of the decomposition JMP1.[28] Leaving aside issues linked to the order of the decomposition, the residual imputation procedure of JMP1 would only capture what I call in Section 5 the effect of changes in the pricing function of unmeasured skills (changes in the $p_t(.)$ function). What JMP call the effect of observable quantities (distribution of covariates) would then be a mix of the true effect of changes in the distribution of covariates plus the change unexplained by the three factors. It is clear from the example of Section 7 that this unexplained change can be quite substantial.

At the other extreme, if the residuals are allowed to depend on the x's in a very flexible way, then the residual imputation procedure should capture all the residual change in the distribution of Section 3. Call this second version of the JMP procedure JMP2. It is easier to see how the procedure works in the cell-by-cell case considered earlier. In this case, the counterfactual residual

$$u_{it}^{b} = F_s^{-1}(\eta_{it}|x_{it})$$

can be obtained by applying the residual imputation procedure within each cell. In other words, the rank η_{it} of the residual u_{it} can be computed within the relevant cell. The imputed residual u_{it}^{b} is the residual at the same rank in the period s distribution of residual for the same cell. As a result, the distribution of residuals for observations with a given value of x_{it} is being replaced by the period s distribution. This is equivalent to starting with the wage distribution in period s and reweighting with the factor $1/\psi_i$ to get the same distribution of x as in period t. The reweighted distribution now has the same distribution of x's as in period t but the distribution of residuals (within each cell) of period s is preserved.

It would be more difficult to implement JMP2 in the case where it is not feasible or desirable to divide the data in a relatively small number of cells. Some alternative estimation procedures discussed below would have to be used to estimate the distribution of the residuals, conditional on the x's. While some estimation issues are also involved in construction the reweighting factor, the estimation problem linked with finding an appropriate functional form for a logit or a probit remains much simpler than carrying over some flexible estimation of conditional wage distributions. Note that Juhn, Murphy, and Pierce (1993) do not discuss how they exactly condition on x when implementing their procedure empirically. It is therefore difficult to know whether they use what I call the JMP2 procedure or something which is in-between JMP1 and JMP2.

In summary, the JMP2 procedure is equivalent to the re-weighting procedure of Section 3. The main advantage of the reweighting approach is that it is much easier to implement empirically. For reasons discussed in Section 5, it does not seem appropriate to attach particular economic interpretations, such as an increase in prices of unmeasured skills, to the residual part of the decomposition that comes out of these procedures. By contrast, it is legitimate to interpret as evidence of an increase in the price of unmeasured skills the outcome of the residual imputation procedure in JMP1. The problem with JMP1 is that the effect of changes in the distribution of covariates and changes that are unexplained by the three factors of Section 5 are mixed together.

8.2. Modelling the conditional distribution of wages.

When the data can be divided in a limited number of cells, the decomposition procedure of Section 3 can be implemented automatically without worrying about specification issues. I also argue that when the data cannot be divided into a small enough number of cells, the same decomposition method can be used by estimating flexible functional forms for the regression equation and the logit or probit re-weighting model. However, both the decomposition results and their statistical properties depend on how those functional forms are chosen. Other decomposition methods that involve different parametric restrictions are also available in the literature.

For example, Gosling, Machin and Meghir (2000) and Machado and Mata (2002) propose an alternative decomposition procedure based on the estimation of quantile regressions. They consider the following regression model for the θ^{th} quantile of y conditional on the covariates x (θ goes from 0 to 1):

$$Q_\theta(y,x) = x\beta(\theta). \tag{17}$$

It is easier to understand how quantile regressions work in the cell-by-cell case.[29] Consider, for example, the median (θ=.5). The median regression can be estimated by first computing the median within each cell and then running a standard regression of the medians on x. This yields an estimate of $\beta(.5)$. Similar regressions can also be run for all other quantiles including the 10^{th}, the 90^{th}, etc. In theory, running regressions for all possible quantiles should describe the whole conditional distribution of wages.[30] It is then possible to use the models estimated for various periods to construct counterfactual distributions.

Donald, Green, and Paarsch (2000) use an alternative method based on the estimation of hazard models. The hazard function for y given x is:

$$h(y|x) = f(y|x)/S(y|x),$$

where $f(y|x)$ is the conditional density and $S(y|x) = 1- F(y|x)$ is the survivor function. Since the survivor is a function of the density, estimates of the density (or of the cumulative distribution) can be inferred back from the hazard. Donald, Green, and Paarsch then rely on the large literature on the estimation of duration models to estimate the hazard function. The estimates can then be used to generate a variety of counterfactual distributions.

It is beyond the scope of the paper to compare in details the performance of the various methods in actual applications. The method proposed in Section 3 is the simplest to implement, but other methods may have particular advantage depending on the setting. For example, when the dependent variable is an actual duration (like the duration of unemployment spells), the Donald, Green, and Paarsch is the most natural method to use since the hazard function plays a role similar to the regression function in wage determination models.

8.3. Index models

When looking at the distribution of wages, a natural generalization of the Mincer-type human capital equation is the following index model:

$$y_{it} = p_t(x_{it}a + \eta_{it}). \tag{18}$$

In economic terms, the function $p_t(.)$ represents a general skill pricing function while $x_{it}a + \eta_{it}$ is human capital (both measured and unmeasured). Relative to the model I consider in Section 5, the flexible pricing function $p_t(.)$ now applies to both measured and unmeasured human capital instead of unmeasured human capital only. Which specification is more accurate depends on the underlying version of the human capital model. The earlier approach of using different pricing schemes for measured and unmeasured human capital is based on the idea that these are two very different types of human capital.

However, this is not necessarily consistent with a Mincer type model. For instance, in the pure OJT model there is only one type of human capital (training). The x variable (potential experience here) captures the mean value while residuals capture dispersion in OJT investments for individuals with same level of experience. Since there is only one type of human capital in this setting, it is appropriate to use a single pricing function.

In terms of equation (18), Mincer focuses on the case where the pricing function p_t is linear when y_{it} is log wages. One possible generalization is to use a Box-Cox specification (see Heckman and Polachek, 1974) which is consistent with a range of functional forms for $p_t(.)$. More recently, Fortin and Lemieux (1998) and Teulings (2000, 2002) have used flexible methods to estimate the function $p_t(.)$. Fortin and Lemieux assume that η_{it} is distributed i.i.d. normal and discretize y_{it} in 200 intervals. In this setting, the model can be estimated using an ordered probit model and the function $p_t(.)$ can be recovered from the estimated thresholds of the ordered probit. Interestingly, they find that $p_t(.)$ is relatively linear except around the value of the minimum wage.

Teulings uses instead high order polynomials as a flexible function form for $p_t(.)$. Instead of assuming a specific distribution for η_{it}, he allows for a flexible distribution for y_{it}. The main advantage of Teulings approach relative to others is that it can be shown to be consistent which an underlying production function where the elasticity of substitution between skill groups is distance-dependent. This enables him to interpret within a well-specified general equilibrium model the large distributional impacts of the minimum wage.

An informal examination of the evidence in Fortin and Lemieux (1998) and Teulings (2002) suggest that these single-index models are more successful at capturing changes in the shape of wage distributions than the regression-based method of Section 5. It is easy to understand why this is the case in the presence of minimum wage. Rewriting equation (18) as:

$$p_t^{-1}(y_{it}) = x_{it}a + \eta_{it}$$

shows that the pricing function allows to transform the distribution of wages in a flexible way, just like $p_t(\eta_{it})$ allowed to transform the distribution of residuals in a flexible way in Section 5. Since the minimum wage has a specific impact at a particular point in the wage distribution, this can be easily captured by the $p_t(.)$ function in the index models. By contrast, the impact of the minimum wage on the distribution of residuals in a standard wage regression depends on the value of the covariates. For workers with low education and experience, the minimum wage has an impact higher up in the distribution of residuals than for more skilled workers. A unique pricing function for the residuals are all groups pooled together cannot, therefore, capture the effect of the minimum wage adequately.

Nonetheless, the decomposition methods suggested in Section 3 and 5 remain a useful first step in analyzing the effect of various factors on changes in the distribution of wages. They help establish a diagnostic for which particular features of the distribution may require further modelling using more complicated approaches like these index models.

9. Concluding comments

The main contribution of this paper is to propose a simple procedure for decomposing changes in the wage or other distributions into three factors: changes in regression coefficients, changes in the distribution of covariates, and residuals changes. The procedure is easy to implement as it only requires estimating standard OLS regressions augmented by a logit or probit model. I also show how the procedure can be extended by modelling residuals as a function of unmeasured skills and skill prices.

The proposed procedure helps clarify some of the conflicting results in the literature on changes in wage inequality over the last decades. Like most other studies, I find that increases in the returns to measured skills like experience and education play a major role in secular increases in wage inequality in the United States. I also find, however, that this explanation does not account well for the changes at the bottom end of the wage distribution. More importantly, I find that much of the increase in residual wage inequality is due to changes in the composition of the workforce. This suggests that increases in the price of unmeasured skills does not play much of a role in the overall growth in wage inequality since 1973. There is an interesting parallel between these findings and those of Lee (1999) and Teulings (2002) who find that, in the 1980s, changes in the minimum wage explain essentially all the change in the lower end of wage distribution and in the residual wage inequality. However, both of these authors still find, as I do, that changes in the returns to measured skills play a significant role in the growth in inequality.

The proposed procedure can also be viewed as an extension to the distributional case of the well-known Oaxaca-Blinder decomposition. I show in an empirical example how differences in the whole wage distribution of women in Alberta and British Columbia can be divided into the effect of the regression coefficients and covariates, just like it is commonly done in decompositions of the mean.

Finally, I only focus on estimation in this paper. What about statistical inference? Given the nature of the decomposition procedure used, the most natural way of conducting statistical inference is to use a bootstrap procedure, i.e. resample with replacement from the data sets for both time periods. The whole decomposition procedure is then repeated on these new samples over and over again to construct an empirical distribution of the statistics of interest. It would be particularly important to go through such procedures in cases where the number of observations available is limited.

References

Acemoglu, Daron (2002) 'Technical Change, Inequality, and the Labor Market,' *Journal of Economic Literature* 40, 7-72

Atkinson, Anthony (1970) 'On the Measurement of Inequality,' *Journal of Economic Theory* 2, 244-63.

Baker, Michael (1997) 'Growth-Rate Heterogeneity and the Covariance Structure of Life-Cycle Earnings,' *Journal of Labor Economics* 15, 338-75

Baker, Michael and Gary Solon (2002) 'Earnings Dynamics and Inequality among Canadian Men, 1976-1992: Evidence from Longitudinal Income Tax Records," forthcoming in the *Journal of Labor Economics*

Barsky, Robert, John Bound, Kerwin Charles, and Joseph Lupton (2001) 'Accounting for the Black-White Wealth Gap: A Nonparametric Approach,' NBER working paper no. 8466

Beaudry, Paul, Fabrice Collard and David Green (2002), 'Decomposing the Twin-Peaks: A Study of Changing World Distribution of Output-per-Worker,' University of British Columbia mimeo.

Beaudry, Paul, and David Green (2002) 'Changes in U.S. Wages 1976-2000: Ongoing Skill Bias or Major Technological Change?,' NBER Working paper no. 8787

Becker, Gary (1975) *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education, 2nd edition* (New York: NBER)

Ben-Porath, Yoram (1967) 'The Production of Human Capital and the Life-Cycle of Earnings,' *Journal of Political Economy* 75, 352-65

Blau, Francine D., and Lawrence M. Kahn (1997) "Swimming Upstream: Trends in the Gender Wage Differential in 1980s", *Journal of Labor Economics* 15, 1-42

Blinder, Alan (1973) 'Wage Discrimination: Reduced Forms and Structural Estimation,' *Journal of Human Resources* 8, 436-455

Bound, John, and George Johnson (1992) 'Changes in the Structure of Wages in the 1980s: An Evaluation of Alternative Explanations,' *American Economic Review* 82, 371-92

Buchinsky, Moshe (1994) 'Changes in the U.S. Wage Structure 1963-1987: Application of Quantile Regression,' *Econometrica* 62, 405-458

Card, David, and John DiNardo (2002) ‘Skill Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles,’ NBER Working paper no. 8769

Card, David, and Thomas Lemieux (1996) ‘Wage Dispersion, Returns to Skill, and Black-White Wage Differentials,’ *Journal of Econometrics*, Vol. 74, No. 2, October 1996, pp. 319-361.

— (2001a) ‘Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis,’ *Quarterly Journal of Economics 116*, 705-46

— (2001b), ‘Dropout and Enrollment Trends in the Post War Period: What Went Wrong in the 1970s?,’ in *An Economic Analysis of Risky Behavior Among Youth*, ed. Jonathan Gruber (Chicago: University of Chicago Press for NBER)

Chamberlain (1994) ‘Quantile Regression, Censoring, and the Structure of Wage,’ in *Advances in Econometrics*, ed. C. Sims (Cambridge: Cambridge University Press)

Chay, Kenneth and David Lee (2000) ‘Changes in Relative Wages in the 1980s: Returns to Observed and Unobserved Skills and Black-White Wage Differentials,’ *Journal of Econometrics* 99, 1-38

Dehejia, R., and S. Wahba (1999) ‘Causal Effects in Non-Experimental Studies: Re-Evaluating the Evaluation of Training Programs,’ *Journal of the American Statistical Association* 94, 1053-62

Deschênes, Olivier (2001) ‘Unobserved Ability, Comparative Advantage, and the Rising Return to Education in the United States: A Cohort-Based Approach,’ Princeton University Industrial Relations Section Working Paper No. 456

DiNardo, John, Nicole M. Fortin, and Thomas Lemieux (1996) ‘Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach,’ *Econometrica 65*, 1001-46

Donald, Stephen, David Green, and Harry Paarsch (2000) ‘Differences in Wage Distributions between Canada and the United States: An Application of a Flexible Estimator of Distribution Functions in the Presence of Covariates,’ *Review of Economic Studies* 67, 609-33

Fortin, Nicole, and Thomas Lemieux (1998) ‘Rank Regressions, Wage Distributions, and the Gender Gap,’ *Journal of Human Resources* 33, 610-43

Green, David, Barton Hamilton and Harry Paarsch (1997) 'Decomposing Trends in Earnings Inequality: Methods and Interpretation,' University of British Columbia mimeo

Gosling, Amanda, and Thomas Lemieux (2002) 'Labor Market Reforms and Wage Inequality in the United Kingdom and the United States,' forthcoming in *Seeking a Premier League Economy*, ed. R. Blundell, D. Card, and R. Freeman (Chicago: University of Chicago Press for NBER).

Gosling, Amanda, Steve Machin and Costas Meghir (2000) 'The changing distribution of male wages in the UK,' *Review of Economic Studies* 67, 635-86

Heckman, James, Hidehiko Ichimura, and Petra Todd (1997) "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme," *Review of Economic Studies 64*, 605-54

Heckman, James and Solomon Polachek (1974) 'Empirical Evidence on the Functional Form of the Earnings-Schooling Relationship,' *Journal of the American Statistical Association* 69, 350-54

Hirano, Keisuke, Guido Imbens, and Geert Ridder (2000) 'Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score,' NBER technical working paper no. T0251

Hirsch, Barry, and Edward Schumacher (2001) 'Match Bias in Wage Gap Estimates Due to Earnings Imputation,' Trinity University mimeo

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce (1993) 'Wage Inequality and the Rise in Returns to Skill,' *Journal of Political Economy* 101, 410-42

Karoly, Lynn (1993) 'The Trend in Inequality Among Families, Individuals, and Workers in the United States: A Twenty-Five Year Perspective,' in *Uneven Tides: Rising Inequality in America*, ed. S. Danziger and P. Gottschalk (New York: Russell Sage Foundation)

Katz, Lawrence, and David Autor (2000) 'Changes in the Wage Structure and Earnings Inequality,' in *Handbook of Labor Economics*, ed. O. Ashenfelter and D. Card (Amsterdam: Elsevier Science)

Katz, Lawrence, and Kevin Murphy (1992) 'Changes in Relative Wages, 1963-1987: Supply and Demand Factors,' *Quarterly Journal of Economics 107,* 35-78

Koenker, Roger and G. Bassett (1978) 'Regression Quantiles,' *Econometrica* 46, 33-50

Lee, David (1999) 'Wage Inequality in the United States during the 1980s: Rising Dispersion or Falling Minimum Wage?,' *Quarterly Journal of Economics* 114, 977-1023

Machado, José and José Mata (2002) 'Counterfactual Decompositions of Changes in Wage Distributions using Quantile Regression,' Universidade Nova de Lisboa mimeo

Mincer, Jacob (1974) *Schooling, Experience, and Earnings* (New York: NBER)

Mincer, Jacob (1997) 'Changes in Wage Inequality, 1970-1990,' *Research in Labor Economics 16*, 1-18.

Oaxaca, Ronald (1973) 'Male-Female Wage Differentials in Urban Labor Markets,' *International Economic Review* 14, 693-709

Quah, Danny (1996) 'Twin-Peaks: Growth and Convergence in Models of Distributional Dynamics,' *Economic Journal* 106, 1045-55

Rosen, Sherwin (1977) 'Human Capital: A Survey of Empirical Research,' in *Research in Labor Economics, volume 1*, ed. R. Ehrenberg (Greenwich Connecticut, JAI Press)

Teulings, Coen (2000) 'Aggregation Bias in Elasticities of Substitution and the Minimum Wage Paradox," *International Economic Review* 41, 359-98

— (2002) 'The Contribution of Minimum Wages to Increasing Wage Inequality,' forthcoming in the *Economic Journal*

Appendix 1

In the May 1973 CPS and the 1979, 1989, and 1999 outgoing rotation group files of the CPS, all workers paid by the hour are asked to report their usual hourly wage on their main job. All wage and salary workers are also asked to provide their usual weekly earnings and usual weekly hours of work on their main job.[31] The measure of hourly wages I use is the hourly wage for workers paid by the hour and average hourly earnings (weekly earnings divided by hours) for others. Since self-employed workers are not asked about their earnings they are not part of these wage samples. I use a broad sample of all workers age 16 to 64 but weight observations by hours of work to avoid putting too much weight on workers marginally attached to the labor market like full-time students. Following other studies, I trim extreme values below 1 dollar or above 100 dollars (deflated to 1979 dollars using the CPI-U) and multiply the wage of workers whose earnings are top-coded by an adjustment factor of 1.4. In the May 1973, the wage is missing for workers who do not report their earnings. In 1979 and 1999, missing earnings are imputed by the U.S. Census Bureau and allocation flags indicate which observations have been imputed. Missing earnings are also imputed in 1989 but accurate allocation flags are not available (see Hirsch and Schumacher, 2001, for a detailed discussion of allocation issues in the CPS). For the sake of consistency, allocated earnings are used to look at changes in the wage distribution between 1979 and 1989 and 1989 and 1999, while allocated earnings are excluded from 1979 to look at changes between 1973 and 1979. Changes from 1973 to 1999 are computed as the sum of the changes for 1973-79, 1979-89, and 1989-99. The six education categories I use in the regression and logit models are 0-8, 9-11, 12, 13-15, 16, and 17+.

Table 1: Summary of the procedure of Section 3 for generating counterfactual distributions

Variable:	Weight:	Resulting distribution:
y_{it}	ω_{it}	Distribution at period t
y_{it}^{a}	ω_{it}	Period t distribution with b of period s
y_{it}	ω_{it}^{a}	Period t distribution with distribution of covariates of period s
y_{it}^{a}	ω_{it}^{a}	Period t distribution with b and distribution of covariates of period s
y_{is}	ω_{is}	Distribution at period s
y_{is}^{a}	ω_{is}	Period s distribution with b of period s
y_{is}	ω_{is}^{a}	Period s distribution with distribution of covariates of period t
y_{is}^{a}	ω_{is}^{a}	Period s distribution with b and distribution of covariates of period t

Note: $y_{it}^{a} = x_{it}b_s + u_{it}$, $y_{is}^{a} = x_{is}b_t + u_{is}$, $\omega_{it}^{a} = \psi_i\, \omega_{it}$, and $\omega_{is}^{a} = (1/\psi_i)\omega_{it}$, where ψ_i is the reweighting factor.

Table 2: Log Wage Distribution of Women in Alberta and British Columbia

	Mean	Variance			Wage gap by percentiles		
		Total	xb	Residual	90-10	50-10	90-50
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1. Alberta	2.522	0.218	0.076	0.142	1.253	0.577	0.676
2. Alberta with BC's b	2.627	0.210	0.068	0.142	1.217	0.576	0.641
3. Alberta with BC's b and X	2.654	0.213	0.066	0.147	1.230	0.599	0.631
4. British Columbia (BC)	2.654	0.199	0.066	0.133	1.196	0.643	0.553
5. Alberta-BC Difference	-0.132	0.019	0.010	0.009	0.057	-0.066	0.123
Effect of:							
6. b (Row 1 - row 2)	-0.105	0.008	0.008	---	-0.036	0.001	0.015
7. x (Row 2 - row 3)	-0.027	-0.003	0.002	-0.005	0.013	-0.023	0.010
8. Residual (Row 3 - row 4)	---	0.014	---	0.014	0.034	-0.044	0.076

Note: Computed using data from January to October 2000 Labour Force Survey. Sample includes all female wage and salary workers age 15-64 with hourly wages between 3 and 75 dollars. Workers of all age groups with 8 years of schooling and less are excluded from the sample. Workers age 15-19 with a university degree and workers age 20-24 with a post-graduate university degree are also excluded from the sample. Total number of observations for Alberta and British Columbia are 19319 and 20587, respectively.

Table 3: Change in wage inequality in the United States, 1973-1999

	Men				Women			
	73-79 (1)	79-89 (2)	89-99 (3)	73-99 (4)	73-79 (5)	79-89 (6)	89-99 (7)	73-99 (8)
A. Variance of wages								
Total chg:	-0.002	0.070	0.011	0.079	-0.026	0.086	0.017	0.077
Effect of:								
b	-0.004	0.043	0.013	0.053	-0.016	0.039	0.016	0.039
x	0.009	0.007	0.008	0.024	0.005	0.013	0.007	0.025
p	-0.007	0.021	-0.009	0.006	-0.015	0.033	-0.005	0.014
Unexpl. chg:	0.000	-0.001	-0.002	-0.003	-0.001	0.001	-0.001	0.000
B. Variance of wage residuals								
Total chg:	-0.003	0.030	0.000	0.027	-0.014	0.043	0.002	0.031
Effect of:								
x	0.004	0.009	0.010	0.023	0.002	0.009	0.007	0.018
p	-0.007	0.021	-0.010	0.004	-0.016	0.034	-0.005	0.014
C. 90-10 gap in wages								
Total chg:	0.003	0.169	0.028	0.200	-0.077	0.328	0.030	0.280
Effect of:								
b	-0.008	0.102	0.035	0.129	-0.054	0.126	0.031	0.103
x	0.021	0.015	0.027	0.063	0.018	0.056	0.029	0.104
p	-0.003	0.056	-0.037	0.016	-0.031	0.096	-0.028	0.037
Unexpl. chg:	-0.006	-0.005	0.003	-0.008	-0.011	0.049	-0.002	0.037
D. 90-10 gap in wage residuals								
Total chg:	0.001	0.104	-0.015	0.089	-0.034	0.183	-0.015	0.134
Effect of:								
x	0.003	0.019	0.025	0.047	0.004	0.038	0.025	0.068
p	-0.003	0.084	-0.040	0.042	-0.038	0.145	-0.040	0.066
E. 50-10 gap in wages								
Total chg:	0.019	0.063	-0.044	0.038	-0.093	0.231	-0.039	0.099
Effect of:								
b	-0.006	0.051	-0.018	0.026	-0.024	0.063	-0.001	0.038
x	0.020	0.008	0.040	0.068	0.022	0.059	0.039	0.121
p	-0.006	0.033	-0.035	-0.008	-0.042	0.059	-0.025	-0.008
Unexpl. chg:	0.012	-0.029	-0.031	-0.047	-0.049	0.050	-0.052	-0.051
F. 90-50 gap in wages								
Total chg:	-0.016	0.106	0.072	0.162	0.016	0.096	0.069	0.181
Effects of:								
b	-0.002	0.052	0.053	0.103	-0.030	0.063	0.032	0.065
x	0.001	0.007	-0.012	-0.005	-0.004	-0.003	-0.010	-0.017
p	0.003	0.023	-0.002	0.024	0.012	0.037	-0.003	0.046
Unexpl. chg:	-0.018	0.024	0.034	0.040	0.038	-0.001	0.050	0.088

Note: Sample contains wage and salary workers age 16-64. See Appendix 1 for more details.

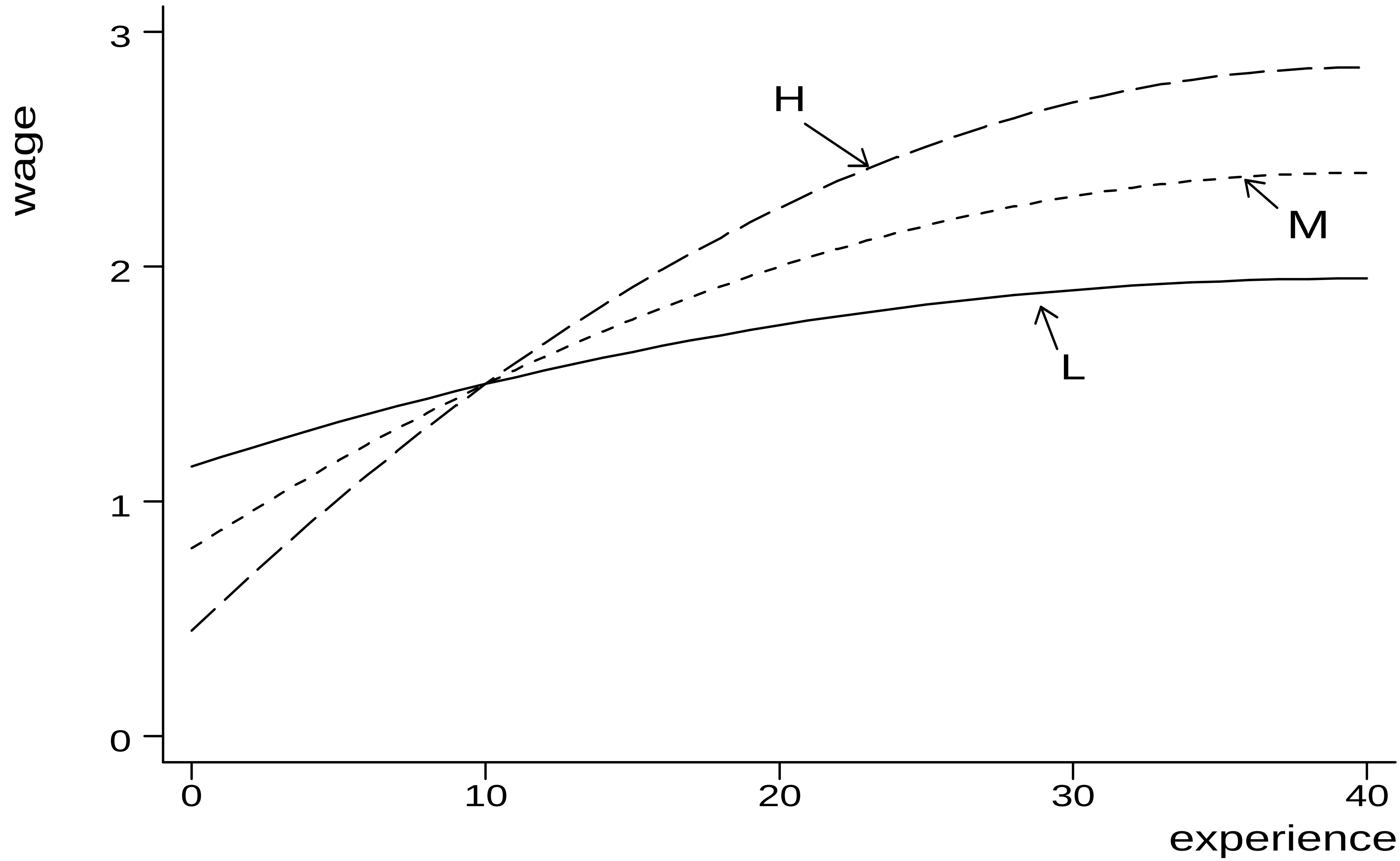

Figure 1: Residual wage dispersion in OJT model

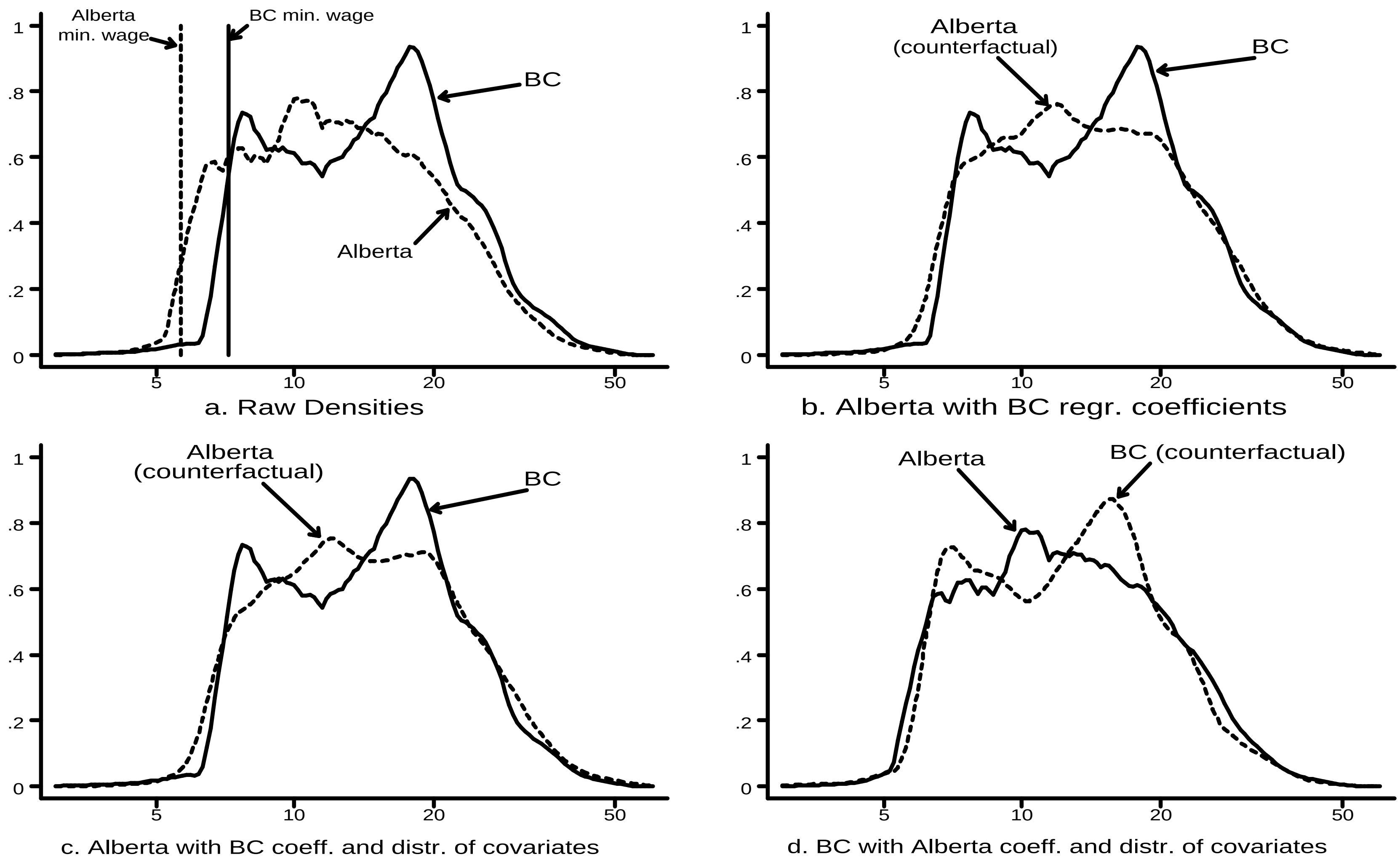

Figure 2: Log wage densities in Alberta and British Columbia, women

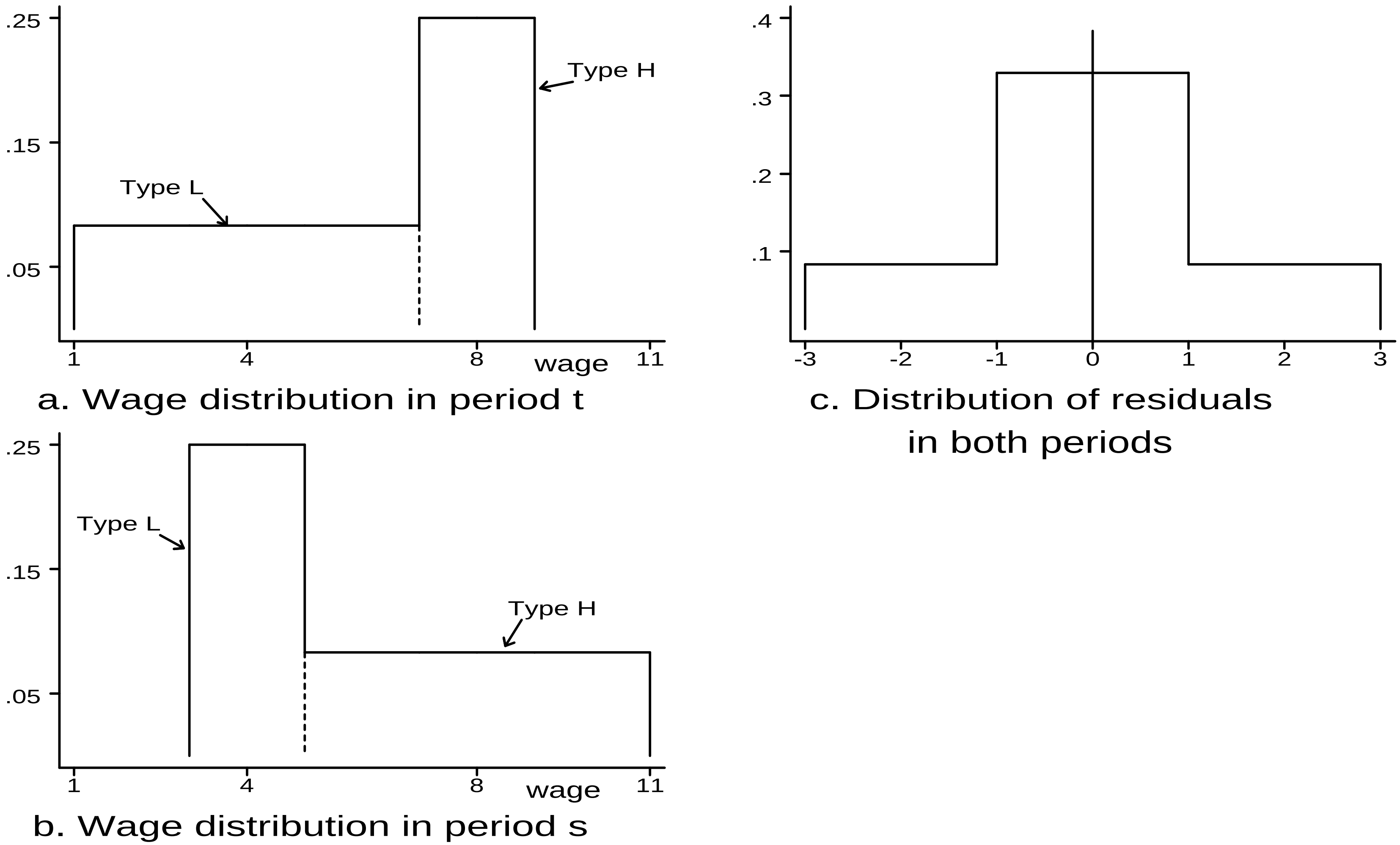

Figure 3: Example with two types of workers

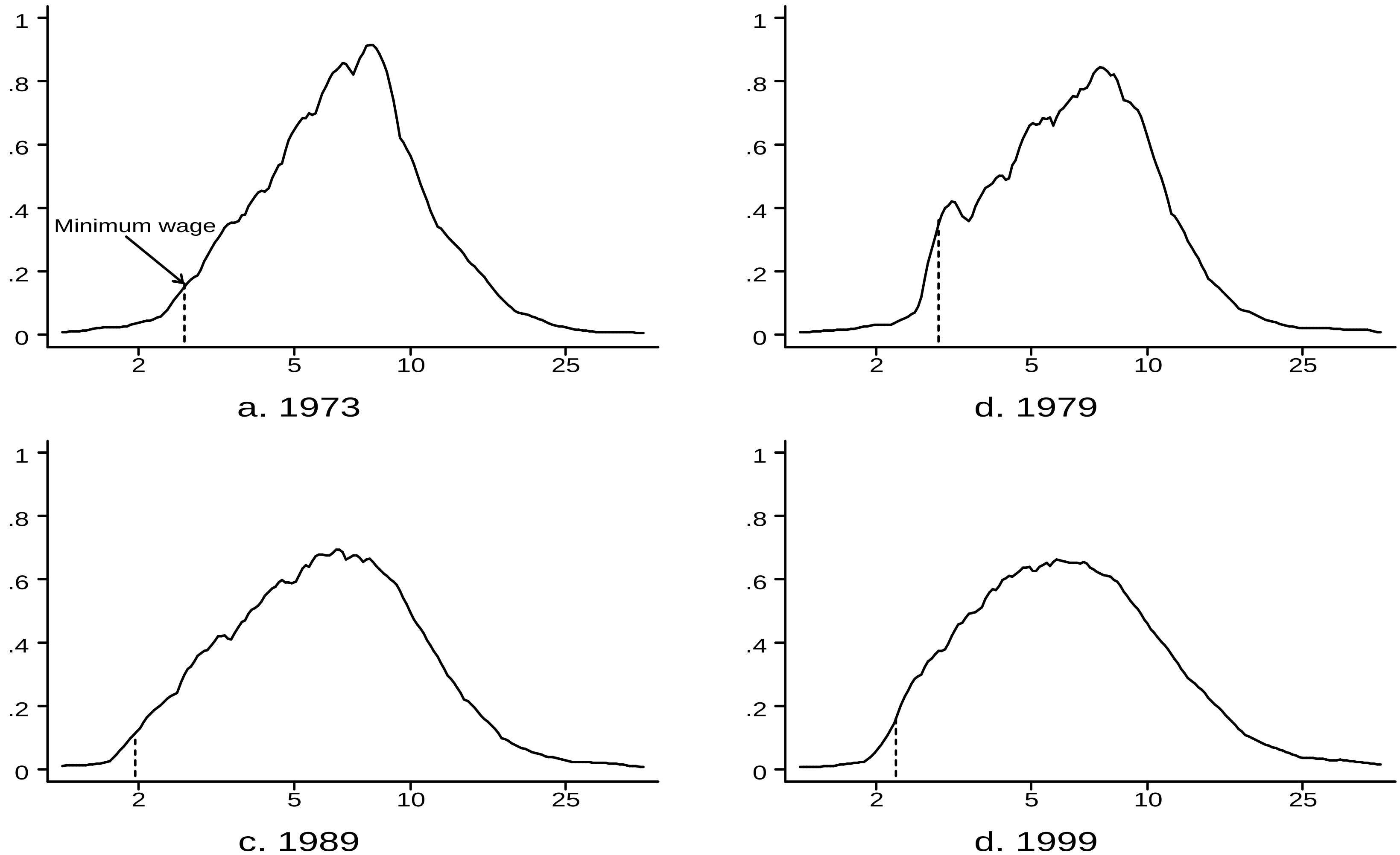

Figure 4: Density of log wages ($1979) for U.S. men

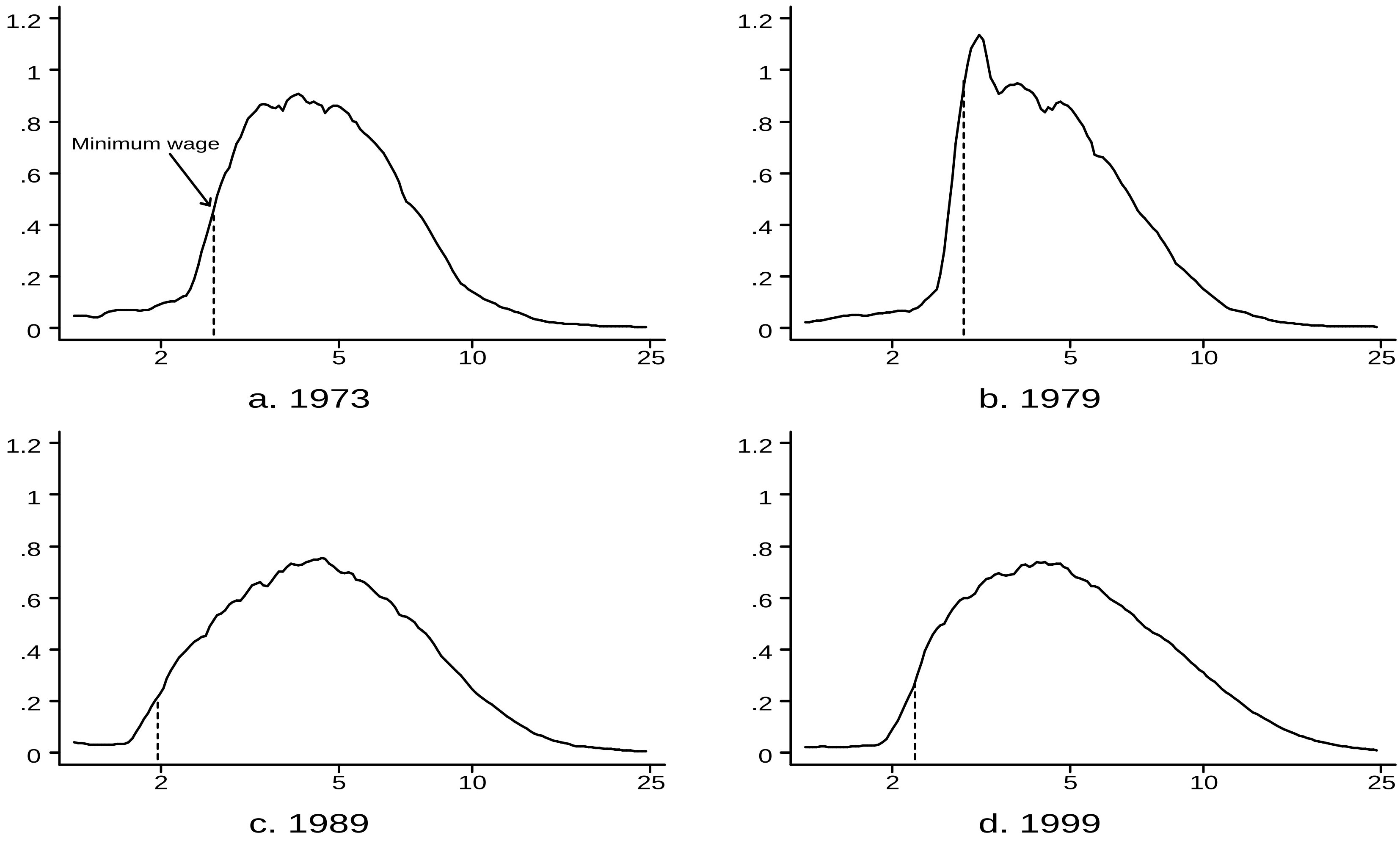

Figure 5: Density of log wages ($1979) for U.S. women

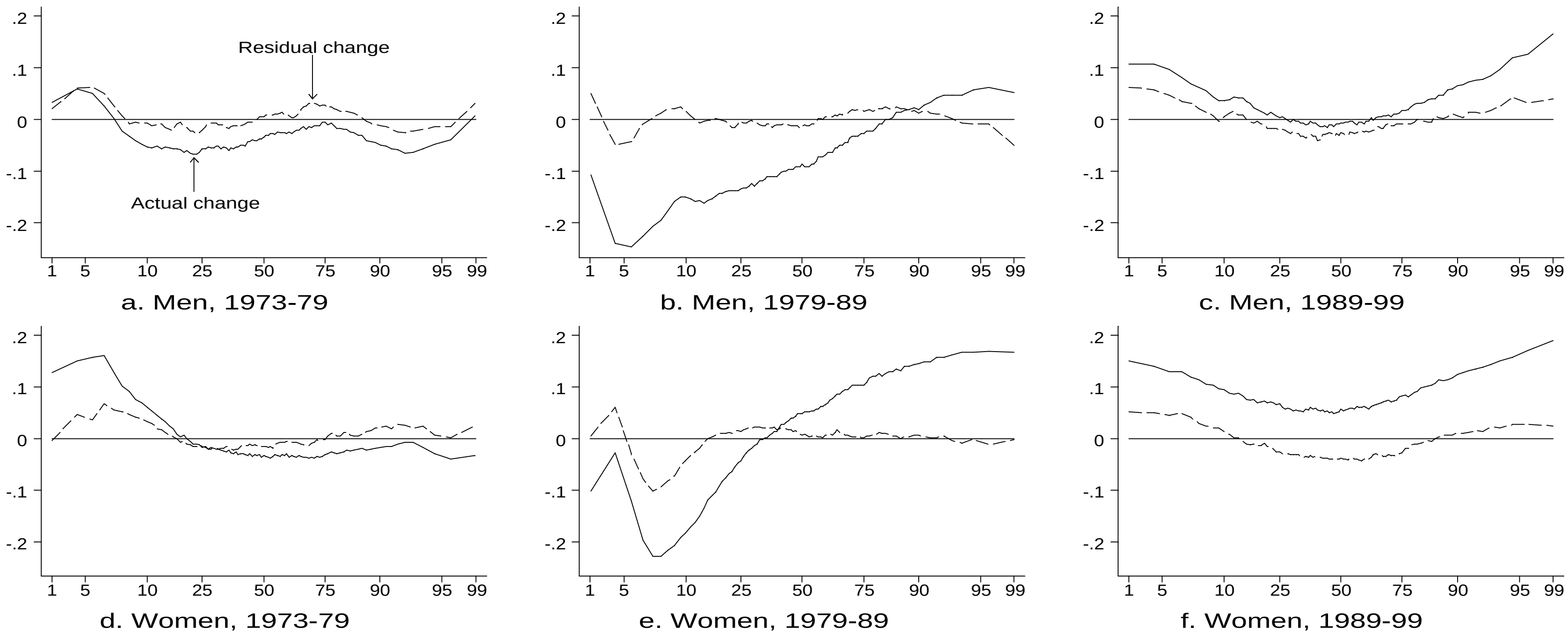

Figure 6: Change in real wages by percentile

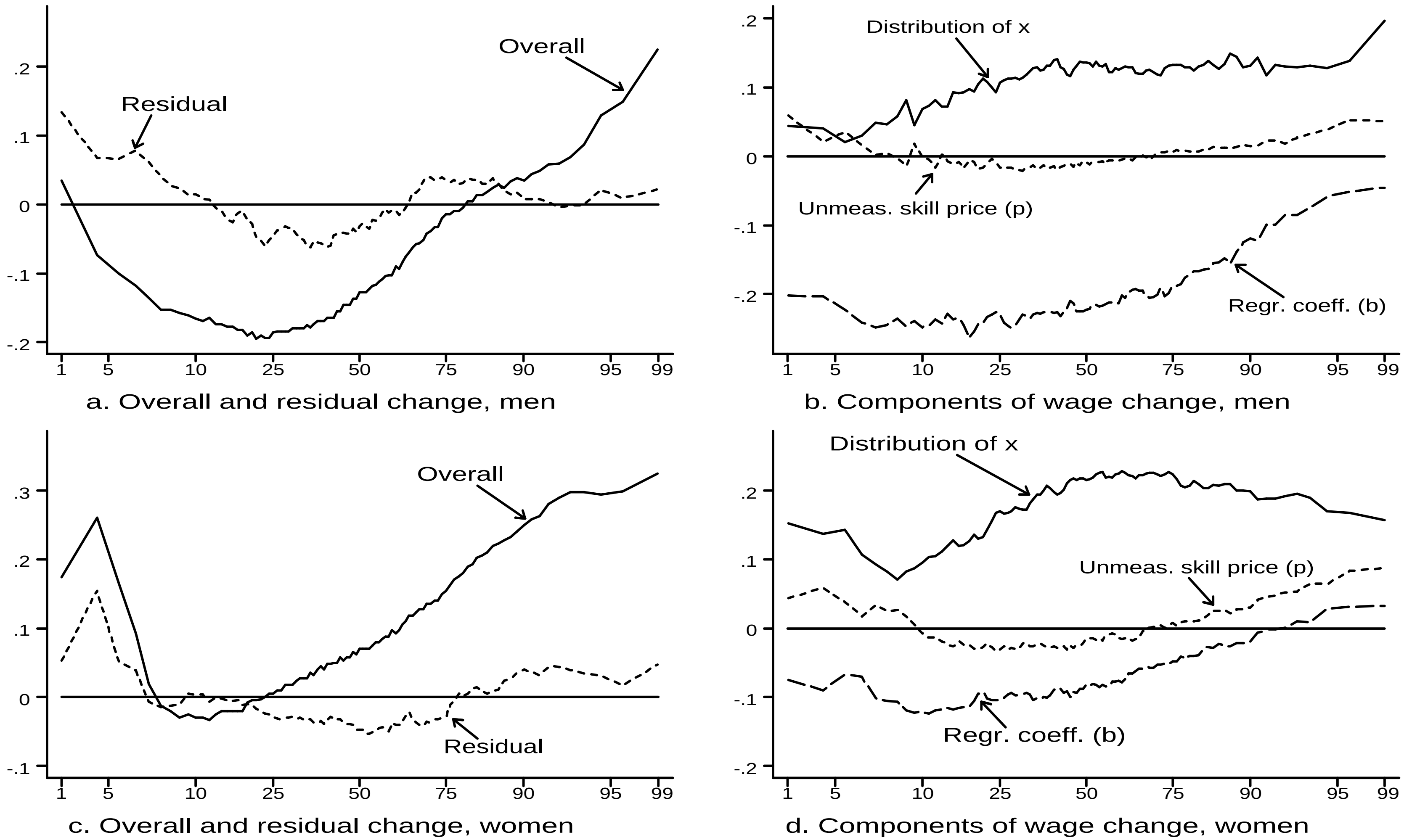

Figure 7: Sources of wage change by percentile, 1973-99

FOOTNOTES

Lead footnote: Innis Lecture delivered at the 2002 Meetings of the Canadian Economic Association in Calgary. I would like to thank Jean-Marie Dufour for inviting me to give this lecture, Paul Beaudry, John DiNardo, David Green, Nicole Fortin and Jennifer Hunt for their helpful comments, David Card for his help with the 1999 CPS data, and SSHRC and NICHD for financial support.

1.In the early studies like Katz and Murphy (1992), skill-biased technological change was viewed as an exogenous factor. More recent papers like Acemoglu (2002) and Beaudry and Green (2002) propose richer models where technological innovation and adoption are endogenous responses to other factors including supply.

2.See Card and DiNardo (2002) and Beaudry and Green (2002) for more recent evidence on these trends.

3.There is a large literature on this starting with Atkinson (1970).

4.Mincer's equation is such a standard tool of empirical labour economics that it is quite difficult to quantify the number of studies that have used it. As a benchmark, the Social Sciences Citation Index reports that between January 1989 and May 2002, over 1,300 studies had cited Mincer (1974)'s book *Schooling, Experience, and Earnings*. Mincer's original study used data from the 1960 U.S. Census. While many studies used more recent data for the United States, many others looked at other developed or developing countries.

5.See also Becker (1975) and Ben-Porath (1967).

6.Both Mincer (1974) and Rosen (1977) make this point. Recent work by Mincer (1997) and Deschênes (2001) shows that the log wage-education relationship has become more convex over the last twenty years.

7.For example, the variance of x_{ijt} is equal to θ_{jt} (1- θ_{jt}). It is easy to show that in this model the between group variance can be written as $\sum_j \theta_{jt} (\bar{y}_{jt} - \bar{y}_t)^2 = b_t' \Omega_{x,t} b_t$.

8. The correction factor P_t / (1-P_t) is of little practical importance since it only changes the reweighting factor in a proportional way and most statistical packages automatically normalize

the sum of weights when computing weighted statistics.

9.Barsky, Bound, Charles and Lupton (2001) suggest using the re-weighting approach as a non-parametric alternative to the usual Oaxaca-Blinder decomposition.

10.Table 1 shows how to perform the decomposition of Section 3 in different orders. This can be used, for example, to compute the decomposition in the same order as JMP who first look at the effect of the regression coefficients, followed by the residuals and the effect of the distribution of covariates. A similar decomposition can be obtained by reweighting the distribution of period s using ω_{is}^{a} (see Table 1) instead of reweighting the counterfactual distribution y_{it}^{a} with ω_{is}^{a}.

11.It is straightforward to look at the impact of changes in a single regression coefficient instead of the whole vector of coefficients by computing a counterfactual wage y_{it}^{a} in which only this particular coefficient has been switched from its period t value to its period s value. Interestingly, a similar procedure can be used to look at changes in the distribution of a single covariate x_j. Say the logit or probit coefficient for this covariate is c_j. The relevant counterfactual is then obtained by setting all the other coefficients but this one to zero when constructing the reweighting factor (using the logit or probit predicted probabilities).

12.See Heckman, Ichimura and Todd (1997) and Dehejia and Wahba (1999) for a recent application of propensity score methods to training programs.

13.More detailed age categories (2 or 3 years bands) are provided for individuals under the age of 30. This information is not used here.

14. Note that experience rather than age is typically used in the Mincer equation. Since measures of actual experience are rarely available in micro data sets, Mincer suggested using potential experience defined as age - S - 6, where S represents years of schooling. Since potential experience is just a linear function of age and education, age-education and experience-education cells can be used interchangeably.

15.Only a very small fraction of the adult population in these two provinces has the lowest coded value of schooling which is 0 to 8 years. All observations with 0 to 8 years of schooling are removed to ensure there is a reasonably large number of observations in each age-education cells.

For similar reasons, I also remove individuals age 15-19 who report holding a university bachelor's degree (second highest education category) and individuals age 15-24 who report holding a post-graduate degree (highest education category). Finally, observations with hourly wages of less than 3 dollars or more than 75 dollars are removed from the sample.

16.These densities are estimated using a normal kernel with a bandwidth of 0.06. See DiNardo, Fortin, and Lemieux (1996) for more details on kernel density estimation of wage distributions and on the choice of bandwidth.

17.The bi-modal feature of the distribution is robust to the choice of bandwidth. The expression "twin-peaks" comes from the growth literature where it is well documented that the distribution of world income went from unimodal to bimodal over the last decades. See, for example, Quah (1996).

18. Though the regression coefficients are not reported here, a closer examination of the evidence indicates that the main difference between the two provinces is that return to education are substantially lower in British Columbia than in Alberta.

19.A low R-square does not necessarily mean that differences in the shape of the distributions cannot be explained by xb. For example, Beaudry, Collard, and Green (2002) find that secular changes in regression coefficients explain well the emergence of the twin-peaks in the distribution of output-per-capita across countries despite the fact that their R-squares are similar to those in Section 4.

20.See Katz and Autor (2000), Acemoglu (2002) and Card and DiNardo (2002) for recent surveys of the large existing literature.

21. One minor computational issue is that it is not generally possible to exactly match period t and period s residuals at a specific rank η_{it}. One exception is when the two samples are equally weighted and have the same number of observations. A simple solution is to discretize the distribution of residuals in k intervals containing an equal (or close to equal) number of weighted observations. For example, in the next Section I use k=500 and replace the actual residuals by the average residual in each interval. This amounts to approximating F_t by a step function with k steps. Since the skill pricing function is defined as the inverse of the cumulative distribution, this

means that the change in the skill pricing function $p_s p_t^{-1}(.) = F_s^{-1}F_t(.)$, is also being approximated by a step function with k steps.

22.The corresponding figures in nominal dollars are $1.60 in 1973, $2.90 in 1979, $3.35 in 1989, and $5.15 in 1999.

23.The findings reported in Table 3 are very similar to those of Card and Lemieux (1996) and DiNardo, Fortin and Lemieux (1996) for the 1973-79 period. Using the May CPS data for 1973 and 1979, Katz and Autor (2000) find that the variance of log wages increased by 0.01 for men and declined by 0.03 for women during the 1970s, which is quite close to the numbers reported in Table 3. Inequality in average hourly wages can also be computed using data from the March CPS starting in 1975. Both Karoly (1993) and Card and DiNardo (2002) find little change in inequality from 1975 to 1980 using this alternative measure of hourly wage inequality.

24.See, for example, Juhn, Murphy, and Pierce (1993), Katz and Autor (2000), and Acemoglu (2002). Prior to 1973, no direct measures of hourly wages were available in the CPS. The closest proxy to an hourly wage before 1973 is weekly earnings of full-time workers (35 hours or more) working full-year. One drawback of this measure is that it is not representative of the whole workforce. Inequality in the wage measure also depends on the distribution of weekly hours of work among full-time workers. Since direct measures of hourly wages have now been available for 30 years, the benefits of extending the sample prior to 1973 are no longer as clear as they were when wage inequality became a major topic of interest in the mid- to late-1980s.

25.Controlling for experience, the variance of the residuals for workers with a post-graduate degree is about twice as high as for high-school graduates or dropouts in a the sample years considered. The residual variance also grows steadily as a function of experience. It increases by about 50 percent in the first 20 years of experience and keep growing at a slower rate thereafter.

26.Interestingly, these composition effects could have been predicted ex-ante to a large extent from the detailed analysis of residual wage dispersion of Mincer (1974). Using data from the 1959 U.S. Census, Mincer found that residual wage dispersion generally grew with experience (past the overtaking point). Mincer also found that education increased dispersion for experienced workers but decreased dispersion for older workers. As is well known, average experience increased substantially since the early 1980s with the aging of the baby-boom cohort. While average education increased rapidly for more experienced workers since the mid-1970s, it remained relatively constant for younger workers throughout this period (Card and Lemieux, 2001a and 2001b). Combined together, all these factors contributed to increasing the variance of residuals.

27.When the normits are not used, plotting wages by centiles simply traces down the inverse of the cumulative distribution which is highly non-linear (unless the distribution is uniform) and difficult to interpret. When the distribution is log normal, then the function is linear which is very convenient. Since most distributions of log wages look much more like a normal than a uniform distribution, it is more convenient to use normits on the x-axis.

28.See Green, Hamilton and Paarsch (1997) who also consider these two possible versions of the JMP decomposition. They also compare the JMP procedure to the method of Donald, Green and Paarsch (2000) discussed below.

29.See Chamberlain (1994) and Buchinsky (1994) for more details. Koenker and Bassett (1978) suggest a now standard estimation method that involves solving a linear programming problem in cases where the data cannot be divided into cells.

30.In practice, Machado and Mata (2002) use a bootstrap approach to stochastically impute a counterfactual (or simulated wage) to each observation. The idea is to draw a random value of θ for each observation, estimate the corresponding quantile regression (in a different period when performing the counterfactual exercise), and compute the predicted quantile at the actual value of x for this observation. One drawback of this approach is that it requires estimating a large number of quantile regressions, which is computationally involved. Gosling, Machin and Meghir (2000) aggregate directly the conditional quantiles into an unconditional wage distribution.

31.Note that starting in 1994, workers have the option of reporting the wages and earnings on a more convenient basis like monthly or annual earnings. Weekly earnings are then obtained by dividing earnings by the appropriate number of weeks. Starting in 1994 workers can also report variable hours instead of a given number of usual hours of work. In the 1999 CPS, I impute hours for these workers using an algorithm suggested by Anne Polivka of the U.S. Census

Bureau.