

Decomposing Wage Distributions with Self-Selection

by

Alexandru M. Lefter
Brian P. McCall
University of Minnesota

May 2005

Introduction

There is ample empirical evidence that the wage structure in the United States changed markedly during the 1980s and early 1990s (for a comprehensive survey of the literature, see Katz and Autor, 1999). Despite unanimous agreement that an important dimension of this change is related to the sharp increase in wage dispersion for both men and women, there has been some debate in the labor economics field as to the roles played by various supply-side, demand-side, and institutional factors in this rampant expansion of wage inequality (detailed discussions of these factors can be found in Topel, 1997; Johnson, 1997; and Fortin and Lemieux, 1997). The empirical strategies used to assess the relative contributions of these factors are typically based on generalizations of the decomposition methodology developed by Blinder (1973) and Oaxaca (1973). One shortcoming of most studies that investigated the determinants of the changes in wage inequality is that they ignored possible changes in the selection process that led some individuals to become part of the labor force and others to stay out of it. In this paper we propose semi-parametric methods that allow assessing the effect of the changes in these self-selection mechanisms on the changes in the entire distribution of wages.

Preliminaries

Since we will be modeling the wage distribution in terms of its hazard function, it is useful to begin our discussion with a brief overview of the properties of hazard functions. The main objective of this outline is to show that the probability density function of a random variable that meets certain general conditions is nothing but a simple transformation of its hazard function. It is worth noting here that although the hazard function is generally used for response variables that come in the form of a duration, wage, earnings and income variables have similar properties (approximately continuous distributions over positive values), and therefore can be analyzed using all techniques available in the duration analysis literature. The benefits of such an empirical strategy are beginning to be acknowledged (see, for example, Donald, Green, and Paarsch, 2000).

Let $T \geq 0$ represent a random duration variable, and t a particular value of T . The cumulative distribution function (c.d.f.) of T is given by $F(t) = P(T \leq t)$, where $t \geq 0$. The survivor function is defined as $S(t) = 1 - F(t) = P(T > t)$. In other words, $S(t)$ represents the probability that an event has not occurred by time t , or that the individual has “survived” past t . Throughout this section, we assume that T is continuous and denote the probability density function (p.d.f.) of T

by $f(t) = \frac{dF(t)}{dt}$. For $\Delta t > 0$, $P(t \leq T < t + \Delta t | T \geq t)$ is the probability of leaving the initial state in the interval $[t, t + \Delta t)$ given survival until time t . The hazard function for T is defined as:

$$\lambda(t) = \lim_{\Delta t \downarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \quad (1)$$

Thus, the hazard function is the instantaneous rate of leaving per unit of time (the “escape” rate). From equation (1) it follows that, for “small” Δt , $P(t \leq T < t + \Delta t | T \geq t) \approx \lambda(t)\Delta t$. The hazard can

then be used to approximate a conditional probability in much the same way that the height of the p.d.f. of T can be used to approximate an unconditional probability. We can express the hazard function in terms of the p.d.f. and c.d.f. of T very simply. First, note that:

$$P(t \leq T < t + \Delta t | T \geq t) = \frac{P(t \leq T \leq t + \Delta t)}{P(T \geq t)} = \frac{F(t + \Delta t) - F(t)}{1 - F(t)} \quad (2)$$

When the c.d.f. is differentiable, we can take the limit of the right hand side of equation (2), divided by Δt , as Δt approaches zero from above:

$$\lambda(t) = \lim_{\Delta t \downarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} \cdot \frac{1}{1 - F(t)} = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)}$$

Because the derivative of $S(t)$ is $-f(t)$, we have:

$$\lambda(t) = -\frac{d \ln S(t)}{dt} \quad (3)$$

Integrating (3) and using the fact that $F(0) = 0$, we can get the c.d.f. of T in terms of the hazard function:

$$F(t) = 1 - \exp\left[-\int_0^t \lambda(s) ds\right] \quad (4)$$

Straightforward differentiation of (4) gives the p.d.f. of T in terms of the hazard function:

$$f(t) = \lambda(t) \exp\left[-\int_0^t \lambda(s) ds\right] \quad (5)$$

Therefore, any probability can be computed using the hazard function. Equation (5) is particularly important for the empirical methods proposed in this paper since it shows that the probability density function of a positive random variable can be easily recovered from its hazard function. This suggests that once selectivity-corrected estimates of the parameters of a hazard function are obtained, it is possible to look at the effects of self-selection on the entire distribution of the variable of interest.

Methods for Decomposing Wage Distribution Changes

In this section we will use the results from hazard models to develop semi-parametric methods that can be used to decompose changes in wage distributions while accounting for self-selection. We assume that the distribution of wages w depends on a J -dimensional vector of observed regressors \mathbf{x} and an unobserved variable v . Rewriting the conditional distribution of wages in terms of (5), we have:

$$f(w | \mathbf{x}, v) = \lambda(w | \mathbf{x}, v)S(w | \mathbf{x}, v) = \lambda(w | \mathbf{x}, v) \exp\left[-\int_0^w \lambda(s | \mathbf{x}, v) ds\right] \quad (6)$$

We will parameterize the hazard function according to the mixed proportional hazard model with time-constant covariates and time-varying coefficients (for details, see McCall, 1996):

$$\lambda(w | \mathbf{x}, v) = v \exp[\alpha(w) + \boldsymbol{\beta}(w)' \mathbf{x}] \quad (7)$$

where $\alpha(\cdot)$ is a continuous function and $\exp[\alpha(w)]$ is the baseline hazard, $\boldsymbol{\beta}(\cdot)$ is a J -dimensional vector of continuous functions measuring the effects of the observed regressors \mathbf{x} , and v is a nonnegative, unobserved, random variable distributed independently of \mathbf{x} .

Substituting (7) into (6) yields:

$$f(w | \mathbf{x}, v) = v \exp[\alpha(w) + \boldsymbol{\beta}(w)' \mathbf{x}] \exp\left\{-\int_0^w v \exp[\alpha(s) + \boldsymbol{\beta}(s)' \mathbf{x}] ds\right\} \quad (8)$$

To see how the unobserved variable v affects the wage distribution, we can use integration by parts to obtain the following equation:

$$E(w | \mathbf{x}, v) = \int_0^\infty S(w | \mathbf{x}, v) dw \quad (9)$$

Thus,

$$E(w | \mathbf{x}, v) = \int_0^\infty \exp\left\{-\int_0^w v \exp[\alpha(s) + \boldsymbol{\beta}(s)' \mathbf{x}] ds\right\} dw = \int_0^\infty \left\{\exp\left\{-\int_0^w \exp[\alpha(s) + \boldsymbol{\beta}(s)' \mathbf{x}] ds\right\}\right\}^v dw \quad (10)$$

Differentiating (10) with respect to v gives:

$$\begin{aligned} \frac{dE(w | \mathbf{x}, v)}{dv} &= \int_0^\infty \frac{d\left\{\exp\left\{-\int_0^w \exp[\alpha(s) + \boldsymbol{\beta}(s)' \mathbf{x}] ds\right\}\right\}^v}{dv} dw = \\ &= \int_0^\infty \left\{\exp\left\{-\int_0^w \exp[\alpha(s) + \boldsymbol{\beta}(s)' \mathbf{x}] ds\right\}\right\}^v \left\{-\int_0^w \exp[\alpha(s) + \boldsymbol{\beta}(s)' \mathbf{x}] ds\right\} dw \\ &= \int_0^\infty [S(w | \mathbf{x}, v = 1)]^v \ln[S(w | \mathbf{x}, v = 1)] dw \end{aligned} \quad (11)$$

Evaluating (11) at $v = 1$ and using integration by parts yields:

$$\left.\frac{dE(w | \mathbf{x}, v)}{dv}\right|_{v=1} = \int_0^\infty S(w | \mathbf{x}, v = 1) \ln[S(w | \mathbf{x}, v = 1)] dw = -\frac{1}{4}$$

In other words, at $v = 1$, a small change in v reduces the mean wage independently of \mathbf{x} .

As is common in the selection literature, we model an individual's labor force participation decision by an index function c . More specifically, we assume that the wage is observed when $c = 1$ and not observed when $c = 0$, where:

$$P(c = 1 | \mathbf{z}, \eta) = 1 - \exp[-\eta \exp(\boldsymbol{\pi}'\mathbf{z})] \quad (12)$$

and \mathbf{z} is an L -dimensional vector of regressors, $\boldsymbol{\pi}$ is an L -dimensional vector of parameters measuring the effect of the regressors on the probability of working, and η is an unobserved random variable. The possibility of selection bias arises when the unobserved variables η and v are correlated. We denote the joint cumulative distribution function of the unobservables η and v by $G(\eta, v)$.

The log-likelihood function for the resulting selectivity-corrected model can be written as:

$$\begin{aligned} \ln(L) = & \sum_{i=1}^N \ln \int \int_{\eta, v} \left\{ \exp[-\eta \exp(\boldsymbol{\pi}'\mathbf{z}_i)] \right\}^{1-c_i} \times \\ & \times \left\{ \left\{ 1 - \exp[-\eta \exp(\boldsymbol{\pi}'\mathbf{z}_i)] \right\} \left\{ v \exp[\alpha(w) + \boldsymbol{\beta}(w)' \mathbf{x}_i] \exp \left\{ -\int_0^w v \exp[\alpha(s) + \boldsymbol{\beta}(s)' \mathbf{x}_i] ds \right\} \right\}^{c_i} \right\} dG(\eta, v) \end{aligned}$$

Maximizing this likelihood function gives estimates of $\alpha(w)$, $\boldsymbol{\beta}(w)$, $\boldsymbol{\pi}$, and $G(\eta, v)$, which can be used to estimate selectivity-corrected probability density functions. McCall (1996) presented conditions under which the parameters of the mixed proportional hazard model with time-varying coefficients can be identified. The only additional condition required for identifiability when self-selection is incorporated in this model is that the vector \mathbf{z} contains at least one variable that is not included in the vector \mathbf{x} .

To illustrate how the empirical strategy outlined above can be used to decompose changes in wage distributions, consider the following selectivity-corrected probability density functions, where the subscripts 1 and 2 indicate time periods:

$$f_1(w | \mathbf{z}_1, \mathbf{x}_1, c_1 = 1) = \int_v v \exp[\alpha_1(w) + \boldsymbol{\beta}_1(w)' \mathbf{x}_1] \exp \left\{ -\int_0^w v \exp[\alpha_1(s) + \boldsymbol{\beta}_1(s)' \mathbf{x}_1] ds \right\} dG_1(v | \mathbf{z}_1, c_1 = 1)$$

$$f_2(w | \mathbf{z}_2, \mathbf{x}_2, c_2 = 1) = \int_v v \exp[\alpha_2(w) + \boldsymbol{\beta}_2(w)' \mathbf{x}_2] \exp \left\{ -\int_0^w v \exp[\alpha_2(s) + \boldsymbol{\beta}_2(s)' \mathbf{x}_2] ds \right\} dG_2(v | \mathbf{z}_2, c_2 = 1)$$

Denote the joint cumulative distribution function of the observables \mathbf{z} and \mathbf{x} at time 1 and time 2 by $Q_1(\mathbf{z}, \mathbf{x})$ and $Q_2(\mathbf{z}, \mathbf{x})$, respectively. Then, the unconditional distributions of wages are:

$$f_1(w | c_1 = 1) = \int_{\mathbf{z}, \mathbf{x}} \left\{ \int_v v \exp[\alpha_1(w) + \boldsymbol{\beta}_1(w)' \mathbf{x}] \exp \left\{ -\int_0^w v \exp[\alpha_1(s) + \boldsymbol{\beta}_1(s)' \mathbf{x}] ds \right\} dG_1(v | \mathbf{z}, c_1 = 1) \right\} dQ_1(\mathbf{z}, \mathbf{x})$$

$$f_2(w | c_2 = 1) = \int_{\mathbf{z}, \mathbf{x}} \left\{ \int_v \nu \exp[\alpha_2(w) + \boldsymbol{\beta}_2(w)' \mathbf{x}] \exp \left\{ - \int_0^w \nu \exp[\alpha_2(s) + \boldsymbol{\beta}_2(s)' \mathbf{x}] ds \right\} dG_2(\nu | \mathbf{z}, c_2 = 1) \right\} dQ_2(\mathbf{z}, \mathbf{x})$$

The differences between these two unconditional distributions can be investigated using a generalized version of the Blinder/Oaxaca decomposition methodology (for details, see DiNardo, Fortin, and Lemieux, 1996):

$$\begin{aligned} f_2(w | c_2 = 1) - f_1(w | c_1 = 1) &= [f_2(w | c_2 = 1) - f_2^{Q_1}(w | c_2 = 1)] + [f_2^{Q_1}(w | c_2 = 1) - f_2^{Q_1, G_1}(w | c_2 = 1)] \\ &+ [f_2^{Q_1, G_1}(w | c_2 = 1) - f_2^{Q_1, G_1, \beta_1}(w | c_2 = 1)] + [f_2^{Q_1, G_1, \beta_1}(w | c_2 = 1) - f_1(w | c_2 = 1)] \end{aligned}$$

where, for example, $f_2^{Q_1, G_1}(w | c_2 = 1)$ is a counterfactual probability density function showing what the distribution of wage would have looked like had the joint distributions of the observables and unobservables remained unchanged between time 1 and time 2. More specifically:

$$f_2^{Q_1, G_1}(w | c_2 = 1) = \int_{\mathbf{z}, \mathbf{x}} \left\{ \int_v \nu \exp[\alpha_2(w) + \boldsymbol{\beta}_2(w)' \mathbf{x}] \exp \left\{ - \int_0^w \nu \exp[\alpha_2(s) + \boldsymbol{\beta}_2(s)' \mathbf{x}] ds \right\} dG_1(\nu | \mathbf{z}, c_1 = 1) \right\} dQ_1(\mathbf{z}, \mathbf{x})$$

A clear advantage of this semi-parametric decomposition technique is to be found in its focus on the entire density of wages, which allows examining the effects of self-selection at any point in the wage distribution. Its main limitation is that the order of the decomposition can influence the magnitude of effects of interest, although this problem can be addressed by reversing the order of the decomposition. Another limitation is related to the absence of exact methods for computing standard errors, but this can be dealt with using a bootstrap procedure.

Conclusions

The purpose of this paper was to discuss a semi-parametric methodology that can be used to investigate changes in wage distributions that are due to changes in labor force participation decisions over time. This methodology borrows techniques from the duration analysis literature and has the advantage of modeling the effects of self-selection on the entire distribution of wages, and not just the mean or some pre-specified percentiles thereof, as is common in regression analysis or quintile regression frameworks. Although the discussion was limited to differences within the same wage distribution between two point in time, with an application to changes in wage inequality, it is worth mentioning that it can be easily extended to explore differences between two wage distributions at the same point in time, with an application to male-female wage differentials. Furthermore, it can be used to analyze changes over time in the differences between two wage distributions, which would allow examining the effects of the changing patterns of self-selection on the narrowing of the gender gap, another important dimension of the change that characterized the U.S. wage structure during the 1980s and early 1990s (Blau, 1998).

References

- Blau, F. (1998), "Trends in the Well-Being of American Women, 1970-1995," *Journal of Economic Literature*, 36(1), 112-165.
- Blinder, A.S. (1973) "Wage Discrimination: Reduced Form and Structural Estimates," *Journal of Human Resources*, 8, 436-455.
- Dinardo, J., N. Fortin, and T. Lemieux (1996), "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semi-Parametric Approach," *Econometrica*, 64(5), 1001-1044.
- Donald S.J., Green, D.A., and H.J. 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, 609-633.
- Johnson, G. (1997), "Changes in Earnings Inequality: The Role of Demand Shifts," *Journal of Economic Perspectives*, 11(2), 41-54.
- Katz, L. and D. Autor (1999), "Changes in the Wage Structure and Earnings Inequality," in *Handbook of Labor Economics Vol. 3*, O. Ashenfelter and D. Card Eds. (North Holland: Amsterdam), 1463-1555.
- Fortin, N., and T. Lemieux (1997), "Institutional Change and Rising Wage Inequality: Is there a Linkage?," *Journal of Economic Perspectives*, 11(2), 75-96.
- McCall, B.P. (1996), "The Identifiability of the Mixed Proportional Hazards Model with Time-Varying Coefficients," *Econometric Theory*, 12, 733-738.
- Oaxaca, R. (1973), "Male-Female Wage Differentials in Urban Labor Markets," *International Economic Review*, 14, 693-709.
- Topel, R.H. (1997), "Factor Proportions and Relative Wages: The Supply-Side Determinants of Wage Inequality," *Journal of Economic Perspectives*, 11(2), 54-74.