Review of Economic Studies (2000) **67**, 609–633 0034-6527/00/00290609$02.00

Differences in Wage Distributions Between Canada and the United States: An Application of a Flexible Estimator of Distribution Functions in the Presence of Covariates

STEPHEN G. DONALD
University of Texas

DAVID A. GREEN
University of British Columbia

and

HARRY J. PAARSCH
University of Iowa

First version received October 1995; *Final version accepted June* 2000 (*Eds.*)

We construct a tractable, flexible-functional-form estimator of cumulative distribution functions for non-negative random variables which admits large numbers of covariates. The estimator adopts and extends techniques from the spell-duration literature for estimating hazard functions to distribution functions for wages, earnings, and income. We apply these methods to investigate sources of wage inequality for full-time male workers between Canada and the United States, finding that the Canadian wage density has a thinner left tail because low-educated workers have higher pay and a thinner right tail because of a lower proportion of highly-educated workers. Unions appear to play a large role in these outcomes.

1. INTRODUCTION

Canada and the United States, in spite of broadly similar institutions as well as human and physical resources, have very different levels of income inequality. Inequality in the Canadian income distribution has been consistently below that in the American distribution over the last 20 years and did not increase as rapidly in the 1980s. In part, this difference is due to differences in the tax and transfer systems of the two countries. However, Canada also has a more equal distribution of male pre-tax annual earnings and, during the 1980s, experienced a smaller increase in inequality in this distribution than the United States.[1] These facts suggest non-trivial differences in the labour markets of the

1. See Hanratty and Blank (1992) as well as Blackburn and Bloom (1993) on these points.

two countries. Pinpointing the sources of the differences can provide insights into the workings of labour markets in the two countries and the effects of differing labour-market institutions. The substantial similarities between the two countries heighten the usefulness of such a comparative exercise because differences attributed to a factor such as relative patterns of unionization are more likely to be a reliable measure of union impacts than one would obtain from a comparison of less similar countries. In this paper, we investigate the sources of differences in the wage distributions of Canada and the United States by decomposing differences in the entire male wage distributions for the two countries in 1989.

To understand the sources of differences in the wage distributions of the two countries, we require a way of examining how differences in the distributions vary by subgroups in the populations. A common approach to this problem is to evaluate scalar measures of dispersion for various subgroups in the population. As discussed in DiNardo, Fortin, and Lemieux (1996), scalar measures mask details in wage, earnings, and income distributions that are needed to identify the effects of labour-market institutions such as minimum wages. Furthermore, scalar measures place different inherent weights on different parts of the distribution.[2] To avoid this sort of masking and prejudging, one needs techniques for estimating the entire distribution in the presence of covariate effects. Our second goal in this paper is to construct a tractable and flexible estimator of conditional density and cumulative distribution functions for non-negative random variables that admits large numbers of either discrete or continuous covariates.

Our approach to conditional-density estimation is to use a hazard-function based estimator to introduce covariates in a consistent manner, and then to make the necessary transformations to obtain an estimate of the associated conditional density function. Below, we argue that, given the large literature and previously-developed estimation routines for hazard-based duration models, this is likely to be a very tractable and accessible approach for many empirical economists. We also demonstrate that our estimator provides very flexible density estimates, in the sense that it imposes a minimal number of restrictions on the shape of the density for any value of the covariate vector, and that the estimates have the desirable statistical property of integrating to one. Just as important, generating standard errors for these estimates is also very straightforward since they are based on the variance–covariance matrix of the parameter estimates derived from the estimation of the hazard-function specification. These advantages come at the cost of some difficulty in interpreting the estimated parameters of the model but, as we shall see in the case of American and Canadian wages, the estimator is nonetheless a useful tool for examining and decomposing wage, earnings, and income distributions in the presence of covariates. Finally, it is worth noting that, while we emphasize the usefulness of the estimator for examining the conditional densities of wages, earnings, and income, the formulae derived here could be equally well used to derive the conditional density functions and cumulative distribution functions (CDFs), along with associated standard errors, for spell durations such as the duration of unemployment. In this sense, our work can also be viewed as a direct extension of the spell-duration literature.

2. DENSITY ESTIMATION IN THE PRESENCE OF COVARIATES

We require an estimator of the density function that incorporates the effects of covariates across the entire range of the dependent variable. We desire an estimator with three main

2. For example, comparisons of Gini coefficients place weight on differences in the middle part of the distribution, while top/bottom-decile differences place weight on differences in the tails.

properties: first, tractability in that it can be adapted to a large number of alternatives and can be easily understood and implemented by researchers with standard training in applied econometrics; second, flexibility in the way covariates affect the shape and location of the density, so that the estimator imposes a minimal number of restrictions on shape and location; and third, consistency, in that estimated probabilities associated with any subset of the dependent variables range lie in the interval [0,1] and that the overall density-function estimate integrates to one.

Our approach is to translate techniques developed for estimating spell-duration distributions to the estimation of wage, earnings, and income distributions.The main building block of this estimation approach is the hazard function. Consider a non-negative random variable Y with associated probability density function $f(y)$ and CDF $F(y)$. The hazard function $h(y)$ is then defined by the conditional probability

$$h(y) = \frac{f(y)}{[1\text{-}F(y)]} \equiv \frac{f(y)}{S(y)}, \tag{2.1}$$

where $S(y)$ is the survivor function. One key result from the literature on spell-duration estimation is that the conditional nature of $h(y)$ makes it easy to introduce flexible functions of the covariates and to entertain complex shapes for the hazard function, while still meeting the consistency objectives set out above. From (2.1), one can see that the hazard function is simply a transformation of the probability density function. Thus, flexible estimates of the hazard function can be translated into flexible estimates of the density function in a manner discussed in detail below. In fact, our direct interest is not in the hazard-function estimates, but in the density-function and CDF estimates we can derive from them. We make no attempt to interpret the hazard-function estimates themselves, but treat the hazard function effectively as a flexible functional form that allows us to generate tractable, flexible, and consistent estimates of the functions in which we are primarily interested.

We introduce covariates using a proportional-hazards model. Specifically, the hazard function for person i conditional on $\mathbf{x}_i$, a particular realization of the covariate vector $\mathbf{X}_i$, is

$$h(y|\mathbf{x}_i) = \exp(\mathbf{x}_i\alpha)h_0(y), \tag{2.2}$$

where $h_0(y)$ is the baseline hazard function common to all individuals and α is a vector of unknown parameters conformable with $\mathbf{x}_i$. Two important shortcomings of this specification are the restrictions that individuals with very different covariate vectors have hazard functions with the same basic shape and that any particular covariate shifts the entire hazard function up or down relative to the baseline specification—strong restrictions indeed. To weaken these restrictions, we partition the range of Y into P subintervals $\Omega_p = [y_L^p, y_U^p)$, where $\Omega_p \cap \Omega_q = 0$ for all $p \neq q$ with $\bigcup_{p=1}^{P} \Omega_p = [0, \infty)$ and allow the covariate effects to vary over these subintervals. In particular, following Gritz and MaCurdy (1992), we replace $\mathbf{x}_i\alpha$ in (2.2) with

$$g[\mathbf{x}_i(\Omega_p),\beta] = \sum_{p=1}^{P} \mathbf{1}(y \in \Omega_p)\mathbf{x}_i(\Omega_p)\beta^p, \tag{2.3}$$

where $\mathbf{1}(y \in \Omega_p)$ is an indicator function of the event $(y \in \Omega_p)$, equalling one if Y is contained in the set Ω_p and zero otherwise, $\mathbf{x}_i(\Omega_p)$ is a $(1 \times k)$ vector of covariates defined on the set Ω_p, β^p is a $(k \times 1)$ vector of unknown parameters, and β is the vector $(\beta^{1\mathsf{T}}, \beta^{2\mathsf{T}}, \ldots, \beta^{P\mathsf{T}})^{\mathsf{T}}$, a $(K \times 1)$ parameter vector with K equalling $(P \times k)$. Within this specification, covariates can shift the hazard function up over some regions and down over others, providing the possibility of quite different shapes for the hazard function for individuals with different covariate vectors. The covariates are written as a function of Ω_p, the relevant range of Y, to allow for covariates that vary with wages, earnings, or income, such as the tax rate or a dummy variable indicating whether Y equals the value of the relevant minimum wage. This is effectively the same as allowing for covariates that vary with spell duration in the spell-duration literature.

For both tractability and flexibility reasons, we approximate $h_0(y)$ using a step function. In particular, we create a set of dummy variables corresponding to each of the segments $[y_j, y_{j+1})$ for $j = 1, \ldots, J$ where J is finite, and estimate a parameter associated with each of these segments. The properties of a proportional-hazards model with this form of baseline specification are discussed in Meyer (1990). The advantage of this approach is that, with a sufficiently large value for J, it can capture complicated shapes for the hazard and associated density functions, including spikes such as those induced by minimum wages. It also makes the transformation from hazard to density estimates very straightforward. The main disadvantage is that spikes induced by the tendency of individuals to report certain round numbers such as \$10 per hour for wages or \$20,000 per annum for earnings are also evident in the final estimated density figures. If these reporting spikes are not of interest in themselves, then they can be a distraction when examining the overall shape of the density. Once the estimated density is calculated, the importance of these spikes can be reduced using smoothing methods.

Our estimation approach is first to select the J "baseline segments" defined above. We choose quite fine segments (\$0.25 segments for wages in the estimation below) between the 5-th and 95-th percentile of the overall wage distribution and broader segments in the two tails. This is similar to adaptive-kernel estimation techniques that permit variable-window widths to reduce the roughness of estimates in the tails of the density. Next, we choose the P "covariate segments" across which the parameter vectors are allowed to vary. The exact way in which we select J and P is discussed in Subsection 2.1. Given these decisions, we maximize the likelihood function corresponding to data for N individuals. The data are $(\mathbf{X}_i, Y_i)_{i=1}^N$ where Y_i is the wage (or earnings) for individual i and $\mathbf{X}_i$ is a $(1 \times k)$ vector of covariates that may be continuous or discrete.

To construct the likelihood function, note that the probability of observing a wage in the j^*-th baseline segment can be expressed as

$$f(y_{j^*}|\mathbf{x}) = S(y_{j^*-1}|\mathbf{x})h(y_{j^*}|\mathbf{x}).$$

It equals the probability of observing a wage at least as large as the upper limit of the (j^*-1)-st segment, the survivor function value at (j^*-1), times the probability of observing a wage in the j^*-th segment conditional on it being at least as large as the upper limit of the (j^*-1)-st segment, the hazard function value at y_{j^*}. If the wage value is top-coded (*i.e.* if all we know is that the wage value is larger than some specified value y_J), then the probability of observing that top-coded wage is the probability that the wage is greater than or equal to the top-code value, $S(y_J|\mathbf{x})$. If one assumes an underlying continous hazard function with the simple proportional-hazards form, then, following Meyer (1990), one can show that the hazard for a discrete baseline segment is given by

$$h(y_{j^*}|\mathbf{x}) = 1 - \exp\left(-\exp\{\gamma_{j^*} + \textstyle\sum_{p=1}^{P} \mathbf{1}[\Omega_{p(j^*)}]\mathbf{x}\beta^p\}\right),$$

where

$$\gamma_j = \log\left[\int_{\mathscr{J}_j} h_0(u)du\right], \tag{2.4}$$

and $\mathscr{J}_j$ is the set of Y-values corresponding to the j-th baseline segment. As before, Ω_p is the set of Y values corresponding to the p-th covariate segment, but notice that we have replaced the p subscript with $p(j)$. We do this to signify that because the covariate segments consist of sets of baseline segments, as j changes, the value of p may also change.

The corresponding survivor function is given by

$$S(y_{j^*}|x) = \exp(-\textstyle\sum_{j=1}^{j_i^*-1} \exp\{\gamma_j + \sum_{p=1}^{P} \mathbf{1}[\Omega_{p(j)}]\mathbf{x}\beta^p\}),$$

so the logarithm of the likelihood function is then

$$\begin{aligned}\mathscr{L}(\beta, \gamma) = \textstyle\sum_{i=1}^{N} [\mathbf{1}(Y_i < y_J) \log\{1 - \exp[-\exp(\gamma_{j_i^*} + g\{X_i[\Omega_{p(j_i^*)}], \beta\})]\}\\ - \textstyle\sum_{j=1}^{j_i^*-1} \exp(\gamma_j + g\{X_i[\Omega_{p(j)}, \beta]\})],\end{aligned} \tag{2.5}$$

where the dependent variable for individual i falls in the j_i^*-th baseline segment, and Y_i is less than y_J if Y_i is not right censored (*i.e.* not top-coded). The vector γ contains the J baseline parameters, where the notation γ_j denotes the element of γ corresponding to the j-th baseline segment. For consistent estimates, we require that covariate values not change within the baseline segments; they may, however, vary across baseline segments.[3]

For the j-th baseline segment, estimates of the hazard- and survivor-function value can be obtained by replacing the β and γ with their estimated values $\hat{\beta}$ and $\hat{\gamma}$ obtained from maximizing (2.5). An estimate of the density is then

$$\hat{f}(y_j|\mathbf{x}) = \hat{S}(y_j|\mathbf{x}) - \hat{S}(y_{j+1}|\mathbf{x}).$$

We now wish to derive an expression for the variance-covariance matrix of the density estimates. Before doing so, we note that the step-function approach to specifying the baseline hazard function involves estimating a large number of parameters corresponding to the baseline segments. In principle, as the number of observations approaches infinity, both J and K, and hence the number of associated coefficients, could also approach infinity. This introduces standard incidental-parameter problems in deriving the variance-covariance matrix of the estimated parameters. In the following derivations, we follow related work (e.g. Heckman and Singer (1984)) by treating the number of baseline segments as fixed.[4] Of course, the key issue is whether the derived formulae provide a good approximation to the actual variance-covariance matrix in standard-sized samples. Below, in Subsection 2.3, we provide some evidence for this point.

Let δ be the $(K+J)$ vector of unknown parameters $(\beta^{\mathsf{T}},\gamma^{\mathsf{T}})^{\mathsf{T}}$, consisting of the K parameters in β and the J baseline parameters in γ. Denote the true value of δ by δ_0 and let $\hat{\delta}$ be the maximum-likelihood estimator (MLE) that maximizes (2.5). Our first assumption concerns the properties of the MLE $\hat{\delta}$.

Assumption 1. Assume the model is described by (2.5) and the MLE $\hat{\delta}$ has the following properties:

(i) $\hat{\delta} \xrightarrow{\text{P}} \delta_0$;
(ii) $\sqrt{N}(\hat{\delta} - \delta_0) \xrightarrow{\text{P}} \mathscr{N}[\mathbf{0}, \mathscr{V}(\hat{\delta})]$;
(iii) a consistent estimator of $\mathscr{V}(\hat{\delta})$ exists.

3. Note that to implement this estimator we require some right censoring. Thus, we typically artificially censor our data at the 99-th percentile, acting as if for values above the 99-th percentile we only know that they are at or above the 99-th percentile and assign the indicator function $\mathbf{1}(Y_I < y_j)$ to zero.

4. One might justify this approach by noting that a natural lower limit exists to how small baseline segments can become because earnings are not paid in fractions of cents. Thus, an upper bound exists on the number of baseline-segment parameters to be estimated.

This assumption can be verified using the standard conditions (such as those sufficient to verify consistency and asymptotic normality of the MLE) discussed in Kalbfleisch and Prentice (1980, Section 4.6). The form of $\mathscr{V}(\hat{\delta})$is also given in Kalbfleisch and Prentice (1980, Section 4.6), and consistent estimates can be obtained by replacing the unknown parameters appearing in $\mathscr{V}(\hat{\delta})$ with consistent MLEs.

We now turn to deriving the limiting properties of the aforementioned density estimator. Toward this end, we first define some notation. Let

$$\Delta_j = \frac{\partial}{\partial \delta} S(y_j|\mathbf{x}),$$

which is a $(K+J)$ column vector with elements

$$\frac{\partial}{\partial \beta^k} S(y_j|\mathbf{x}) = -S(y_j|\mathbf{x}) \textstyle\sum_{i=1}^{j-1} \mathbf{1}[\Omega_{p(i)}] \exp(\gamma_i) \exp\{\sum_{p=1}^{P} \mathbf{1}[\Omega_{p(i)}]\mathbf{x}\beta^p\}\mathbf{x}^{\mathsf{T}},$$

and

$$\frac{\partial}{\partial \gamma_l} S(y_j|\mathbf{x}) = \begin{cases} -S(y_j|\mathbf{x})h(y_l|\mathbf{x}) & \text{for } l<j; \\ 0 & \text{otherwise.} \end{cases}$$

Also, let Θ_j denote $(\Delta_j - \Delta_{j+1})$. The asymptotic properties of the estimators for the survivor and density functions are given as follows:

Lemma 1. *Given Assumption* 1,

$$\hat{S}(y_j|\mathbf{x}) \xrightarrow{\mathrm{p}} S(y_j|\mathbf{x}),$$

$$\sqrt{N}[\hat{S}(y_j|\mathbf{x}) - S(y_j|\mathbf{x})] \xrightarrow{\mathrm{d}} \mathscr{N}[\mathbf{0}, \Delta_j^{\mathsf{T}}\mathscr{V}(\hat{\delta})\Delta_j],$$

$$\hat{f}(y_j|\mathbf{x}) \xrightarrow{\mathrm{p}} f(y_j|\mathbf{x}),$$

$$\sqrt{N}[\hat{f}(y_j|\mathbf{x}) - f(y_j|\mathbf{x})] \xrightarrow{\mathrm{d}} \mathscr{N}[\mathbf{0}, \Theta_j^{\mathsf{T}}\mathscr{V}(\hat{\delta})\Theta_j].$$

The proof of this lemma involves straightforward applications of the continuous mapping theorem and the delta method, and follows given Assumption 1(i) and the nature of the dependence of $\hat{S}(y_j|\mathbf{x})$ on $\hat{\delta}$.

2.1. *Discussion of the estimator's general features*

To understand better the nature of the density estimator, consider an example with no covariates. In this case, the estimator is just the Kaplan–Meier estimator of the hazard function at each baseline segment. For a given segment j, the Kaplan–Meier estimate equals the number of observations with wage values in the baseline segment range d_j (called "failures" in the duration literature), divided by r_j, the number of observations with wage values in the j-th segment range or in segments with higher values (called the "risk set" in the duration literature). Using the estimated hazard function and the formulae given above, one can show that the estimated mass in the j-th segment equals d_j divided by the total sample size N. This equals the sample histogram estimate for the j-th segment. In the presence of covariates, we could estimate a histogram for every available value of the covariate vector. With complete flexibility (*i.e.* with J equalling P and each possible combination of covariate values represented by a dummy variable), implementing our estimator is the same as estimating a separate histogram for each possible combination of covariate values. We implement a restricted version in which the impact of covariates are assumed to follow a particular (flexible) functional form. Thus, our estimator can be seen as a transformation of the sample histogram where the transformation is used to ensure that covariate effects are introduced in a consistent manner.

The fact that our estimator is strongly related to the histogram estimator is useful in considering both its strengths and its weaknesses. Like the histogram, this estimator can approximate any density-function shape and has no difficulty incorporating outliers or top-coding; it goes beyond the histogram by incorporating covariate effects in a flexible way. At the same time, it is quite tractable. In our experience, convergence for this likelihood function can be obtained easily and quickly.[5] The transformations from $\hat{\delta}$ to estimates of the hazard function and then to the density functions described above are straightforward.

The estimator also shares some of the disadvantages of histograms. One difficulty in implementing these types of estimators is that the choice of bin size (or baseline segment size, in our case) is somewhat subjective. This problem is amplified in our context by the fact that we estimate parameters associated with hazard functions, but are ultimately interested in the density function. One possible response to this problem is to use the formulae for optimal bin size developed in the histogram literature; *e.g.*, see Scott (1979).We do not adopt this approach because we prefer to use varying segment widths rather than the constant widths assumed in the optimal-bin-width literature and because the introduction of covariates complicates choosing a segment size based on the range of the unconditional distribution of the dependent variable. Deriving a formula for optimal segment widths is beyond the scope of this paper, and the large numbers of observations and parameters we use make techniques of cross-validation computationally prohibitive at present. Instead, based on substantial Monte Carlo evidence, some of which is reported below, and hands-on experience in applying the estimator to real data, we have developed a procedure that provides a rough rule-of-thumb for selecting segment widths as follows: first, regress the dependent variable, in our case the wage, on the full set of covariates using least squares; second, recover the residuals from the first regression and regress the squared residuals on the covariate vector; and third, use the estimates from the second regression to form estimates of the conditional variance for a set of 10 to 20 common covariate vector values. Apply a middle value from the range of calculated conditional variances to Scott's formula for the optimal bin width with Gaussian data

$$h_{\text{opt}} = 3{\cdot}49 s N^{-1/3}$$

where s equals the square root of the chosen middle conditional variance and N is the total sample size. Employ this bin width between the 5-th and 95-th percentiles of the unconditional wage distribution, but use wider segments in the tails. This approach sets segment widths which are narrow enough to fit even relatively narrow conditonal density functions and admits baseline segment parameters which are estimated using all of the observations.

A second potential difficulty common to both histograms and our estimator is that resulting density estimates can be distractingly "spikey." For our estimator, this is a direct

5. The exception to this claim is that finding useful starting values for the parameters estimates sometimes takes effort. A quick way to get starting values is to transform Kaplan-Meier estimates for the baseline segment values and set all parameters for the covariates to zero.

result of using a discrete baseline-hazard function and is another way in which our estimator is similar to the sample histogram. One response is to use standard techniques to smooth the estimated densities after they are calculated. Of course, the extent to which one pursues such smoothing depends on the question being investigated. For example, in work where minimum-wage effects are of interest, the researcher is most interested in the location and the size of spikes in the wage distribution, so smoothing is unwarranted.

Finally, we note that the estimates of the hazard function for wages, earnings, or income are difficult to interpret. For weekly unemployment-spell data, the hazard function at week τ is the proportion of spells "at risk" to end in week τ (*i.e.* that have not ended or been censored before τ) that actually end in week τ. In the earnings context, it is unclear whether a \$50,000-a-year job was "at risk" of being a \$10,000-a-year job. Thus, some standard interpretations from the duration literature do not carry over to this context. However, the conditional probability given in (2.1) is still defined and maintains its attractive properties as a basis for flexible density estimation in the presence of covariates. In particular, software currently used to estimate duration models and the estimation experience developed with such models can be directly translated to this context. Thus, the estimator represents a trade-off: interpretability of parameters for tractability and flexibility. The key disadvantage of poor interpretability of parameters is that those parameters, and potential restrictions on them, are difficult to relate to economic theory. Nonetheless, the estimator is useful for uncovering patterns in the data, as we demonstrate in the empirical investigation below.

2.2. *Comparisons to alternative estimators*

Of course, ours is not the first attempt to incorporate covariates in ways that affect more than the first moment of a distribution. Approaches to this problem vary from fully parametric, where a specific functional form is assumed for the density of interest and the associated parameters are made to be functions of covariates, to fully non-parametric, where the data are divided into subsamples corresponding to all possible combinations of covariate values and a density estimate is generated for each subsample using a non-parametric approach such as kernel-density estimation. The parametric approach is tractable but risks distorting the information in the data by forcing them into an inappropriate functional form. The non-parametric approach is flexible, but suffers from the "curse of dimensionality" because as new covariate combinations are introduced, cell sizes quickly become too small to provide a source of useful inference. It is also difficult to introduce continuous covariates. For these reasons, researchers often turn to semi-parametric and flexible-function-form approaches, trading some restrictions for ease of implementation.

A feasible alternative is quantile regression; *e.g.* see Buchinsky (1994). This approach admits the estimation of the impact of covariates on any quantile of the distribution and, given recent advances in software, can be implemented relatively easily, even with continuous covariates. The key difficulty is a potential lack of consistency: the predicted value of the p-th quantile for a given covariate vector could be larger than the predicted $(p+1)$-th quantile value for the same covariate vector. This could be remedied by imposing appropriate restrictions in the estimation, but then standard software packages could no longer be used. We suggest that our approach, based on hazard function estimation, is more familiar and thus more accessible to most empirical researchers.

One could also form a flexible estimate of the CDF using an "ordered-probit"-type specification. This would also require imposing restrictions to ensure one always generates positive probabilities, especially with small categories and covariate effects varying over the earnings range. Fortin and Lemieux (1995) have used such an estimator.

Another interesting approach is the weighted-kernel estimator introduced by DiNardo, Fortin, and Lemieux (1996), hereafter DFL. The DFL approach focuses on decomposing differences in the densities for two populations according to the effects of specific covariates. To do this, they weight the kernel function for a particular individual by the relative probability that a covariate of interest takes a specific value in the two populations, conditional on the values of other covariates. Thus, in examining the impact of declines in unionization on changes in the American wage distribution between 1979 and 1987, they create a ratio of the probabilities that an individual is unionized in 1979 and in 1987, given the individual's covariate vector. This relative probability is then multiplied by the individual's kernel-function values at each wage rate to create an estimate of what the 1987 union-wage distribution would have been had the 1979 unionization pattern been maintained.

DFL argue convincingly for the need to examine the entire distribution when considering explanations for changes in wage dispersion over time or across countries. We build on their work; our approach has three advantages over the DFL approach. First, the DFL technique is primarily a way of decomposing differences in distributions; it does not provide a method of graphing density functions or CDFs for specific covariate vectors. In contrast, our approach allows one both to graph such functions and to carry out the same kind of decompositions as in DFL. Second, our approach is more tractable, using techniques more familiar to most researchers. Third, DFL do not provide standard errors for their density estimates, while in our approach standard errors are straightforward to calculate.

2.3. *Some Monte Carlo results on the performance of the hazard-based estimator*

We now turn to implementing the hazard-based estimator using artificially-generated data. We do this in part to provide evidence on how well the estimator fits known density functions and in part to investigate the properties of the estimator as we vary its main estimation choice parameters, the number of baseline segments and the number of covariate segments. We generate data using a mixture of log-normal distributions as follows: first,we draw 1000 pseudo-random observations from a standard-normal distribution. Next, we draw 1000 pseudo-random observations from a uniform [0,1] distribution. Thus, we have 1000 observations each with a standard-normal-variate value for the *i*-th observation, call it U_i, and a uniform-variate value for observation *i*, call it X_i. Without loss of generality, we use the first 500 observations to create the value of the dependent variable according to

$$Y_i = \exp(0{\cdot}5 + 0{\cdot}5U_i),$$

while for the second 500 observations we create the value according to

$$Y_i = \exp(0{\cdot}5 + 0{\cdot}5X_i + 0{\cdot}5U_i).$$

Together, the 1000 observations on Y constitute a random variable which is distributed as a fifty-fifty mixture of log-normal distributions. The advantage of using a mixture is that it generates a random variable with a distribution the shape and location of which changes considerably for different values of X. For values of X near zero, the conditional distribution is very similar to the log-normal distribution. As X increases, the conditional distribution becomes increasingly bimodal. Fitting conditional distributions that vary this much with the covariate value is a challenge for our and any other estimator.

Our approach to evaluating the hazard-based estimator is as follows. We first generated 100 samples of 1000 observations each in the manner described above. We then selected a number of baseline segments J and a number of covariate segments P. We maximized the logarithm of the likelihood function given in (2.5) for each of the 100 samples separately and recovered the estimated parameter vector $\hat{\delta}$ in each case. We then set the covariate value to 0·7, a value that generates a visibly bimodal density function, and created an estimate of the conditional density using the formulae presented above based on each of the one hundred $\hat{\delta}$ vectors. We established a uniform grid with 50 intervals running from the 1-st percentile of the unconditional distribution of Y (approximately 1·35) to the 99-th percentile (approximately 2·80). For each of the 100 fitted density functions, we took the difference between the fitted and true conditional densities at the right endpoint of each of the 50 intervals just mentioned. Using these, we constructed the mean-squared-errors for each of the 100 fitted densities. We then averaged the mean-squared-errors over the samples to obtain an overall measure of closeness of fit corresponding to the chosen values of J and P. We repeated this exercise for a range of values for J and P. The resulting root-mean-squared-error values are reported in Table 1.[6]

TABLE 1

Root-mean-squared-error values
Monte Carlo experiments of the hazard-based estimator

	$P = 1$	$P = 5$	$P = 10$	$P = 20$
$J = 10$	0·0587	0·0499	0·0499	—
$J = 20$	0·0471	0·0310	0·0324	0·0328
$J = 25$	0·0456	0·0302	0·0305	0·0324
$J = 30$	0·0454	0·0315	0·0312	0·0329
$J = 35$	0·0458	0·0322	0·0314	0·0331
$J = 40$	0·0462	0·0323	0·0317	0·0334

The results in Table 1 indicate that, for any number of covariate segments, increasing the number of baseline segments improves the fit of the estimated density up to a point; past that point additional baseline segments erode the fit. This is the same trade-off witnessed in choosing bin widths for histograms: with a few, broad segments it is impossible to capture nuances in the shape of the density and the fit is poor, but with too many, narrow segments the number of observations used to identify each baseline segment parameter is small and excessive variability results. Increasing the number of covariate segments (moving across rows in the table) leads to improvements at first, particularly in moving from a P of 1 to a P of 5; however, adding too many covariate segments worsens the fit. The fact that the root-mean-squared-error value changes little between a P of 5 and a P of 20 suggests that the penalty for using too many covariate segments is not large. The very poor relative fit of the cases with one covariate segment show that it is indeed necessary to use something more complicated than the simple proportional-hazards model

6. In each case, we allowed one baseline segment in the left tail of the distribution covering the range of Y up to 1·4 and baseline segments with left endpoints 2·65, 2·73, 2·8 and 2·9 in the right end of the distribution. The baseline segments between 1·4 and 2·65 consist of (J-5) even-sized intervals. The right endpoints of the P covariate segments are placed in each case at approximately P even quantile points from the unconditional distribution of Y. Thus, when P equalled 10, we divided the covariate segments at the decile points. In each case, we estimated using both the generated X variable and that variable squared. We have also evaluated the root-mean-squared-errors at different values of the X vector with very similar results to those presented in Table 1.

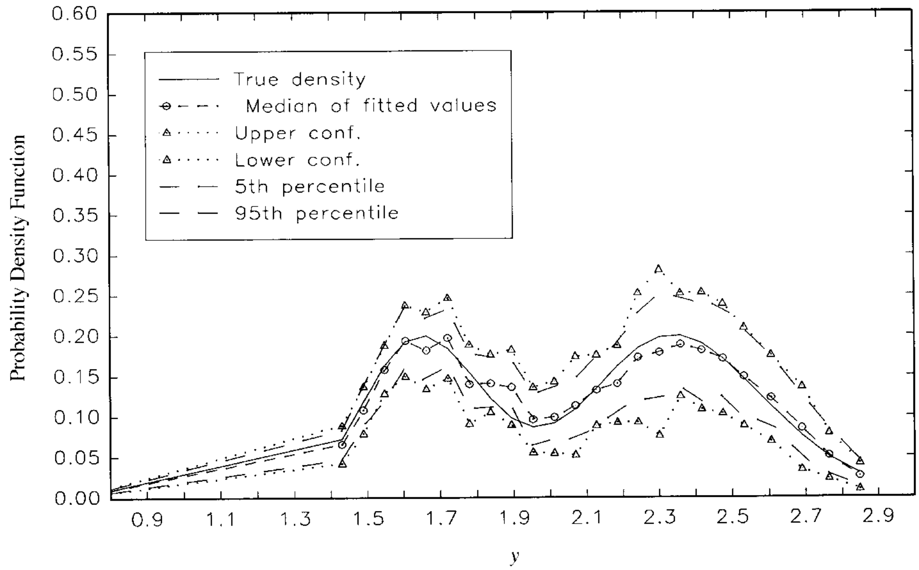

FIGURE 1

Plots of mixture log-normal and fitted densities: 25 baseline segments, 5 covariate segments, $X = 0{\cdot}7$

when introducing covariates. The J-of-20, P-of-5 specification provides the minimum value of the average root-mean-squared-error, though a range of similar values for J and P appear to provide comparable fits.[7]

Visual depiction of the fit of the estimated densities is given in Figures 1, 2, and 3. In Figure 1, we plot the true density when X is 0·7, and a set of fitted densities and associated standard errors based on the estimations with J of 25 and P of 5, the preferred specification based on Table 1. The line labelled median is the median of the 100 fitted density values at each baseline segment. Similarly, the 5-th- and 95th-percentile lines correspond to the 5-th and 95-th percentiles of the fitted-density values at each segment. We also plot an estimate of the 95% confidence band around the median fitted line. The confidence band equals the median fitted value at each baseline segment plus or minus 1·96 times an estimate of the standard error at that segment.[8] In Figure 2, we recreate Figure 1 for an X of 0·2.

Both Figures 1 and 2 indicate that the the hazard-based estimator fits the true density very well. This is noteworthy since we have been able to fit very different shapes, associated with different values of the covariate, from the same estimations. It is also clear from the figures that the estimated standard errors calculated from our formulae are close approximations to the 5-th and 95-th percentiles of the distribution of the fitted density values. Thus, both the density estimates and the estimated standard errors perform well in this example. In contrast, in Figure 3 we recreate Figure 1 based on estimates of a J of

7. Using the rough rule-of-thumb methodology described in Subsection 21, the estimated conditional standard deviation of Y conditional on X being 0·5 is 0·195, which implies an optimal segment width of approximately 0·06 from Scott's formula. Even division of the range between the 5-th and 95-th percentiles of the unconditional distribution of Y for the J-of-25, P-of-5 case implies segments with widths of 0·055.

8. To construct the standard errors, we calculated the standard errors at each baseline segment for each of the 100 density-function estimates using the formulae presented in the text. The standard error for baseline segment j used in the figure is the standard error for segment j associated with the fitted density chosen as the median at that segment.

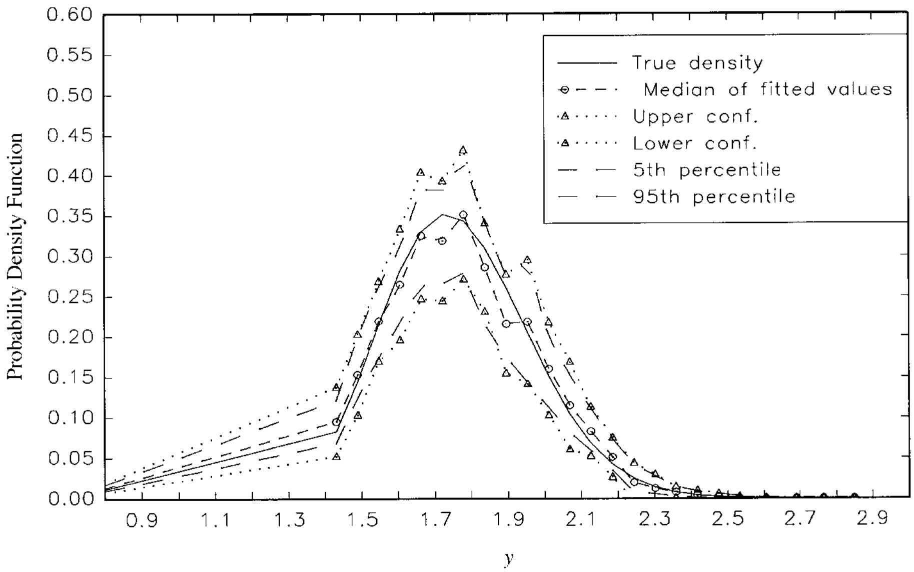

FIGURE 2

Plots of mixing log-normal and fitted densities: 25 baseline segments, 5 covariate segments, $X = 0{\cdot}2$

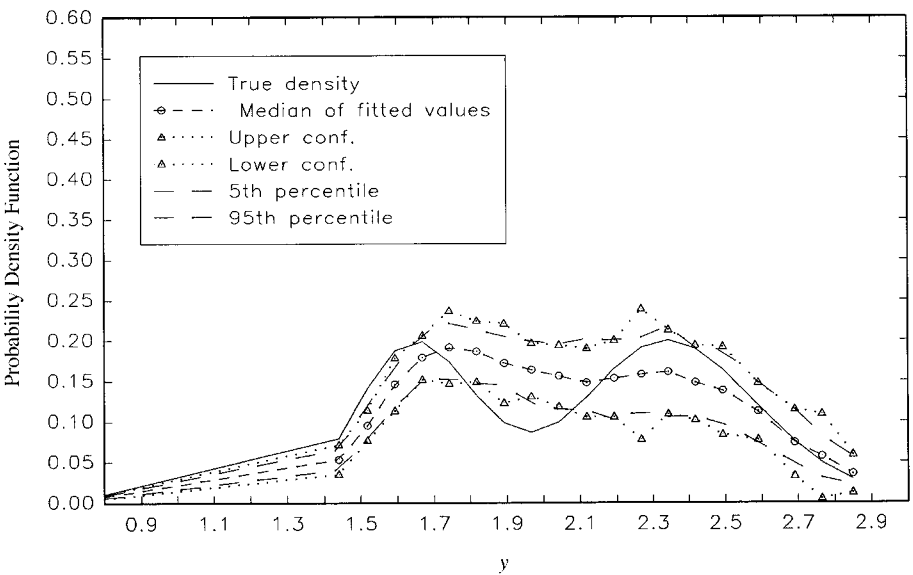

FIGURE 3

Plots of mixture log-normal and fitted densities: 20 baseline segments, 1 covariate segment, $X = 0{\cdot}7$

20 and a P of 1. From Figure 3, one can see that insufficient flexibility, in the form of too few baseline and covariate segments, leads to visibly poorer approximations to the true density. Finally, in Figure 4 we plot the true and fitted cumulative distribution functions for the J-of-25, P-of-5, and X-of-0·7 case. As one might expect from Figure 1, the median

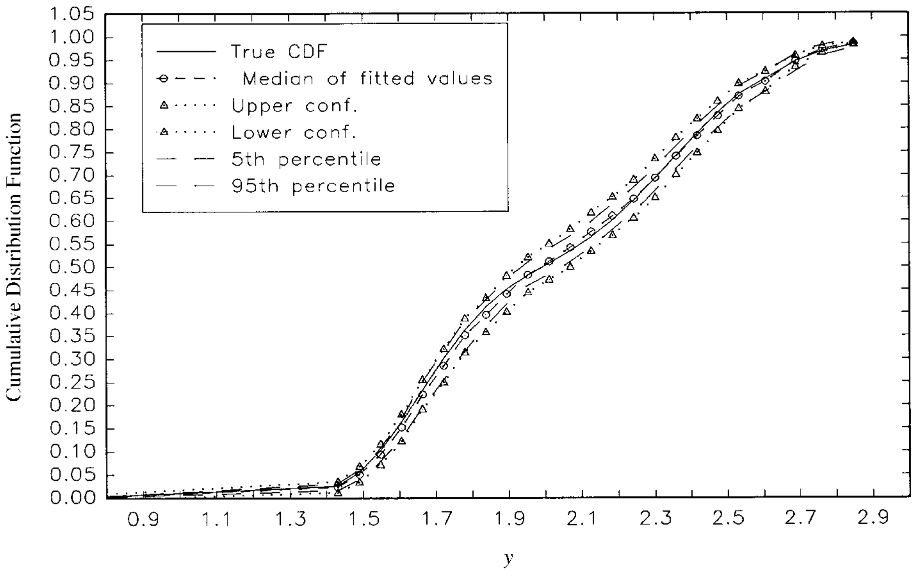

FIGURE 4

Plots of mixture log-normal and fitted distribution functions: 25 baseline segments, 5 covariate segments, $X = 0{\cdot}7$

fitted CDF is very close to the true value, and the constructed confidence band is close to the 5-th and 95-th percentiles of the distribution of fitted CDF values.

3. AN EMPIRICAL EXAMPLE: DIFFERENCES IN WAGE DISTRIBUTIONS BETWEEN CANADA AND THE UNITED STATES

We now turn to our examination of differences between the 1989 American and Canadian wage distributions. The Canadian data are from the 1989 *Labour Market Activity Survey* (LMAS). The American data are from the outgoing rotation group for the March 1989 *Current Population Survey* (CPS). Both data sets are large samples of individuals containing information on income sources as well as individual demographic information such as age and education. The two data sets are very similar in their target populations and questions asked. Neither is strictly representative of the working population in the relevant country, and assigned sampling weights from the surveys are used throughout our estimation. In the American data, we obtain our measure of usual hourly wages by dividing usual-weekly-wages by usual-hours-per-week. The LMAS questions respondents in January and February concerning up to five jobs held in the previous calendar year. The information gathered includes union status, industry, occupation, usual hourly wage and usual hours per week for each job. To match the American data, we use data from jobs held in the first two weeks of March, 1989.[9] The LMAS data are from the surveys taken in January and February of 1990 and thus will include recall bias not present in the American data. Otherwise, the two data sets are very comparable. We restrict the sample to full-time (35 or more hours per week), non-self-employed, paid jobs held by males age

9. For individuals with multiple jobs in these weeks, we use the job with the highest number of hours per week.

20 to 64. We convert Canadian values to their American equivalents using the 1990 bilateral Purchasing-Power-Parity (PPP) exchange rate for household consumption created by the OECD; see OECD (1993). This exchange rate is translated to 1989 using relative changes in the CPI for the two countries. The actual exchange rate we use is 1·257. While the results presented here are based on this PPP exchange rate, we also generated another set of results using the average spot-exchange rate for 1989, 1·180. We discuss differences in results arising from differences in the exchange rates at the end of the paper.

3.1. *Empirical results*

In Figure 5, we plot the CDFs for American and Canadian male wages in 1989. It is clear from the plots that the Canadian distribution is more equal. The means of the two distributions are very similar, US$11.67 for Canada and US$12.27 for the United States. Thus, the shapes and the locations of the distributions provide a picture that accords with the mythologies of the two countries: the more egalitarian Canadian society provides relatively better earnings for poorer workers, while the more incentive-oriented American society provides better returns for those at the top of the distribution; see Lipset (1990).

Plausible explanations for the differences in the two distributions include higher unionization rates and lower variability in school quality in Canada. Another possibility is that the more-generous Canadian social safety net may mean both that individuals do not have to accept low-paying, full-time jobs and that the associated higher taxes make taking the risks leading to much higher earnings less attractive. Under these explanations, one might expect the pattern depicted in Figure 5 to be repeated for different skill groups in the populations. Alternatively, the distributions within skill groups may be similar between the two countries, but the proportions of the workforce in the various skill groups may differ. Evaluating whether Figure 5 predominantly reflects one of these alternatives or some combination of the two provides insight into the sources of inequality in the two

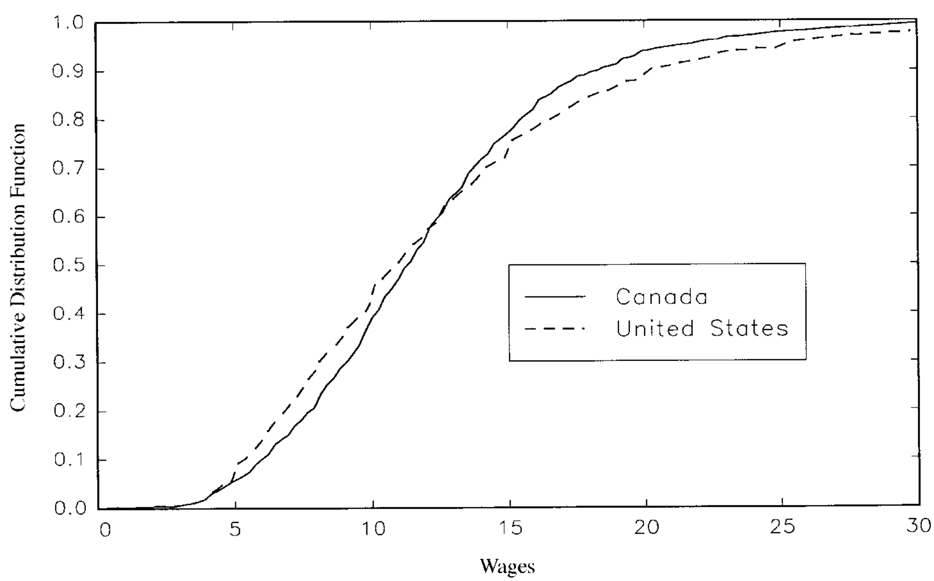

FIGURE 5

Cumulative distribution functions: hourly wages countries and into who benefits most from the earnings structure in each. In the remainder of this paper, we carry out such an evaluation.

We estimate the hazard-based, wage-generation model described in section 2 separately for the two samples. Using the rule of thumb described in Subsection 2.1, we chose to use 94 baseline segments, with uniform segment widths between the 5-th and 95-th percentiles of the unconditional wage distribution obtained by pooling the wage data from the two countries and wider segments in the two tails. Based on the Monte Carlo results in Table 1, combined with the fact that we have more observations than were used in those experiments, we chose 10 covariate segments with dividing points at approximately the deciles of the unconditional pooled-wage distribution. The covariate vector includes: dummy variables corresponding to the education categories, less-than-high-school or some-but-not-completed-high-school, some- or completed-post-secondary, and completed-university (the base category is completed-high-school);[10] categorical age variables; categorical industry variables; an indicator variable for visible minorities; an indicator variable for union members; the percentage of workers unionized in the individual's two-digit industry; and the latter variable interacted with union status.[11] Sample means for the covariates are presented in Table 2. Note that estimates of the densities and the CDFs are non-linear functions of the covariates and their associated coefficient estimates. Thus, as in standard Probit estimation, marginal effects on the ordinates from varying covariates are functions of values for all covariates. In such situations, presenting results using interesting person-types greatly eases exposition and interpretation.

In Figures 6, 7, 8, and 9, we present plots of CDFs and the associated confidence intervals for Canada and the United States for each of four person-types: not-completed-high-school, age 20 to 24; completed-high-school, age 20 to 24; completed-university, age 25 to 34; and completed-university, age 45 to 54. The age categories are chosen to correspond to low-experience workers in the first three cases and high-experience workers in the fourth case. In each case, we use only white workers to avoid complications due to racial differences between the two countries. The confidence intervals are formed as plus or minus 2 times the estimated standard errors for the CDF estimates at each baseline segment.[12] The plots for low-experience, not-completed-high-school workers are striking: the Canadian CDF for this worker type lies everywhere to the right of the American CDF. The confidence bands for the two CDFs do not overlap, and the difference appears economically substantial: the 80-th percentile of the Canadian distribution occurs at US$9.38 compared to only US$7.13 in the American distribution. The CDFs for low-experienced workers who have completed high school, plotted in Figure 7, also follow a pattern with Canada dominating. In this case, however, the two CDFs are not point-wise statistically significantly different below the 40-th percentile, and the difference thereafter

10. The some- or completed-post-secondary category includes those obtaining a post-secondary diploma or certificate other than a B.A. and those who attended, but did not complete a post-secondary programme. The completed-university category includes all workers with a B.A. or more. These groupings are directly reported in the Canadian data. In the American data, the educational categories are defined as follows: less-than-high-school contains workers with less than 12 years of school attended or 12 years attended but only 11 years completed; high-school-completed contains those with exactly 12 years of education completed; post-secondary contains those with 13 to 15 years of education attended or 16 years attended but only 15 years completed; and completed-university contains those with 16 years attended and completed or more than 16 years attended. Individuals obtaining a GED are not captured in this division for the United States.

11. Age is not a continuous variable in the LMAS. We use all the age categories available in the LMAS and group the American data to make them comparable.

12. To form these plots, we draw the subsamples of individuals who are in the specified age and education groups and who are white. We calculate fitted CDFs for each individual in the subsample using estimated parameters from the hazard-based estimation and then calculate the average of these fitted CDFs.

TABLE 2

Sample descriptive statistics
full-time males, 1989

Variable	Canada	Variable	United States
Wage	11.67	Wage	12.27
Age:		Age:	
20-24	0·09	20-24	0·10
25-34	0·33	25-34	0·34
35-44	0·30	35-44	0·28
45-54 (Base)	0·18	45-54 (Base)	0·18
55-64	0·10	55-64	0·10
Education:		Education:	
Less Than High School	0·28	Less Than High School	0·15
High School (Base)	0·23	High School (Base)	0·37
Some Post-Secondary	0·33	Some Post-Secondary	0·22
University	0·16	University	0·26
Union Member or Covered by Collective Agreement	0·48	Union Member or Covered by Collective Agreement	0·24
Visible Minority	0·04	Visible Minority	0·11
Occupation:		Occupation:	
White-Collar Professional	0·29	White-Collar Professional	0·33
White-Collar Other	0·19	White-Collar Other	0·21
Blue-Collar (Base)	0·52	Blue-Collar (Base)	0·46
Industry:		Industry:	
Primary and Construction	0·11	Primary and Construction	0·11
Manufacturing (Base)	0·30	Manufacturing (Base)	0·30
Transportation	0·13	Transportation	0·10
Trade	0·15	Trade	0·17
Service	0·21	Service	0·25
Public Administration	0·10	Public Administration	0·07
Sample Size	15,142	Sample Size	6,384

is much smaller than that observed in Figure 6. For high-school-educated workers, the 80-th percentile in the American distribution occurs at US\$9.13 compared to US\$10.30 in the Canadian. The superior CDF for less-than-high-school educated workers in Canada combined with the more similar CDFs for high-school completers in the two countries points to a smaller premium to completing high school in Canada. The difference between the 80-th percentiles of the high-school completers and the non-completers wage distributions is US\$0.92 in Canada, but US\$2.00 in the United States. Finally, for both high-school completers and non-completers, the Canadian CDFs are flatter than those for the American, indicating that there is greater inequality in Canadian wages for these skill groups.[13] The patterns in the CDFs for university-educated workers, shown in Figures 8 and 9, are very different from their less-educated counterparts. The CDFs for younger, university-educated workers presented in Figure 8 are virtually coincident over much of the wage range and are never point-wise statistically significantly different. For older university-educated workers, the CDFs for the two countries are very similar and are not point-wise statistically significantly different at any point up to the 60-th percentile. The American distribution is superior in the upper tail, with point-wise statistically significant

13. It is possible that the greater variability in Canadian wages is due to recall bias that is surely absent in the American data. However, if this were true then one might expect greater variability in the distributions for more educated workers, which, as we shall see, is not true.

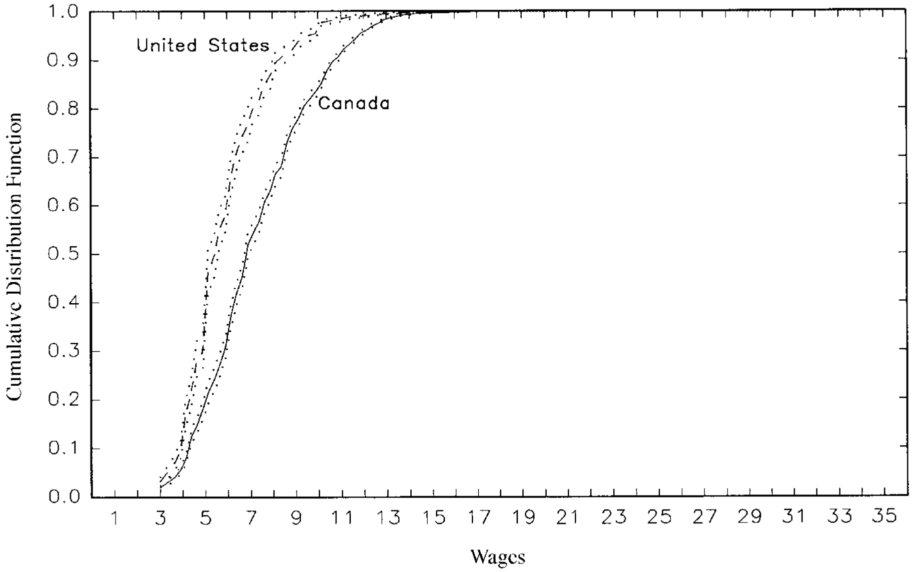

FIGURE 6
Cumulative distribution function fitted plots: wages for less-than-high-school-educated workers, age 20-24

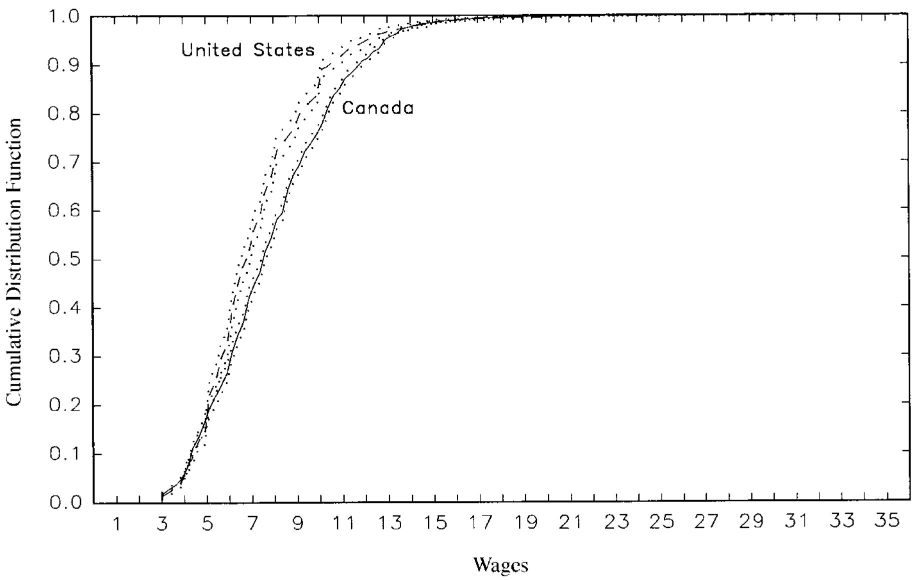

FIGURE 7
Cumulative distribution function fitted plots: wages for high-school educated workers, age 20-24

differences between the two countries occurring between the 60-th and the 80-th percentiles.

Overall, no evidence exists suggesting that Canadian institutions compress the wage distributions of workers over *all* skills groups. Rather, these Figures suggest that low-skilled Canadians have superior earnings, but face a more unequal distribution. The most

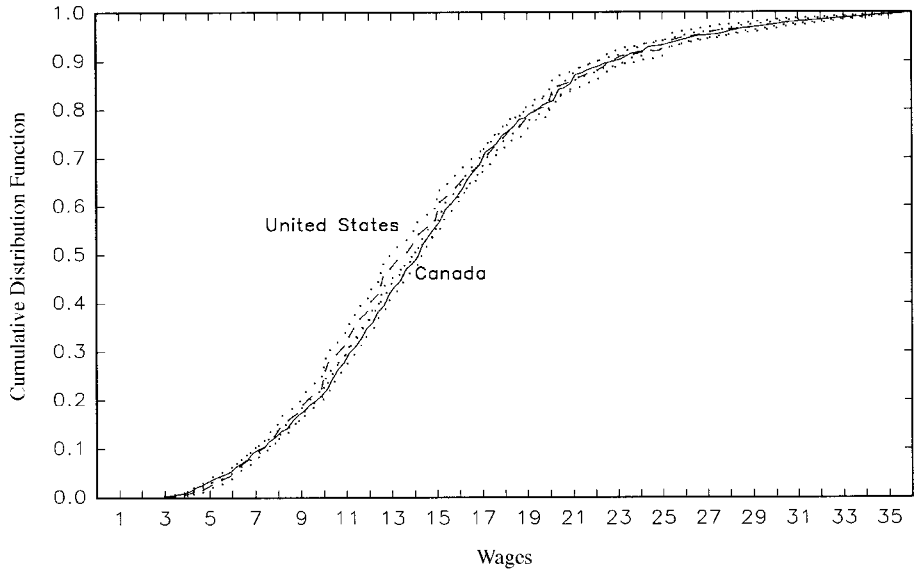

FIGURE 8

Cumulative distribution function fitted plots: wages for university-educated workers, age 25-34

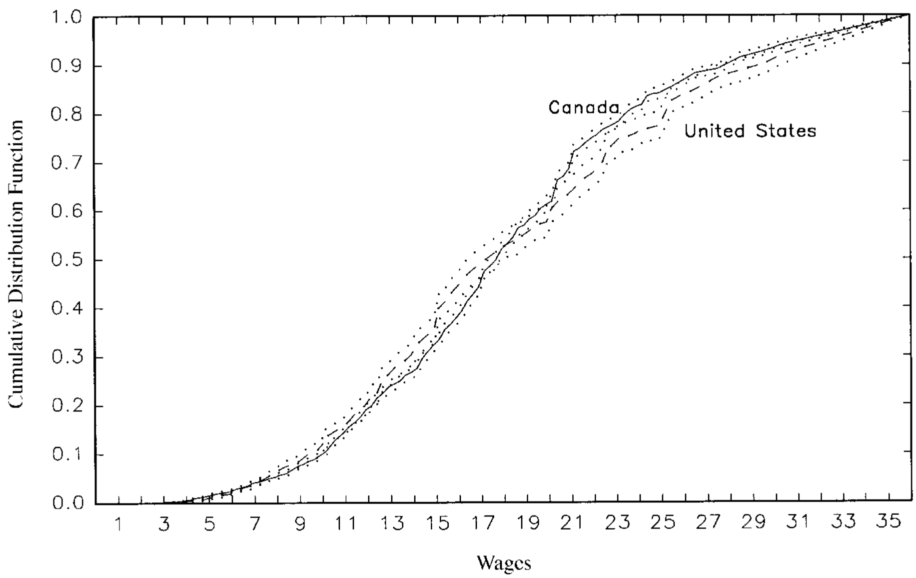

FIGURE 9

Cumulative distribution function fitted plots: wages for university-educated workers, age 45–54

educated, in contrast, fare slightly better in the United States. Finally, the American densities appear less dispersed than the Canadian densities; the opposite of the mythical more-egalitarian Canadian distribution.

The markedly superior CDF for the wages of high-school non-completers in Canada versus the United States combined with the more similar CDFs for higher-skilled workers indicates that a key to understanding why the low end of the American and Canadian overall wage distributions are different is understanding why low-skilled American workers are paid so much less than their Canadian counterparts. Differences in unionization rates appear to be a likely place to start when trying to explain the wage differences of low-paid workers in the two countries. Sample means presented in Table 2 indicate that Canadian workers are twice as likely to be unionized as their American counterparts. In Figures 10 and 11, we present the fitted wage densities for non-union and union less-than-high-school educated workers, respectively.[14] These densities are presented smoothed, using a simple three-point moving average, to make comparisons of overall patterns in the densities easier to spot. Non-smoothed plots (available on request) show no evidence of substantial spikes in either the American or Canadian densities at minimum-wage points. These results are consistent with the findings of DFL; *viz*. that the real minimum wage declined enough in the 1980s in the United States to make it irrelevant as a determinant of the male wage distribution. Thus, it appears that differences in wages for the low-educated in the two countries cannot be attributed to differences in minimum wages.

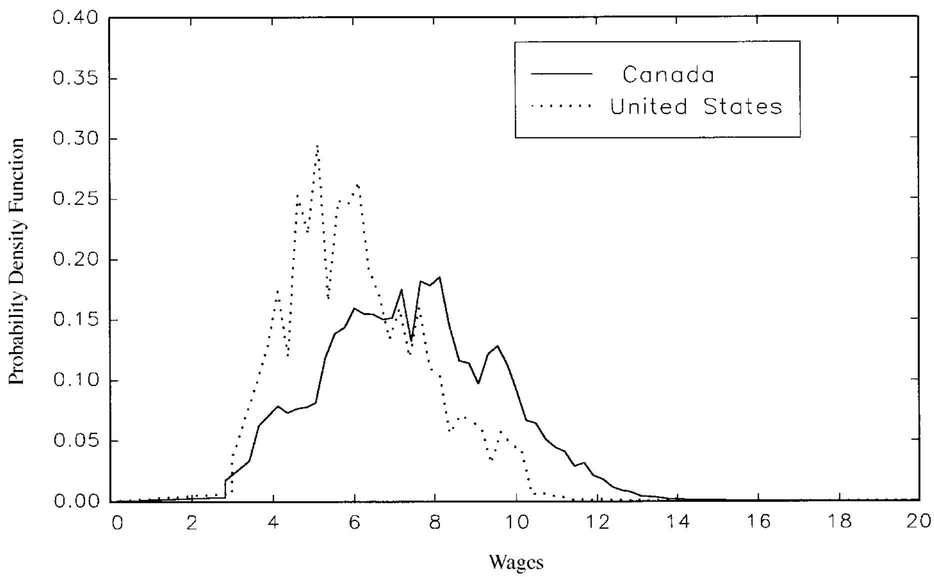

FIGURE 10

Fitted density plots: nonunion, less-than-high-school-education, age 20-24

Union wages have less dispersion than non-union wages in these data sets, as has been found in studies such as Lemieux (1993). The variance in the logarithm of wages for American males with less-than-high-school education in the manufacturing sector is 0·11 for union workers and 0·19 for non-union workers. However, once one draws the densities, as in Figures 10 and 11, one notes interesting and complex differences in union effects for the two countries. In the American case, the non-union density is more skewed than the union density, with more mass near \$5 but also significant mass up to \$10. The union density is more symmetric and concentrated near \$8. This result fits with the idea that unions generate greater wage equality for a given skill type mainly by creating the largest

14. Apart from union status, the worker characteristics used in creating these plots are: white, age 20 to 24, less-than-high-school educated, in a blue-collar occupation in the manufacturing sector, and in an industry with average union concentration for the manufacturing sector.

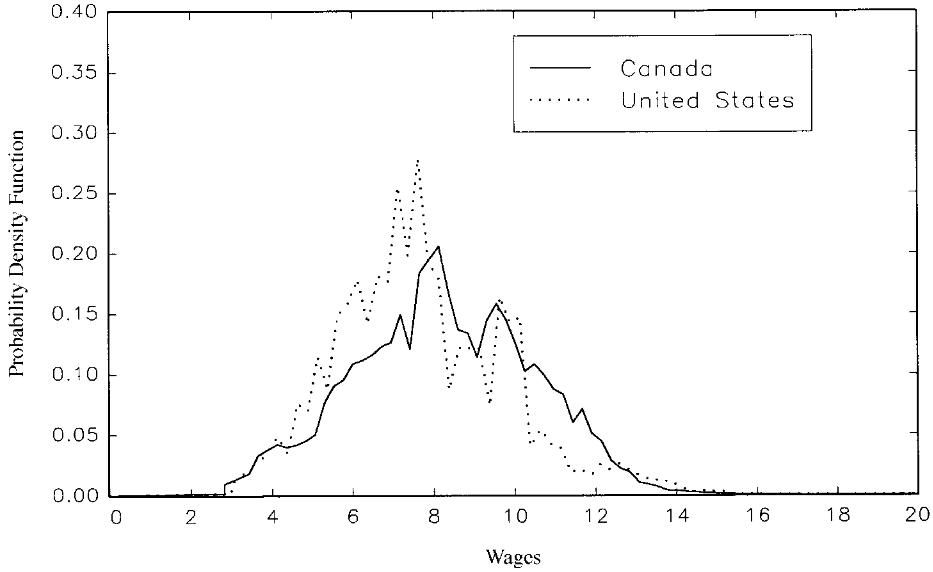

FIGURE 11
Fitted density plots: union, less-than-high-school-education, age 20-24

wage increases for the least well-paid. In contrast, the Canadian non-union density appears to have the same basic shape as the union density, but with more dispersion.

Comparing American and Canadian densities reveals that the Canadian non-union density has more mass above \$8 and less below \$6 than its American counterpart, and appears to be more symmetric. The American and Canadian union densities are much more similar, with only a small superiority for the Canadian density. Thus, the question of why the less-educated are paid more in Canada can, to some extent, be refined further: Why do the least-educated non-union workers fare so much better in Canada? One possible answer is that the much higher unionization rate in Canada means more pervasive and credible union-threat effects that bring benefits to the non-union distribution. This is also a potential explanation for the more symmetric "union-like" Canadian non-union wage density. In Figure 12, we present the non-union fitted densities for an American of the same type as in the previous two Figures evaluated both with the union-concentration variable set to its average for the manufacturing sector in the United States (0·28) and in Canada (0.53). Increasing the concentration rate raises average earnings and makes the non-union wage density more symmetric. A reduction in the mass up to \$5 is offset largely by an increase in the mass in the \$6 to \$8 range, with little effect at \$9 or above.[15] The greater benefit for low-wage earners relative to medium- or high-wage earners makes the density more symmetric, moving it toward the shape of the union density. Of course, increasing unionization rates means moving workers from the non-union to the union sector and these workers would not be selected at random. This implies additional changes in density shapes that may not be fully captured in our results. Thus, a counter-factual exercise of this sort does not provide a full picture of what the American non-union density would look like with higher unionization rates. It does, however, provide a graphic

15. The extra symbols on the Canadian union-concentration line denote points at which points on the two lines differ by more than two standard errors.

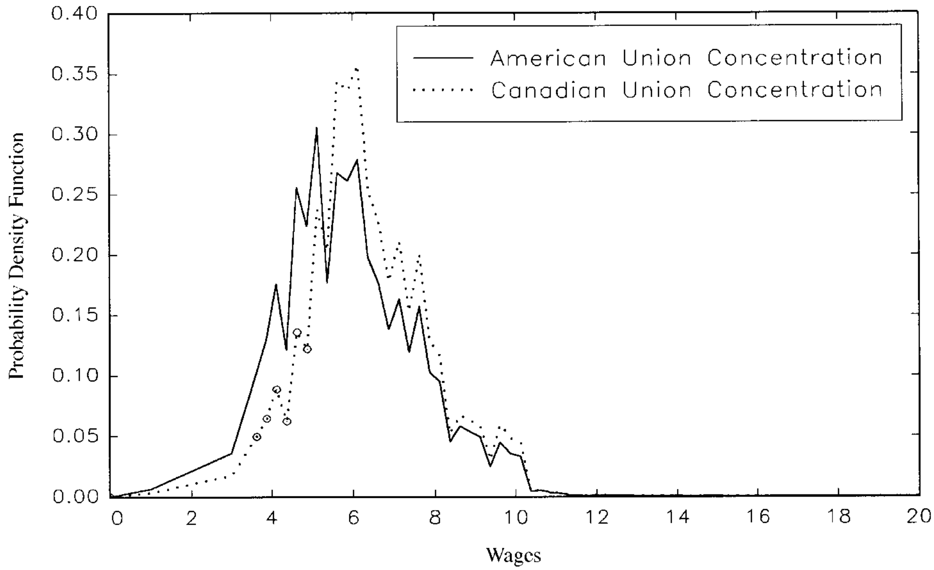

FIGURE 12
Fitted density plots: U.S., nonunion with average and high union concentration

illustration of who in the non-union sector, at the margin, benefits from increased unionization. The distribution of these benefits implies a movement of the American non-union wage distribution toward the Canadian distribution as concentration rates increase.[16] A plot of the fitted CDF corresponding to an American facing Canadian concentration rates in Figure 13 shows that differences in union-concentration rates over-explain the Canadian-American difference in CDFs below \$6 for this person type, but explain virtually none of the difference above \$8. In contrast, density plots for university-educated workers show little impact of the worker being unionized or of changes in union-concentration rates.[17]

Based on both the direct and indirect effects of unionization on the wage density, it is plausible that differences in unionization could play a large role in explaining differences at the low end of the wage distribution between Canada and the United States. To pursue this possibility, we construct a fitted CDF in which Americans face the wage structure in the United States (*i.e.*, $\hat{\delta}$ for the American sample is used), but the Canadian union structure. Thus, using the Canadian sample, we estimate a Probit on union status as a function of all the covariates used in the wage estimation, apart from union status itself and the union-concentration rate. Using parameters from this estimation, we form a probability of being covered by a collective agreement for each member of the American sample using their own covariate values, and then replace their union dummy variable with this fitted probability. We also assign each worker in the American sample the Canadian union-concentration rate for their two-digit industry. We then form a CDF estimate for each American worker and then average at each baseline point to create the final fitted CDF.

16. To validate these results, we selected a subsample similar to the workers depicted in Figure 12: specifically, a subsample of American, non-union, blue-collar workers. Using this subsample, in a regression of the logarithm of wages on the same set of regressors used in the hazard-based estimation, the estimated coefficient on the union-concentration variable was 0·545, with a standard error of 0·084. This implies that an increase in union concentration from 0·28 to 0·53 would raise average wages for these workers by approximately 14%.

17. These figures are available from the authors on request.

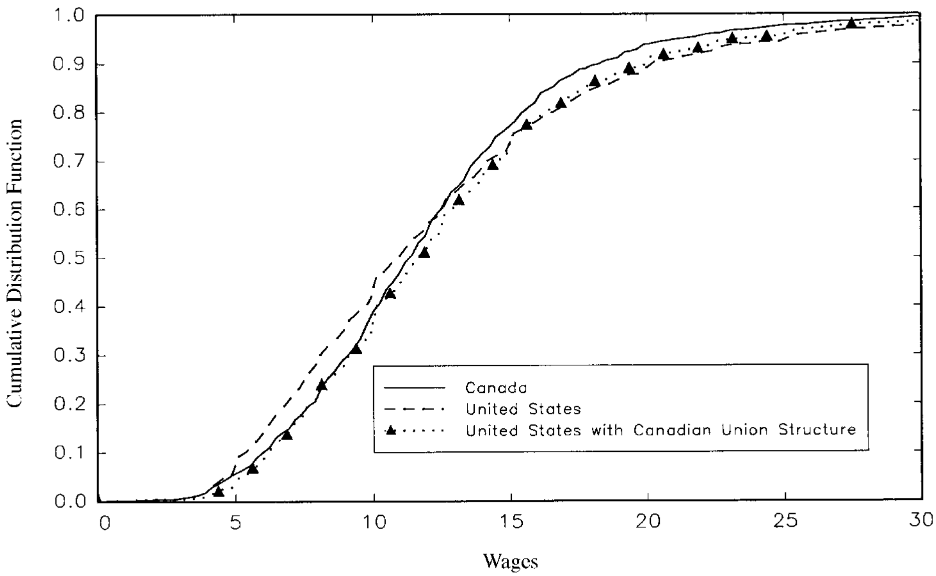

FIGURE 13
Cumulative distribution functions: hourly wages with U.S. given Canadian union structure

The resulting CDF is plotted in Figure 13 along with the simple wage CDFs for Canada and the United States. Based on these estimates, differences in union structure over-explain the gap in the CDFs for the two countries between \$5 and \$10 by 24%, and explain 30% of the gap above \$15. Assigning Americans the probability of being unionized taken from the Canadian economy, but not the Canadian union-concentration rates, yields a fitted CDF that explains approximately 76% of the CDF gap below \$10 and 11% above \$15. In comparison, Lemieux (1993) found that differences in union status between Canada and the United States explained approximately 40% of the difference in the variance in the logarithm of wages between the two countries.

Finally, in Figure 14 we present the fitted-wage CDF formed using the Canadian sample, but the wage-structure parameter estimates from the United States. Comparing this to the simple Canadian CDF shows the effect of differences in wage structure (*i.e.* differences in covariate effects and the shape of the underlying baseline hazard) in explaining differences in the distributions between the two countries. Comparing the fitted CDF to the CDF of the United States shows the effect of differences in the distribution of covariate characteristics, holding the wage structure constant. Based on these comparisons, differences in covariate distributions explain virtually all of the gap in the CDFs between \$5 and \$10, and 80% of the gap above \$15.[18] Thus, compared to Figure 13, moving beyond just giving American workers the Canadian union structure to giving them the whole Canadian covariate distribution leads to a worsening across the whole distribution. In a simulation in which American workers were given the Canadian union structure as above and predicted industry and occupation values based on parameters from a multinomial-Logit model of industry and occupation choice using Canadian data, the percentage of the gap explained was virtually unchanged relative to the case where

18. We replicated the exercises depicted in Figures 13 and 14 using only white workers in both countries and obtained extremely similar results. Thus, our conclusions are not confounded by different racial effects in the two countries. The alternative versions of Figures 13 and 14 are available from the authors on request.

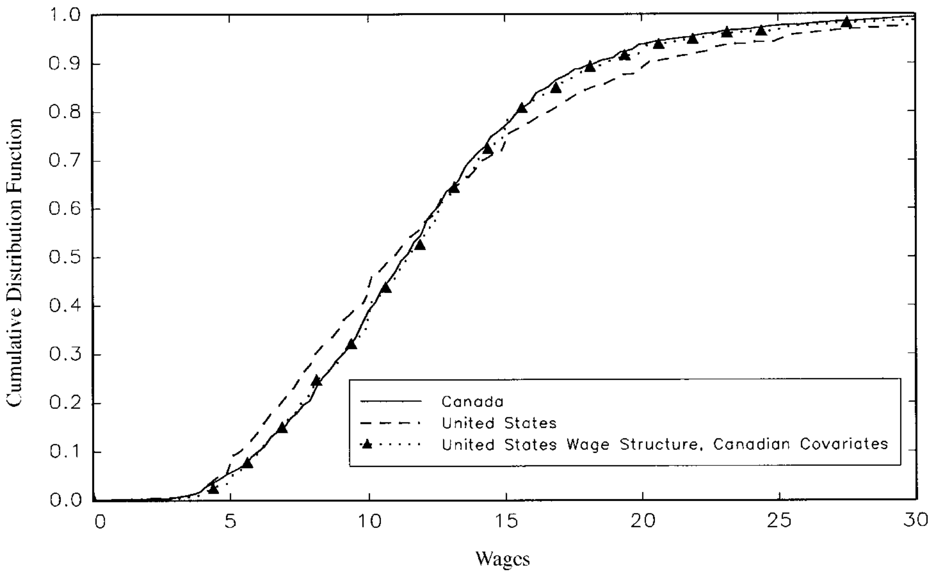

FIGURE 14
Cumulative distribution functions: hourly wages with U.S. wage structure and Canadian covariates

they were given only the Canadian union structure. Thus, the worsening in the fitted CDF in Figure 14 is due to the remaining covariates: the education variables. The overall implication is that Canada has a superior wage distribution at the low end of the wage range largely because of differences in unionization between the two countries. Lower education levels in Canada partially offset this effect at the lower end and cause a significant worsening at the upper end. It seems plausible, although there is no direct evidence on this point in this paper, that the two outcomes are related: unions have a beneficial impact in improving the wages of low-paid workers in both the union and non-union sectors, but that this improvement may have a negative impact on educational attainment. Seeing the small return to completing high school in Canada relative to the United States, it is not surprising that fewer individuals complete high school in Canada. In all of this, one cannot rule out the possibility that Canadian wages for the least-skilled are more symmetrically distributed with a higher mean for a reason such as some other facet of labour legislation and that unions find workers in this situation easier to organize. In this case, the results above cannot be attributed to direct union actions, but still indicate that the union/non-union distinction is a useful starting place in searching for the root cause of differences between the two countries. Addressing the selection issue directly is of clear importance, but beyond the scope of this paper.

Several researchers have noted previously the dramatic decline in unionization in the United States and have examined its impact on the rapid growth in wage inequality in the United States. Card (1992), Freeman (1991), and DiNardo *et al.* (1996) all find that declines in unionization account for about 20% of the increase in the variance of male wages in the United States over the 1970s and 1980s. DiNardo and Lemieux (1997) find that differences in unionization-rate changes between Canada and the United States over the 1980s account for two-thirds of the differential in wage inequality growth between the two countries in that period. Our results reveal that differences in unionization rates have impacts primarily at the low end of the wage distribution. Further, we show that unions have an indirect impact on non-union wages, making the Canadian non-union distribution less skewed, with more mass at higher wages relative to its American counterpart. By drawing the entire wage distribution for various skill types, one can see that a key difference between the distributions in the two countries is the high concentration of American non-union workers at wages below \$6 per hour. This appears to be related to union-threat effects and is an element of the differences between the two countries that is missed by focussing on scalar measures such as the variance in the logarithm of wages.

The exchange rate is of central importance in many parts of the cross-country comparison carried out above. In the calculations above, we use a PPP exchange rate for Canada and the United States. We have also generated all of our results using 1·18, the average spot-exchange rate for 1989. In our view, the PPP rate is clearly superior since it sets the comparisons in terms of goods that can be consumed. Using the spot exchange rate yields three main results: first, the smaller return to completing high school in Canada remains since this is based on comparisons within countries; second, the less-than-high-school educated Canadian workers have an even more substantial advantage over their American counterparts; and third, American and Canadian workers with higher education face almost identical earnings and wage distributions. The main difference in interpretation is that with the alternate exchange rate, the American dominance in the upper part of the distribution is due almost entirely to differences in the educational composition of the two work forces. By comparison, in the results above, the American dominance at the upper end is due to both higher returns to education and a more educated workforce.

4. SUMMARY AND CONCLUSIONS

In this paper, we investigated the sources of differences in the wage distributions between Canada and the United States in 1989. We found, perhaps not surprisingly, that the Canadian wage distribution has less dispersion than the American, and a slightly lower mean. Decomposing wage distributions by skill groups based on age and education, we found that the pattern evident in the overall distribution is not present within skill groups. Thus, Canadian institutions do not generate greater equality and lower mean earnings for each skill group relative to their American counterparts. Rather, the effect seen in the overall distribution arises from a combination of lower penalties to not completing high school in Canada and a larger proportion of Canadian workers having low education. Also, higher unionization rates in Canada appear to play a very large role in generating superior wages for those at the lower end of the distribution in Canada relative to the United States. Lower education levels among Canadian workers offset this union effect at the low end of the distribution and lead to a substantially poorer Canadian distribution at the upper end.

Reaching these conclusions depended crucially on using an estimator that permitted estimating complex differences in the shapes of density functions for different values of a covariate vector. We proposed and implemented a hazard-based estimator of wage, earnings, and income densities in the presence of covariates. Because the estimator makes use of techniques that are standard in the analysis of spell durations, it should be familiar to many empirical researchers. The estimator allows one to examine the effects of covariates of interest on the entire density function, while holding the effects of other covariates constant. This provided important added information when studying the effects of institutions such as minimum wages and labour unions. The estimator is potentially useful in other circumstances where studying the entire distribution of wages, earnings or income is required.

Acknowledgements. Green and Paarsch wish to thank the SSHRC of Canada for financial support. The authors are also grateful to Charles M. Beach, Paul Beaudry, Michael I. Cragg, James B. Davies, Barton H. Hamilton, James J. Heckman, Joel L. Horowitz, Thomas E. MaCurdy, Thomas A. Mroz, James L. Powell, Jeffrey Smith, and Frank A. Wolak as well as two anonymous referees for useful comments and helpful suggestions. Benoit Delage provided excellent research assistance.

REFERENCES

BLACKBURN, M. and BLOOM, D. (1993), "The Distribution of Family Income: Measuring and Explaining Changes in the 1980s for Canada and the United States", in D. Card and R. Freeman (eds.) *Small Differences that Matter*: *Labor Markets and Income Maintenance in Canada and the United States* (Chicago: University of Chicago Press), 233-266.

BUCHINSKY, M. (1994), "Changes in the U.S. Wage Structure 1963–1987: Application of Quantile Regression", *Econometrica*, **62**, 405-458.

CARD, D. (1992), "The Effects of Unions on the Distribution of Wages: Redistribution or Relabelling?" (Working Paper No. 4195, National Bureau of Economic Research).

DiNARDO, J., FORTIN, N. and LEMIEUX, T. (1996), "Labor Market Institutions and the Distribution of Wages, 1973–1992: A Semiparametric Approach", *Econometrica*, **64**, 1001-1044.

DiNARDO, J. and LEMIEUX, T. (1997), "Diverging Male Wage Inequality in the United States and Canada, 1981–1988: Do Institutions Explain the Difference?", *Industrial and Labor Relations Review*, **50**, 629-651.

FORTIN, N. and LEMIEUX, T. (1995), "Rank Regressions, Wage Distributions, and the Gender Gap", *Journal of Human Resources*, **33**, 610-643.

FREEMAN, R. (1991), "How Much has De-Unionization Contributed to the Rise in Male Earnings Inequality?" (Working Paper No. 3826, National Bureau of Economic Research).

GRITZ, M. and MaCURDY, T. (1992), "Unemployment Compensation and Episodes of Nonemployment", *Empirical Economics*, **17**, 183-204.

HANRATTY, M. and BLANK, R. (1992), "Down and Out in North America: Recent Trends in Poverty Rates in the United States and Canada", *Quarterly Journal of Economics*, **107**, 233-254.

HECKMAN, J. and SINGER, B. (1984), "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data", *Econometrica*, **52**, 271-320.

KALBFLEISCH, J. and PRENTICE, R. (1980), *The Statistical Analysis of Failure Time Data* (New York: John Wiley & Sons).

LEMIEUX, T. (1993), "Unions and Wage Inequality in Canada and the United States", in D. Card and R. Freeman (eds.) *Small Differences that Matter*: *Labor Markets and Income Maintenance in Canada and the United States* (Chicago: University of Chicago Press), 69-107.

LIPSET, S. (1990), *Continental Divide*: *The Values and Institutions of the United States and Canada* (New York: Routledge).

MEYER, B. (1990), "Unemployment Insurance and Unemployment Spells", *Econometrica*, **58**, 757-782.

OECD (1993), *Purchasing Power Parity and Real Expenditures for Canada and the United States*, 1990 (Paris: OECD).

SCOTT, D. (1979), "On Optimal and Data-Based Histograms", *Biometrika*, **66**, 605-610