Scottish Journal of Political Economy

Scottish Journal of Political Economy, DOI:10.1111/sjpe.12038, Vol. 61, No. 2, May 2014

BAYESIAN UNCONDITIONAL QUANTILE REGRESSION: AN ANALYSIS OF RECENT EXPANSIONS IN WAGE STRUCTURE AND EARNINGS INEQUALITY IN THE US 1992–2009

Michel Lubrano and Abdoul Aziz Junior Ndoye***

Abstract

We develop Bayesian inference for an unconditional quantile regression model. Our approach provides better estimates in the upper tail of the wage distribution as well as valid small sample confidence intervals for the Oaxaca–Blinder decomposition. We analyze the recent changes in the US wage structure using data from the CPS Outgoing Rotation Group from 1992 to 2009. We find that the largest part of the recent changes is explained mainly by differences in returns to education while the decline in the unionization rate has a small impact, and that earnings inequality is rising more at the top end of the wage distribution.

I Introduction

Introduced by Koenker and Bassett (1978), quantile regression models aim at modeling the effect of the explanatory variables on the conditional distribution of the outcome variable. Quantile regression have been increasingly used in empirical labor market studies, to describe parsimoniously the entire wage conditional distribution (see e.g., Buchinsky, 1994; Chamberlain, 1994; Machado and Mata, 2001). Several competing methods of estimation in both classical and Bayesian frameworks have been recently developed (see for instance Yu and Moyeed, 2001; Kozumi and Kobayashi, 2011, or Kottas and Krnjajic, 2009 for the Bayesian side with a semiparametric approach for the last reference). As any quantile can be used in any part of the outcome distribution, the quantile regression models are more flexible and more robust to outliers than the classical mean regression models.

While the conditional quantile regression models can be useful, they are very restrictive. *First*, a change in the distribution of covariates may change the interpretation of the coefficient estimates. This point is illustrated for instance in Powell (2011). To overcome this restriction, Firpo *et al.* (2009)

*Aix-Marseille University (AMSE) and GREQAM-CNRS
**Aix-Marseille University (AMSE) and GREQAM

have proposed a new regression method which evaluates the impact of changes in the distribution of the explanatory variables on the quantiles of the unconditional distribution of the outcome variable. *Second*, the property that, in the popular Oaxaca–Blinder decomposition method of a simple linear regression, differences in unconditional means are equal to differences between conditional means is no longer valid for conditional quantile regressions. As explained in, e.g., Firpo *et al.* (2011), with conditional quantile regressions, the difference in unconditional quantiles is not equal to difference in conditional quantiles. This question has received several answers in the literature, see e.g., Juhn, Murphy and Pierce (1993), DiNardo *et al.* (1996), or Machado and Mata (2005), but none of these methods can be used to decompose general distributional measures in the same way that the means can be decomposed using the conventional Oaxaca–Blinder method. However, the method of Melly (2005) and the recentered influence function method of Firpo *et al.* (2009) (*RIF* regression) can perform a detailed decomposition very much in the spirit of the traditional Oaxaca decomposition for the mean (Firpo *et al.*, 2011).

In this study, we develop a Bayesian inference method for the *RIF* regression model of Firpo *et al.* (2009) in which we estimate the log wage distribution by a mixture of normal densities. The mixture of normal densities is pursued so as to produce a better fit in the tails of the wage distribution which are essential to have a precise evaluation of higher quantiles. As documented in Bahadur and Savage (1956), in the presence of a heavy tail distribution, a nonparametric approach using kernel smoothing can lead to unreliable inference. As a consequence, the presence of a heavy right-hand tail in the wage distribution can make less reliable the usual density kernel estimate used in the *RIF*-OLS method of Firpo *et al.* (2009). A semiparametric approach using mixtures of distributions would provide better estimates of the *RIF* regression coefficients for the upper quantiles. Finally, a Bayesian approach takes a better account of parameter uncertainty of the density estimation in the first stage of estimation and is pursued so as to propose valid confidence intervals for the Oaxaca–Blinder decomposition.

We illustrate our approach, analyzing the recent trends in US wage structure and earnings inequality. The recent rise in earnings dispersion in United States is remarkable. The literature dealing with the causes of this wage dispersion has exponentially increased over the past decades. Several competing explanations have been offered. Bound and Johnson (1992) attribute the changes to the skill-biased technological progress which increases the rate of growth of the relative demand for highly educated and ‘more-skilled’ workers (see also Mincer, 1993;Katz and Autor, 1999). Murphy and Welch (1992) stress the impact of the globalization which increases the rate of unskilled immigration workers and led to a decrease in the growth of the relative supply of skills (see also Katz and Murphy, 1992). DiNardo *et al.* (1996) focus on changes in labor market institutions, in wage setting norms including the decline in unionization, on the erosion of the real and relative value of the minimum wage.

Atkinson (2008) is inclined to be careful about these now traditional explanations and suggests to take seriously the new models of earnings formation. In his book, he reviews other alternative explanations such as Rosen's (1981) model of superstars and hierarchical models. He provides a complete descriptive analysis for the changing distribution of earnings in different OECD countries. He argues that 'while the race between technology and education is appealing, a constantly rising demand for educated workers does not lead to a constantly rising wage premium but to a stable wage differential, the size of which depends on the speed of a country's response to shortages of qualified workers'. Investigating Atkinson's proposals is on our research agenda. For the while, we aim simply at measuring the alternative role of some factors such as union, education, experience, and gender for explaining the recent changes in the US wage structure and in earnings inequality.

The study is organized as follows. In Section II, we review the conditional quantile regression models when using a likelihood function that is based on the asymmetric Laplace distribution (Yu and Moyeed, 2001), and we show the limitations of MCMC methods and Oaxaca–Blinder decomposition procedure used for conditional quantile regression. In Section III, we present a reliable Bayesian inference for the *RIF* regression of Firpo *et al.*, 2009 in which we estimate the log wage distribution by a mixture of normal densities. We provide an Oaxaca–Blinder decomposition procedure using our *RIF* regression method, and we show how to obtain reliable standard errors for each component of the decomposition using the draws of the *RIF* regression coefficients together with a procedure of Rao-Blackwellization. Section IV illustrates the approach using the CPS-ORG sample from 1992 to 2009 to analyze wage inequality in the United States. A sensibility analysis is considered. Section V concludes.

II Conditional Quantile Regression Models

Consider the usual linear regression model

$$y_i = x_i'\beta + \epsilon_i, \tag{1}$$

where (y_i, x_i), $i = 1, 2,\ldots,n$ are independent observations, y_i being the response variable, and $x_i' = (1, x_{i1}, \ldots, x_{ik})$ being the $(k + 1)$ known covariates. $\beta' = (\beta_0,\ldots,\beta_k)$ represents the $(k + 1)$ unknown regression parameters, and ε_i, $i = 1,\ldots,n$ are the error terms which are supposed to be independent and identically distributed. The unbiased estimation of β in this regression model requires that $\mathrm{E}(\epsilon_i|x_i) = 0$, without making any specific assumption on the parametric form of the distribution of ε_i.

A quantile regression model considers a similar linear regression as in Equation (1), but adds the fact that this regression can be estimated for every predefined quantile τ of the endogenous variable. So for the τ^{th} quantile, we have now the new regression model:

$$y_i = x_i'\beta_\tau + \epsilon_i, \tag{2}$$

where the parameter to be estimated is the $\beta'_\tau = (\beta_{0_\tau}, \ldots, \beta_{k_\tau})$. A coherent definition of this regression requires no longer that $E(\epsilon_i|x_i) = 0$, but that the τ^{th} quantile of ε is equal to zero. If $f_\tau(.)$ is the density of ε, this means that

$$\int_{-\infty}^{0} f_\tau(\epsilon_i|x)d\epsilon_i = \tau. \tag{3}$$

The quantile regression estimator for $\beta_\tau, \hat{\beta}_\tau$ first proposed in Koenker and Bassett (1978) does not consider a specific distribution for ε (so that $f(.)$ is left unspecified). It is simply given as the solution of the following minimization problem

$$\min_\beta \frac{1}{N}\sum_{i=1}^{n}\rho_\tau(y_i - x'_i\beta_\tau), \tag{4}$$

where $\rho_\tau(\cdot)$ is the check function or loss function defined as

$$\rho_\tau(u) = u \times (\tau - \mathbb{1}(u<0)), \tag{5}$$

$\mathbb{1}(\cdot)$being the indicator function. As this loss function is not differentiable (as a quadratic loss function would be), one has to use linear programming techniques to solve this problem.

Using the asymmetric Laplace distribution

Yu and Moyeed (2001) have proposed to specify the distribution of ε using the asymmetric Laplace distribution (ALD):

$$f(\epsilon_i|\tau) = \frac{\tau(1-\tau)}{\sigma}\exp\left\{-\frac{1}{\sigma}\rho_\tau(\epsilon_i)\right\}. \tag{6}$$

This density automatically fulfill the quantile restriction condition (3). For a symmetric Laplace process, the maximum likelihood estimator of the central tendency parameter is equal to the sample median. This property is generalized here for all quantiles so that the maximum likelihood estimator based on the complete likelihood

$$L(y_i|\beta_\tau, \sigma_\tau, \tau) = \sigma_\tau^{-n}\tau^n(1-\tau)^n \exp\left\{-\frac{1}{\sigma_\tau}\sum_i \rho_\tau(y_i - x'_i\beta_\tau)\right\} \tag{7}$$

provides exactly the same value as that provided by the estimator proposed in Koenker and Bassett (1978) for β_τ. With, however, the same difficulties as with the loss function $\rho_\tau(u)$, as Equation (7) is not differentiable at zero. A Bayesian approach does not lead to the same difficulties, as in a Bayesian approach the likelihood function (times the prior) has to be integrated and differentiability plays no role in integration.

Bayesian inference for conditional quantile regression

For inference on the parameters of interest β_τ and σ_τ, given τ and the observations on (X, Y), one has to specify a prior density. So the posterior distribution of $\beta_\tau, \pi(\beta|y)$ is proportional to

$$\pi(\beta, \sigma|y) \propto L(y|\beta)\pi(\beta, \sigma),$$

where $L(y|\beta)$ is the likelihood function given in Equation (7) and $\pi(\beta,\sigma)$ is the prior distribution of β and σ. Yu and Moyeed (2001) show that for any type of prior, including an improper prior, the posterior moments exist in this particular problem. They choose an improper prior as no conjugate prior is available when the model is presented in this form. The posterior density has to be integrated out and a simple random walk MCMC method is the more straightforward method to use. The method is available as package bayesQR in R (see Benoit, Al-Hamzawi, Yu and Van den Poel, 2012). As noted in Kozumi and Kobayashi (2011), the random walk Metropolis may be difficult to tune because a different tuning parameter has to be chosen for every value of τ so as to get an acceptation rate of around 25%.

Kozumi and Kobayashi (2011) propose a location-scale mixture representation of the asymmetric Laplace distribution that allows to find analytical expressions for the conditional posterior densities of the model. With these tools, they can propose first a conditional natural conjugate prior and second a Gibbs sampler. The merit of the Gibbs sampler is to avoid the specification of a candidate density and of a tuning parameter. The normal-inverted-gamma prior combines nicely with the conditional likelihood in the Gibbs sampler. We can note, however, that it seems difficult to elicit an informative prior because we should specify different hyperparameters for each quantile. The Gibbs sampler has an important drawback in this particular case compared with a direct Metropolis approach which is its extreme slowness due to the fact that one has to draw random numbers in an inverted generalized Gaussian density for each observation separately and this is a very slow operation.[1]

The assumption of an asymmetric Laplace distribution for the error term might seem restrictive. Two alternative solutions were proposed in the literature. Lancaster and Jun (2010) make use of a Bayesian empirical likelihood based on the results of Schennach (2005). Kottas and Krnjajic (2009) extend the results of Kottas and Gelfand (2001), and propose a nonparametric error distributions based on Dirichlet process mixture models. Both approaches make use of variants of the Metropolis–Hastings algorithm.

Oaxaca–Blinder decomposition and quantiles

The popular Oaxaca–Blinder decomposition (Blinder, 1973; Oaxaca, 1973) makes use of the property that, in a linear regression, the difference in unconditional means is equal to the difference between conditional means. If $y_i = x_i'\beta + \epsilon_i$, then $E(y_i) = E(x_i')\beta$. Applying this simple result to a Mincer wage equation where y is the log wage, we can explain the mean wage gap between, for instance, males and females as

[1] This remark concerning the Gibbs sampler is not a general remark concerning the respective performance of the Gibbs sampler vs. the Metropolis–Hastings sampler. For instance, every time we are in data augmentation problem, the Gibbs sampler appears as the natural solution.

$$
\begin{aligned}
E[y_{mi} - y_{fi}] &= E[x'_{mi}\beta_m + e_{mi}] - E[x'_{fi}\beta_f + \epsilon_{fi}] \\
&= E[x'_{mi}]\beta_m - E[x'_{fi}]\beta_f \\
&= [E(x'_{mi}) - E(x'_{fi})]\beta_m + E(x'_{fi})[\beta_m - \beta_f].
\end{aligned}
$$

This decomposition is estimated by replacing the expected value of the covariates by their sample mean and the β by their regression estimates. In a classical framework, this will be the OLS estimator, in a Bayesian framework the posterior expectation is used as a first approximation. This equation means that mean wage differences are explained first by the difference in average characteristics multiplied by the male coefficient (composition effect) and second by the difference in yield of female average characteristics expressed by $\hat{\beta}_m - \hat{\beta}_f$ (structure effect).

This result is not directly transposable to quantile regression as in a quantile regression $E(\varepsilon_i) \neq 0$. We would like to explain the difference between two unconditional quantiles as a function of the conditional quantiles. As recalled in Firpo *et al.* (2011), the difference in unconditional quantiles is not equal to the difference in conditional quantiles. This question has received several answers in the literature (see, e.g., Juhn *et al.*, 1993; DiNardo *et al.*, 1996; Machado and Mata, 2005, or Melly, 2005), but none of these methods can be used to decompose general distributional measures in the same way as means can be decomposed when using the conventional Oaxaca–Blinder method.

Juhn *et al.* (1993) have proposed a ‘plug-in’ procedure of Oaxaca decomposition which allows for the distribution of the error term to depend on the covariates. They make a strong assumption of ‘Conditional Rank Preservation’ which is hard to maintain especially in the presence of heteroskedasticity. DiNardo *et al.* (1996) have proposed a reweighing procedure using a kernel density estimation. However, if there are too many variables, it becomes impossible to estimate counterfactual distributions nonparametrically. Machado and Mata (2005), but also Melly (2005), have proposed a simulation method to compute the wage structure subcomponents of the detailed decomposition using a Monte Carlo approach. These components are computed by sequentially switching the coefficients of the quantile regressions for each covariate from their estimated valued. But, this method does not provide a consistent effect as the effect of the reweighed covariate of interest gets confounded by other covariates correlated with that same covariate.

Firpo *et al.* (2011) show that the method based on the estimation of *RIF* regressions proposed in Firpo *et al.* (2009) is fully appropriate for estimating the detailed components of both the wage structure effect and the composition effect. This is the method that we shall discuss in the next section and use as a basis for a Bayesian implementation.

III Unconditional Quantile Regression

The influence function (*IF*), first introduced by Hampel (1974), describes the influence of an infinitesimal change in the distribution of a sample on a real-valued functional distribution or statistics $v(F)$, where F is a cumulative distribution function. The (IF) of the functional v is defined as

$$IF(y, v, F) = \lim_{\varepsilon \to 0} \frac{v(F_{\varepsilon,\Delta_y}) - v(F)}{\varepsilon} = \frac{\partial v(F_{\varepsilon,\Delta_y})}{\partial \varepsilon}\Big|_{\varepsilon=0}, \tag{8}$$

where $F_{\varepsilon,\Delta_y} = (1-\varepsilon)F + \varepsilon\Delta_y$ is a mixture model with a perturbation distribution Δ_y which puts a mass one at any point y. The expectation of IF is equal to zero.

Firpo *et al.* (2009) make use of Equation (8) by considering the distributional statistics $v(.)$ as being the quantile function ($v(F) = q_\tau$) to find how a marginal quantile of y can be modified by a small change in the distribution of the covariates. They make use of the recentered influence function (RIF), defined as the original statistics plus the IF, so that the expectation of the RIF is equal to the original statistics.

Considering the τ^{th} quantile q_τ defined implicitly as $\tau = \int_{-\infty}^{q_\tau} dF(y)$, Firpo *et al.* (2009) show that the IF for the quantile q_τ of the distribution of y is given by

$$IF(y, q_\tau(y), F) = \frac{\tau - \mathbb{1}(y \leq q_\tau)}{f(q_\tau)},$$

where $f(q_\tau)$ is the value of the density function of y evaluated at q_τ. The corresponding RIF is simply defined by

$$RIF(y, q_\tau, F) = q_\tau + \frac{\tau - \mathbb{1}(y \leq q_\tau)}{f(q_\tau)}, \tag{9}$$

with the immediate property that

$$E(RIF(y, q_\tau, F)) = \int RIF(y, q_\tau, F)f(y)dy = q_\tau.$$

The illuminating idea of Firpo *et al.* (2009) is to regress the RIF on covariates, so that a change in the marginal quantile q_τ is going to be explained by a change in the distribution of the covariates by means of a simple linear regression:

$$E[RIF(y, q_\tau, F|X)] = X\beta. \tag{10}$$

They propose different estimation methods: an ordinary least square (OLS) regression (RIF-OLS), a logit regression (RIF-Logit) and a nonparametric logit regression. An estimate of the coefficients of the unconditional quantile regressions, $\hat{\beta}_\tau$, obtained by a simple OLS regression is as follows:

$$\hat{\beta}_\tau = (X'X)^{-1}X'\widehat{RIF}(y; q_\tau, F). \tag{11}$$

The practical problem to solve is that the RIF depends on the marginal density of y. Firpo *et al.* (2009) propose using a kernel estimator for the density and the sample quantile for q_τ so that an estimate of the RIF for each observation is given by

$$\widehat{RIF}(y_i; q_\tau, F) = \hat{q}_\tau + \frac{\tau - \mathbb{1}(y_i \leq \hat{q}_\tau)}{\hat{f}(\hat{q}_\tau)}.$$

Standard deviations of the coefficients are given by the standard errors of the regression.

However, the *RIF* regression models of Firpo *et al.* (2009) present some limitations.

- *First*, if the wage distribution is characterized by a heavy right-hand tail, the kernel density estimation may undersmooth the tail density estimates, leading to unreliable inference for the upper quantile regression coefficients. To overcome this problem, we propose a semiparametric approach to estimate the distribution of log wages using a mixture of normal densities.
- *Second*, the classical *RIF*-OLS estimation is a two stage estimation method. When estimating β_τ, it does not take into account the uncertainty introduced by the use of a point estimate for $f(q_\tau)$ in the first stage. A Bayesian approach should consider the step simultaneously and thus should help us to remove this difficulty.

Fergusson (1983) has shown, but see also Escobar and West (1995), that a mixture of normal densities with a sample determined number of components can approximate any type of densities. Quoting Escobar and West (1995), 'this model produces predictive distributions qualitatively similar to kernel techniques, but catering for differing degree of smoothing across the sample space through the use of possibly differing variances'. This means that an approach using mixtures is as precise as kernel smoothing for the bulk of the density, but manages to model the tails of the distribution in small samples.

Bayesian inference for the RIF regression model

We model the distribution of the observed log wages by a mixture of K normal densities $f(y|\theta)$ indexed by $\theta = (\theta_k)_{k=1,\ldots,K}$, where $\theta_k = (\mu_k, \sigma_k^2, p_k)$, and (μ_k, σ_k^2) are the component-specific mean and variance. If each component is sampled with probability p_k, then the density function $f(y|\theta)$ is written as:

$$f(y|\theta) = \sum_{k=1}^{K} p_k f(y|\theta_k), \tag{12}$$

where

$$f(y|\theta_k) = \frac{1}{\sigma_k\sqrt{2\pi}} \exp\left(-\frac{(y-\mu_k)^2}{2\sigma_k^2}\right).$$

Bayesian inference for mixture of normal densities relies on a rewriting of the likelihood function using a data augmentation representation which leads to a Gibbs sampler. For each observation y_i we associate a missing variable z_i that indicates its component. Formally, this means that we have a hierarchical structure associated with the model:

$$z_i|p \sim \mathrm{M}_k(p_1, \ldots, p_k), \qquad y_i|z_i, \mu, \sigma^2 \sim N(\cdot|\mu_{z_i}, \sigma^2_{z_i}),$$

where $z_i \in \{1,\ldots,k\}$, and $M_k(\cdot)$ represents the multinomial distribution. Details of the approach can be found in, e.g., Robert and Casella (1999) or Frühwirth-Schnatter (2006).

To clearly combine the two steps of the procedure, we reformulate as follows the *RIF* for a quantile regression:

$$RIF(y_i; q_\tau) = y(\theta, \tau) = \hat{q}_\tau + \frac{\tau - \mathbb{1}(y \leq \hat{q}_\tau)}{f(\hat{q}_\tau|\theta)},$$

where $\hat{q}_\tau$ remains the natural estimator of the τ^{th} quantile while θ is explicitly treated as an unknown parameter. The quantile regression model is now expressed conditionally on θ:

$$y(\theta, \tau) = X\beta(\theta, \tau) + \epsilon, \tag{13}$$

where ε is normal with zero mean and variance σ^2. Equation (13) is a conditional linear regression, conditional on the value of θ. In fact, this problem can be treated sequentially. We first estimate the marginal density of y by means of a Gibbs sampler for the given mixture of normal densities in Equation (12). For each draw of θ, noted θ_j, we run the linear regression Equation (13). Marginal moments of β_τ are then obtained by averaging over the draws of θ each obtained value β^j_τ. More precisely, the conditional posterior density of β in Equation (13) is Student[2] with

$$\varphi(\beta|\theta, \tau, y, X) = f_t(\beta|\beta_*(\theta), s_*(\theta), M_*, n). \tag{14}$$

If we suppose a noninformative prior for β and σ^2, the hyperparameters in Equation (14) are given by:

$$\begin{aligned} M_* &= X'X, \\ \beta_*(\theta) &= M_*^{-1} X' y(\theta, \tau), \\ s_*(\theta) &= y(\theta, \tau)'(I_N - X(X'X)^{-1}X')y(\theta, \tau). \end{aligned} \tag{15}$$

Marginal moments are obtained by integrating out θ. This integration can be approximated easily when we have posterior draws of θ:

$$E(\beta|y, \tau) = \int \beta_*(\theta, \tau)\varphi(\theta|y)d\theta \simeq \frac{1}{m}\sum_{j=1}^{m} \beta_*(\theta_j), \tag{16}$$

$$\mathrm{Var}(\beta|y, \tau) = \frac{M_*^{-1}}{n-2}\int s_*(\theta, \tau)\varphi(\theta|y)d\theta \simeq \frac{1}{m(n-2)} M_*^{-1} \sum_{j=1}^{m} s_*(\theta_j). \tag{17}$$

Let us give a brief sketch of the procedure for estimating θ. For each observation y_i of $(y_1,\ldots,y_n)$ from Equation (12), we associate a missing variable z_i that indicates its component of origin. The conditional likelihood function of the sample is

[2] The Student density is noted $f_t(\beta|b, s, M, v) = C^{-1}|M|^{1/2}[s + (\beta - b)'M(\beta - b)]^{-(v+k)/2}$. See Bauwens, Lubrano and Richard (1999, Appendix A) for more details.

$$L(\mu_k, \sigma_k^2|y, z) \propto \sigma_k^{-n_k} \exp -\frac{1}{2\sigma_k^2}(s_k^2(z) + n_k(\mu_k - \bar{y}_k(z))^2), \tag{18}$$

where the sufficient statistics are

$$\bar{y}_k(z) = \frac{1}{n_k}\sum_{i\in Z_k} \log y_i, \qquad s_k^2(z) = \frac{1}{n_k}\sum_{i\in Z_k}(\log y_i - \bar{y})^2, \qquad n_k = \sum \mathbb{1}(z_i = k).$$

We can specify conjugate prior densities for all the parameters with a conditional normal prior for μ_k, an inverted gamma2 prior for σ_k^2 and a Dirichlet prior on p_k. Combining these prior densities with the conditional likelihood function (18), we obtain a conditional Student posterior density for μ_k, and an inverted gamma[3] conditional posterior density for σ^2:

$$\varphi(\mu_k|x, z) \propto f_t(\mu_k|\mu_{*k}, s_{*k}, n_{*k}, v_{*k}) \tag{19}$$

$$\varphi(\sigma_k^2|x, z) \propto f_{i\gamma}(\sigma_k^2|v_{*k}, s_{*k}), \tag{20}$$

where

$$v_{*k} = v_0 + n_k, \qquad s_{*k} = s_0 + s_k^2(z) + \frac{n_0 n_k}{n_0 + n_k}(\mu_0 - \bar{x}_k(z))^2,$$

$$\mu_{*k} = \frac{n_o\mu_0 + n_k\bar{y}_k}{n_{*k}}, \qquad n_{*k} = n_0 + n_k,$$

and where μ_0, n_0, s_0, and v_0 are the hyperparameters of the prior densities for the mixture. We propose the following MCMC algorithm which combines inference for θ and β_τ in a sequential process.

1. Set as starting values $p^{(0)}$, $\mu^{(0)}$, $\sigma^{2^{(0)}}$, the number of draws m and select τ.
2. Compute the τ^{th} quantile $q(\tau)$ of the log wages and $M = (X'X)^{-1}$
3. Begin loop on $j = 1,\ldots,m$
   (a) Begin loop on $k = 1,\ldots,K$
      i. For each observation i, generate $z_i^{(j)}$ from
      $$\mathbb{P}\left(z_i^{(j)} = j|p_k^{(j-1)}, \mu_k^{(j-1)}, \sigma_k^{2(j-1)}, y_i\right) \propto p_k^{(j-1)} f\left(y_i|\mu_k^{(j-1)}, \sigma_k^{2(j-1)}\right)$$
      ii. Compute $n_k^{(j)} = \sum_{i=1}^n \mathbb{1}_{z_i^{(j)}=k}, s_k^{(j)} = \sum_{i=1}^n \mathbb{1}_{z_i^{(j)}=k} y_i$
      iii. Generate $p_k^{(j)}$ from $\mathrm{D}\left(\gamma_1 + n_1^{(j)}, \ldots, \gamma_k + n_k^{(j)}\right)$,
      iv. Generate $\mu_k^{(j)}$ from $\varphi(\mu_k^{(j)}|z^{(j)}, y)$
      v. Generate $\sigma_k^{2(j)}$ from $\varphi(\sigma_k^{2(j)}|y, z^{(j)})$
   (b) End loop on k
   (c) Compute $y(\tau)^{(j)} = \hat{q}(\tau) + \frac{\tau - \mathbb{1}(y \le \hat{q}(\tau))}{\sum_k p_k^{(j)} f(\hat{q}(\tau)|\mu_k^{(j)}, \sigma_k^{2(j)})}$
   (d) Store $\beta_*^{(j)} = MX'y(\tau)^{(j)}$
   (e) Store $s_*^{(j)} = y'(\tau)^{(j)} y(\tau)^{(j)} - y'(\tau)^{(j)} XMX'y(\tau)^{(j)}$

[3] The inverted gamma density is noted $f_{i\gamma}(\sigma^2|v, s) = C^{-1}(\sigma^2)^{-(v+2)/2}\exp(-s/(2\sigma^2))$. See Bauwens *et al.* (1999, Appendix A) for more details.

4. End loop on j
5. Compute the mean of β_*
6. Compute the mean of $s_* \times \frac{M}{n-2}$

As a by-product of this algorithm, we obtain draws from an approximation to the posterior density of θ, $\phi(\theta)$ that will be useful for the derivation in Oaxaca–Blinder decomposition.

Selecting the optimal number of mixture components

The number of components in a mixture is usually determined from the data, using an information criteria. From a Bayesian perspective, the choice of the optimal number of mixture components is based on the maximization of the marginal likelihood (*ml*) obtained by integrating the likelihood function with respect to the prior density (Gelfand and Dey, 1994; Newton and Raftery, 1994; Chib, 1995; Kass and Raftery, 1995). The BIC criterion was devised by Schwarz (1978) as an asymptotic approximation to the log integrated likelihood

$$BIC(k) = \log L(x|\hat{p}_j, \hat{\mu}_j, \hat{\sigma}_j^2) - \frac{\eta_k}{2}\log(n), \tag{21}$$

where η_k is the number of free parameters of the model with k components and the $\hat{p}_j$, $\hat{\mu}_j$, and $\hat{\sigma}_j^2$ are parameter estimates. Using the Gibbs output, we shall find the maximum of the log likelihood over the MCMC draws.

Oaxaca–Blinder decomposition and RIF-OLS

The Oaxaca–Blinder method is very useful for decomposing differences in mean wages between two periods into a wage structure effect and a composition effect. For the unconditional quantile regression, the Oaxaca–Blinder decomposition procedure based on the *RIF* regression model provides a detailed decomposition of the differences in mean wages between two periods (Firpo *et al.*, 2011). If we label A and B as the two different periods, the *RIF* regressions for each period g, (g = A,B) are given by

$$y_g(\theta, \tau) = X_g \beta_g(\theta, \tau) + \epsilon_g, \qquad g = A, B. \tag{22}$$

The differences in mean quantile wages between the two periods are then given by

$$\begin{aligned} \underbrace{E(y_B(\theta,\tau)|X_B) - E(y_A(\theta,\tau)|X_A)}_{\Delta_O(\theta,\tau)} = {} & \underbrace{\bar{X}_B(\beta_B(\theta,\tau) - \beta_A(\theta,\tau))}_{\Delta_\beta(\theta,\tau)} \\ & + \underbrace{(\bar{X}_B - \bar{X}_A)\beta_A(\theta,\tau)}_{\Delta_X(\theta,\tau)}. \end{aligned} \tag{23}$$

as $E(\epsilon_g|X) = 0$ in the *RIF* regression. The first right-hand component, $\Delta_\beta(\theta,\tau)$, is interpreted as the difference in yields of given individual characteristics corresponding to the second period (the wage structure effect). The second right-hand term, $\Delta_X(\theta,\tau)$ is the component associated with differences in the characteristics themselves (the composition effect) as they have evolved between the two periods.

The three quantities in Equation (23) are conditional on θ which has now to be integrated out. Formally,

$$\int \Delta_O(\theta,\tau)\varphi(\theta)d\theta = \bar{X}_B \int [\beta_B(\theta,\tau) - \beta_A(\theta,\tau)]\varphi(\theta)d\theta + (\bar{X}_B - \bar{X}_A)\int \beta_A(\theta,\tau)\varphi(\theta)d\theta. \quad (24)$$

We want now to compute the posterior marginal expectation and posterior marginal variance of the two components of the Oaxaca–Blinder decomposition. With Equation (24), we still do not have an estimator. We can produce an estimator if we replace $\beta_g(\theta,\tau)$ by $\beta^*_g(\theta,\tau)$ in Equation (24), which means replacing the parameter by its conditional posterior expectation which will be then marginalized by integration. This operation is called a Rao-Blackwellization. The marginal expectation of the composition and wage structure effects can now be evaluated in a straightforward way:

$$E[\Delta_\beta(\tau)] = \bar{X}_B\left(\frac{1}{m}\sum_{j=1}^{m}(\beta^B_{*j} - \beta^A_{*j})\right) \quad (25)$$

$$E[\Delta_X(\tau)] = (\bar{X}_B - \bar{X}_A)\left(\frac{1}{m}\sum_{j=1}^{m}\beta^A_{*j}\right), \quad (26)$$

where $\left(\beta^A_{*j};\beta^B_{*j}\right) = \left(\beta^A_*(\tau,\theta^{(j)});\beta^B_*(\tau,\theta^{(j)})\right)$ are the draws of the *RIF* regression coefficients obtained from the Gibbs output $\theta = (\theta^{(j)})^m_{j=1}$. The expectation of the total effect is just the sum of the two components expectations.

Standard errors for the Oaxaca–Blinder decomposition

Most empirical studies which use the Oaxaca–Blinder decomposition procedure do not indicate how standard errors are obtained. As $E(y(\theta,\tau)|X) = \bar{X}'\beta_*(\theta,\tau)$, a well defined approximate estimator for the variance of the conditional mean of the *RIF* is given by:

$$\begin{aligned} \mathrm{V}(E[y(\theta,\tau)|X]) &= \mathrm{V}[\bar{X}'\beta_*(\theta,\tau)] \\ &= \bar{X}'\mathrm{V}[\beta_*(\theta,\tau)]\bar{X}, \end{aligned}$$

as $\bar{X}$ is supposed to be constant (see Jann, 2008 for a classical approach and an alternative derivation when $\bar{X}$ is supposed to be random). As this is a conditional expectation, we have to integrate out θ to obtain the marginal variance of β as given in Equation (17).

Following the lines given in Oaxaca and Ransom (1998), the conditional variances of $\Delta_\beta(\theta,\tau)$ and $\Delta_X(\theta,\tau)$ are easily obtained and when θ is integrated out, we get the following estimates which are transformations of Equation (17):

$$\begin{aligned} V(\bar{X}_B(\beta_*^B - \beta_*^A)) &= \bar{X}_B' V(\beta_*^B - \beta_*^A)\bar{X}_B \\ &= \bar{X}_B'\big(V(\beta_*^B) + V(\beta_*^A)\big)\bar{X}_B \end{aligned} \tag{27}$$

$$V((\bar{X}_B - \bar{X}_A)\beta_*^A) = (\bar{X}_B - \bar{X}_A)' V(\beta_*^A)(\bar{X}_B - \bar{X}_A), \tag{28}$$

provided β_*^B and β_*^A are independent. Standard deviations reported in Tables of the next section are obtained using this method.

Remark: We could have proceeded in another way. Conditionally on a draw of θ, say θ_j, we could compute the hyperparameters in equation (15) and then using equation (14), we could have got m posterior draws of β_τ^g and consequently m draws for $\Delta_O(\tau)$, $\Delta_\beta(\tau)$, and $\Delta_X(\tau)$. Once we have m draws for these three quantities, we can compute the mean and variance of the Oaxaca decomposition. This method is suggested in Radchenko and Yun (2003) in the framework of the usual linear regression (not the quantile regression). Note that this method is less precise than ours as it makes use of less information (no Rao-Blackwellization).

IV Application: Trends in US Wage Structure and Earnings Inequality 1992–2009

Over the past two decades, the United States experienced a sharp rise in wage inequality accompanied by a large increase in wage differentials by skill groups. A large and growing empirical literature attempts at explaining these changes in the US wage structure by using a variety of datasets. As stressed by Firpo, Fortin and Lemieux (2007), these various explanations can all be summarized in terms of the respective contributions of various sets of factors such as education, experience, unions, and gender in a Mincer equation evaluated over different quantiles. In this study, we follow the same route, using a Bayesian approach.

The data

We use the hourly wage data from the Outgoing Rotation Group (ORG) supplements of the Current Population Surveys (CPS)[4] following Lemieux (2006) and Firpo *et al*. (2007). The CPS is the monthly household survey conducted by the Bureau of Labour Statistics to measure labor force participation and employment. A total of 50–60,000 households per month are queried. This is not really a panel survey as households are not followed if they move. The CPS files include the March CPS file and the Outgoing Rotation Group (ORG) files.[5] The ORG CPS files provide a better dataset for measuring changes in hourly wage distribution than the March CPS as they give a better

[4] Atkinson (2008) p. 401) reports different bibliographic sources showing that the ORG data are the most accurate source to study the evolution of the US wage structure.

[5] The ORG files correspond to the set of every household that enters the CPS interviewed each month for 4 consecutive months, and then ignored for 8 months.

representation of the dispersion of wages for each and every hour worked in the labor market, regardless of who is supplying this hour (Bernstein and Mishel, 1997).

We take the monthly earnings files for January 1992 through May 2009. We decide to focus our attention on 3 years (1992, 2001, 2009) to cover the main features of the recent period and their evolution. We use the weekly wage divided by the number of hours worked to get an homogeneous definition of hourly wages.[6] We deflate these wages by the annual average CPI which is, respectively, 140.2, 177.1, and 214.5 for these 3 years. Since January 1992, the CPS has changed the coding scheme of its education attainment question from completed years to degree actually acquired. The new coding scheme details 16 categories for education[7] which include the highest level of school completed or the highest degree received. Kominski and Siegel (1993) show that the new educational attainment item provides more relevant and useful data for current and future analysis. Our education variable will indicate the official number of years needed to reach the acquired education level. It will represent the efficient number of years of schooling.

Most of the studies concerning wage dispersions in the US cover the period 1973–1989 to provide a comparison basis between the different articles. We found marked differences between our sample period 1992–2009 and the previous period 1973–1989. For instance, Melly (2005) indicates that mean and median real wages declined between 1973 and 1989. For the new period, between 1992 and 2009, we have a constant rise of real wage together with a sharp increase in inequality at the end of the period. See Table 1 for detailed figures. This evolution is also depicted in the estimated wage densities. In Figure 1, we display a nonparametric estimate of the wage density. We notice that the distribution of real wages is characterized by a heavy right tail in 2009.

We give in Tables 2 and 3 the descriptive statistics of the covariates that we shall use in the regression model. The median age increases, the third quartile of education level increases also. But the proportion of unionized decreases steadily over the period. The proportion of females remains relatively stable.

The model

The formulation we adopt is a standard Mincer equation:

$$\ln(y_i) = \beta_{0\tau} + \beta_{1\tau}\text{Educ}_i + \beta_{2\tau}\text{Exp}_i + \beta_{3\tau}\text{Exp}_i^2 + \beta_{4\tau}\text{Union}_i + \beta_{5\tau}\text{Fem}_i + \epsilon_{i\tau}, \quad (29)$$

[6] The ORG files are often used because they include a direct observation of the hourly wage, which thus has not to be computed as the ratio between the weekly wage and the number of hours worked. However, many individuals did not answer to that question, so we prefer to compute a ratio to keep the maximum number of observations. And anyway, apart from a few aberrant values, our ratio series gave similar figures as the one given by the hourly series.

[7] The 16 categories include: no diploma; high school graduate; some college but no degree; associate degree in college (occupational or vocational program); associate degree in college (academic program); bachelor's degree, master's degree, professional school degree; and doctorate degree.

Table 1
Hourly real wage dispersion for the US recent period

	1992	2001	2009
$q_{0.10}$	6.90	7.45	8.00
$q_{0.25}$	9.36	10.30	10.00
Mean	18.00	20.05	24.49
Median	14.52	15.53	15.62
$q_{0.75}$	22.02	23.98	26.49
$q_{0.90}$	31.37	36.01	46.12
Gini	0.352	0.369	0.455
N	62107	63409	47837

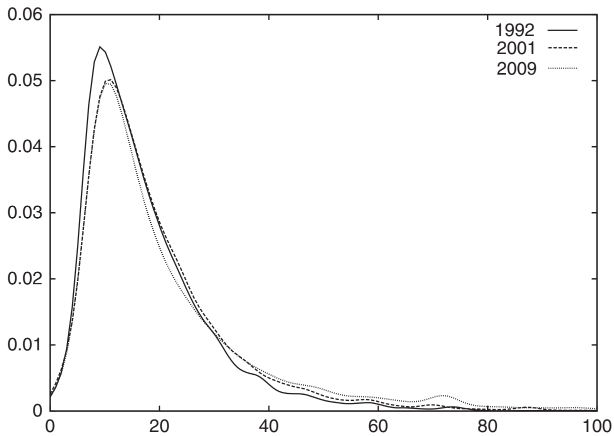


Figure 1. Real wage density estimates.

where (y_i, $i = 1,\ldots,n$) is the hourly real wage for workers. We have introduced education (number of years), experience and its square, the union status, and gender. Potential experience is calculated as the age minus the assigned years of education minus 6, rounded down to the nearest integer value, min(age−education−6,age−18). Education is the official number of years needed to reach the acquired education level.

In Figure 1, we see that it is quite difficult to obtain a smooth estimate for the right tail of the wage-level density with a unique window size. Figure 2 indicates that a nonparametric density estimate of the log wages is also problematic. This lack of smoothness may disturb the classical *RIF*-OLS. The adjusted mixture of normal densities (see also Figure 2) provides of course a much smoother picture, but requires a large number of components as determined by a BIC.

Table 4 gives the value of the BIC to determine the optimal number of mixture components for the 3 years covered the sample. For the years 1992–2001, the BIC is minimized when $k = 6$, whereas for the year 2009 the BIC is minimized

Table 2
Summary statistics for age and education

	Age			Education		
Year	1992	2001	2009	1992	2001	2009
Min	16	16	16	1	1	1
Q1	29	30	30	9	9	9
Med.	40	43	45	9	9	10
Mean	43	45	45	9	10	10
Q3	56	56	58	10	12	13
Max	90	90	85	16	16	16

Table 3
Summary statistics for union and gender

	Union			Female		
Year	1992	2001	2009	1992	2001	2009
Proportion	18.4	15.0	13.7	53.03	52.43	52.34

when $k = 8$. The distribution for the year 2009 requires more components as it is characterized by the heaviest right-hand tail as shown in Figure 1.

Using a noninformative prior for β and σ, we give in Table 5 posterior means and standard deviations for the parameters of the quantile regression. We use 10,000 draws for each year and the same quantiles $\tau =$ 0.10, 0.50, and 0.90. As a point of comparison, we have first estimated this equation using the procedure of Firpo *et al.* (2009) and we reported the results in the same Table 5.

The comparison of the two sets of columns in Table 5 motivates the following comments. *First*, the posterior means are very comparable to the classical estimates in the body of the log wage distribution (10th and 50th percentile). However, there is a difference between the coefficient estimates of the covariates in the right tail (90th percentile) especially for the year 2009 that we can explain by the difference in smoothness between the two different methods for estimating the log wage density. The presence of a fat right tail in the distribution of 2009 might be the main explanation. In fact, the kernel density estimation may undersmooth the tail of the distribution when it is characterized by a heavy tail. As a consequence, the classical *RIF* regression coefficients (90th percentile) are overestimated for the constant term and underestimated for the other coefficients, in 2009. This might have an impact on the results of the Oaxaca–Blinder decomposition as shown in Table 6. *Second*, the posterior standard deviations are most of the time larger than their classical counterpart. In the Bayesian approach, we take into account the uncertainty contained in the first step estimation of the log wage density. This might have consequences on the significance of wage inequality

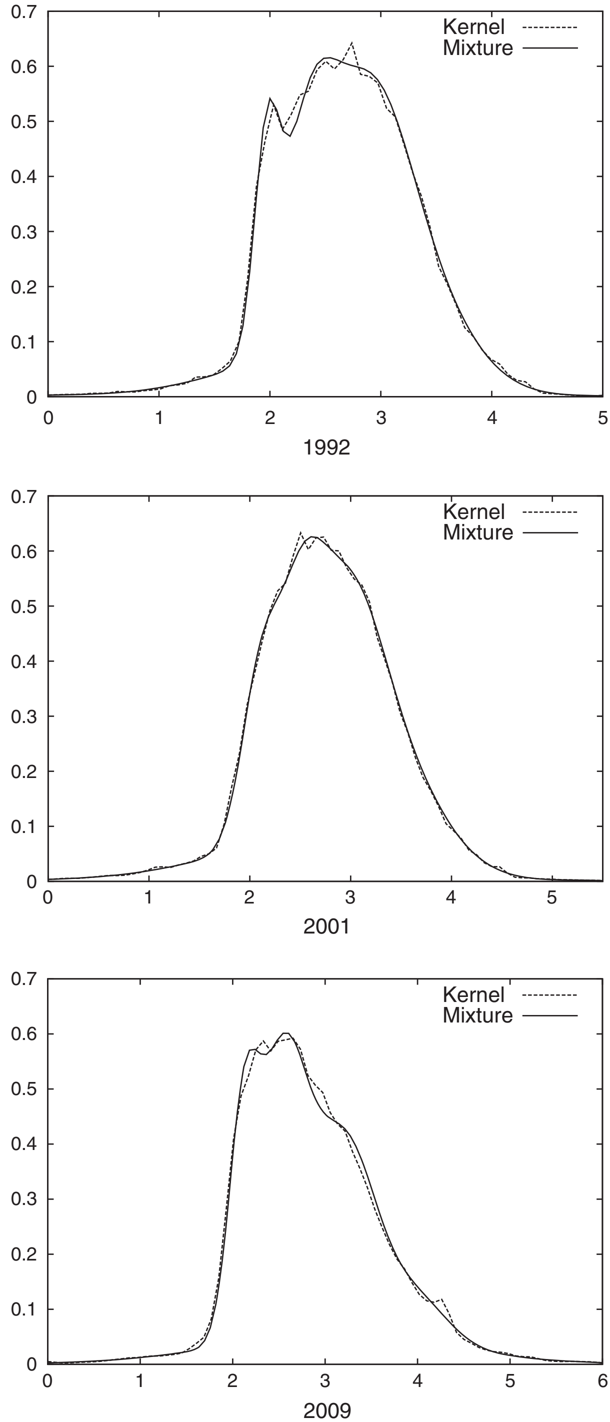


Figure 2. Fitting a mixture of normal densities on real log wages.

Table 4
BIC for selecting the optimal number of mixture components

	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$	$k = 8$	$k = 9$
1992	117740	115835	115416	115316	**115278**	115312	115317	115332
2001	126485	125604	125512	125408	**125346**	125387	125409	125446
2009	106913	103649	102949	102657	102579	102554	**102550**	102595

Bold numbers correspond to the optimal value.

decomposition. However, all the coefficients have rather small standard errors.

Economic interpretation

Let us now detail the economic interpretation of Table 5. Between 1992 and 2009, the return to education has increased in all parts of the distribution. But the yield rose sharply for the median wages (4.6% to 7%)[8] and for higher wages (4.8% to 9%). This provides an explanation of the rise in wage inequality (at constant education composition). The return to experience is much lower than that of education, even if it has risen over the period for all the categories. It is much higher for the first decile and for the median than for the last decile. This should reduce wage inequalities. The evolution of the yield of being member of a union is not uniform over time and over the quantiles. In 1992, it was very profitable for median wages to be a union member with a wage differential of 49%. The yield of being unionized decreased while climbing up the wage ladder. It becomes negligible (3% on average) for high wages. When we now look at the end of the period, the yield of being unionized has decreased for low wages, a fact already noticed in the literature, but has increased slightly for median and high wages. The last covariate concerns gender. Being a woman has always meant having a lower wage. This is especially true here for median and high wages, but not so for low wages. This gender discrimination has risen over the period for all the categories. As a final comment, the constant term for the lowest quantile is traditionally interpreting as measuring the effect of the minimum wage. The minimum wage was raised slightly before 1992 and 2009, but not around 2001. The constant term for 2001 is lower than for 1992, showing the readjustment of the labor market. The rise of the constant term in 2009 reflects nicely the next rise of the minimum wage.

Oaxaca–Blinder decomposition

The results of the Oaxaca–Blinder decomposition are given in Table 6 for both Bayesian and classical estimates to facilitate comparison.

[8] As underlined in Bazen (2011, Table 1.1, p. 21), in a log linear regression, coefficients can be interpreted as percentages only for small values. For higher values, one has to use the formula $\exp(\beta_i) - 1$.

Table 5
Classical and Bayesian RIF regression estimates

	Classical with kernel density estimation			Bayesian with mixture models		
Quantiles	10th	50th	90th	10th	50th	90th
Cst						
1992	1.591	2.058	2.927	1.628	2.049	2.898
	(0.0124)	(0.0143)	(0.0175)	(0.0111)	(0.0145)	(0.0185)
2001	1.540	1.990	2.890	1.559	1.974	2.862
	(0.0158)	(0.0138)	(0.0201)	(0.0151)	(0.0141)	(0.0210)
2009	1.723	1.883	2.774	1.744	1.860	2.701
	(0.0130)	(0.0173)	(0.0335)	(1.744)	(0.0178)	(0.0358)
Educ						
1992	0.021	0.047	0.047	0.019	0.047	0.050
	(0.0010)	(0.0012)	(0.0014)	(0.0009)	(0.0012)	(0.0015)
2001	0.031	0.060	0.065	0.030	0.062	0.067
	(0.0013)	(0.0011)	(0.0017)	(0.0012)	(0.0012)	(0.0017)
2009	0.021	0.065	0.101	0.020	0.067	0.108
	(0.0011)	(0.0014)	(0.0028)	(0.0010)	(0.0015)	(0.0030)
Exp						
1992	0.008	0.010	0.007	0.007	0.010	0.007
	(0.0005)	(0.0006)	(0.0007)	(0.0005)	(0.0006)	(0.0008)
2001	0.010	0.010	0.006	0.010	0.010	0.006
	(0.0007)	(0.0006)	(0.0009)	(0.0006)	(0.0006)	(0.0009)
2009	0.011	0.016	0.010	0.010	0.017	0.011
	(0.0006)	(0.0007)	(0.0014)	(0.0005)	(0.0008)	(0.0017)
Exp^2*100						
1992	−0.009	−0.010	−0.005	−0.008	−0.010	−0.005
	(0.0007)	(0.0009)	(0.0011)	(0.0007)	(0.0009)	(0.0011)
2001	−0.011	−0.009	−0.003	−0.011	−0.009	−0.003
	(0.0010)	(0.0008)	(0.0012)	(0.0009)	(0.0009)	(0.0013)
2009	−0.013	−0.020	−0.012	−0.013	−0.021	−0.013
	(0.0008)	(0.0011)	(0.0022)	(0.0008)	(0.0011)	(0.0023)
Union						
1992	0.171	0.409	0.025	0.152	0.415	0.027
	(0.0071)	(0.0081)	(0.0100)	(0.0063)	(0.0083)	(0.0105)
2001	0.212	0.310	0.016	0.203	0.318	0.017
	(0.0098)	(0.0085)	(0.0125)	(0.0094)	(0.0087)	(0.0130)
2009	0.140	0.394	0.041	0.132	0.405	0.044
	(0.0083)	(0.0110)	(0.0213)	(0.0078)	(0.0113)	(0.0228)
Female						
1992	−0.024	−0.127	−0.117	−0.022	−0.129	−0.124
	(0.0055)	(0.0063)	(0.0077)	(0.0049)	(0.0064)	(0.0817)
2001	−0.065	−0.136	−0.151	−0.063	−0.140	−0.157
	(0.0070)	(0.0061)	(0.0089)	(0.0067)	(0.0062)	(0.0093)
2009	−0.060	−0.160	−0.245	−0.057	−0.164	−0.262
	(0.0057)	(0.0076)	(0.0146)	(0.0054)	(0.0078)	(0.0157)

We used a Gaussian kernel and the method of Silverman to determine the window size to estimate the data density for the classical regression.

Table 6
Oaxaca–Blinder decomposition RIF regression, 1992–2009

	Classical with kernel			Bayesian with mixtures		
	10th	50th	90th	10th	50th	90th
Total effect	0.148	0.070	0.386	0.148	0.069	0.386
	(0.00394)	(0.0049)	(0.00825)	(0.0036)	(0.0050)	(0.0088)
Structure						
Total	0.136	*0.048*	0.348	0.137	*0.048*	0.346
	(0.0356)	(0.0444)	(0.0749)	(0.0328)	(0.0454)	(0.0799)
Cst	0.132	−0.175	−0.153	0.116	−0.189	−0.196
	(0.0180)	(0.0224)	(0.0378)	(0.0166)	(0.0229)	(0.0403)
Educ	−0.002	0.186	0.545	0.009	0.197	0.588
	(0.0149)	(0.0186)	(0.0315)	(0.0137)	(0.0190)	(0.0336)
Exp	0.081	0.175	0.102	0.089	0.184	0.112
	(0.0228)	(0.0285)	(0.0480)	(0.0210)	(0.0291)	(0.0512)
Exp^2	−0.052	−0.119	−0.081	−0.055	−0.123	−0.087
	(0.0134)	(0.0168)	(0.0285)	(0.0124)	(0.0172)	(0.0304)
Union	−0.004	−0.002	0.002	−0.003	−0.001	0.002
	(0.0015)	(0.0019)	(0.0032)	(0.0014)	(0.0019)	(0.0034)
Female	−0.019	−0.017	−0.067	−0.018	−0.018	−0.072
	(0.0041)	(0.0051)	(0.0087)	(0.0038)	(0.0053)	(0.0092)
Composition						
Total	0.012	0.021	0.037	0.0108	0.0214	0.0394
	(0.00100)	(0.00118)	(0.0015)	(0.00092)	(0.00120)	(0.0015)
Educ	0.014	0.032	0.032	0.0126	0.0319	0.0334
	(0.00068)	(0.00078)	(0.0010)	(0.00061)	(0.00080)	(0.0010)
Exp	0.009	0.012	0.008	0.0084	0.012	0.0085
	(0.00062)	(0.00072)	(0.0009)	(0.00056)	(0.00073)	(0.0009)
Exp^2	−0.004	−0.004	−0.002	−0.0032	−0.004	−0.0021
	(0.00031)	(0.00035)	(0.0004)	(0.00027)	(0.00036)	(0.0005)
Union	−0.008	−0.019	−0.001	−0.0072	−0.020	−0.0013
	(0.00033)	(0.00038)	(0.0005)	(0.00030)	(0.00039)	(0.0005)
Female	0.0002	0.001	0.001	0.0002	0.0009	0.0008
	(0.00004)	(0.00004)	(0.0001)	(0.00003)	(0.00004)	(0.0001)

The two methods (classical and Bayesian) give very comparable results for the composition effect. The most important differences appear for the wage structure effect, partly because the two inference methods produce rather different estimates for β_{2009}. However, these difference concern mainly the tails (10th and 90th percentiles). From now on, we shall report results obtained with the Bayesian approach.

Total effects are all significant. We note that there was a large increase of 16% for the first percentile, that the increase is very moderate for median wages (7%), and comparatively huge for the last percentile (47%) over a period of 18 years.

Composition effects represent around 30% of the total effect for the median group, but only around 10% for the lowest and highest groups. Composition effects cannot explain the large increases at both ends of the earning distribution. Nevertheless, we can notice that education represents the major part of the composition effect especially for the highest quantile. The other composition effects play a weaker role, while being still significant. The decline in the unionization rate is significant for all quantiles but cannot be regarded as a main explanation of wage inequality, contrary to what was a convincing explanation in a previous period (see DiNardo *et al.*, 1996): the rates of decline are rather small.

Most of the explanation about the evolution of wages inequality over the period relies on structure effects. We must first notice that the total wage structure effect is not significant for the median quantile, so we shall concentrate on results concerning the two extremes of the distribution. The constant term is only significant for the lowest decile, depicting the influence of the minimum wage, completed by a strong influence of experience, a weaker influence of education. Unionization rate is not significant. The large wage increase in the highest quantile is due to a much higher reward of education (40%), compensated by a slight increase in female discrimination (5%). The other factors are either not significant or have a very small coefficient.

Let us now consider the Oaxaca decomposition computed for the more recent period 2001–2009 covering 9 years to see if there was an acceleration in the trends of wage inequality. The results are reported in Table 7.

For the lowest quantile, the total increase is in line with the total period and significant. The wage structure effect is again not significant. For median wages, there is no significant total change. On the contrary for the highest quantile, the increase is strongly significant and denotes a large acceleration in wage increase corresponding to a large increase in the yield of higher education compensated partly by an increase in gender discrimination.

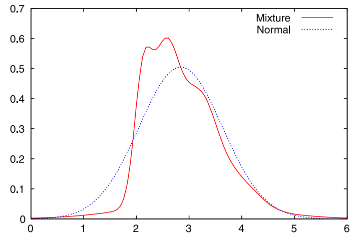
Table 7
Oaxaca–Blinder decomposition: acceleration of wage inequalities? (2001–2009)

	10th percentile	50th percentile	90th percentile
Total effect	0.072 (0.0043)	*0.006 (0.0050)*	0.248 (0.0091)
Wage structure			
Total	*0.064 (0.039)*	*−0.009 (0.045)*	0.227 (0.082)
Cst	0.185 (0.020)	−0.114 (0.023)	−0.161 (0.042)
Educ	−0.104 (0.016)	0.055 (0.019)	0.412 (0.035)
Exp	*0.014 (0.025)*	0.184 (0.029)	0.145 (0.053)
Exp^2	*−0.024 (0.015)*	−0.133 (0.017)	−0.117 (0.031)
Union	−0.010 (0.002)	0.012 (0.002)	*0.004 (0.004)*
Female	*0.003 (0.005)*	−0.013 (0.005)	−0.055 (0.010)
Composition			
Total	0.0080 (0.00042)	0.0157 (0.00039)	0.0203 (0.00058)
Educ	0.0084 (0.00035)	0.0172 (0.00033)	0.0188 (0.00049)
Exp	0.0028 (0.00019)	0.0030 (0.00018)	0.0018 (0.00026)
Exp^2	−0.001 (0.00006)	−0.0006 (0.00005)	−0.0002 (0.00008)
Union	−0.003 (0.00012)	−0.0041 (0.00011)	*−0.0002 (0.00017)*
Female	0.000 (0.00001)	0.0001 (0.00001)	0.0001 (0.00001)

Italics correspond to coefficients for which 0 is contained in an HPD interval.

Table 8
Error committed by using a wrong density estimate (2009 data)

quantile	10th	50th	90th
Cst	11.30	5.65	−10.55
Educ	−55.00	−11.94	25.00
Exp	−23.08	−5.56	11.11
Exp^2	−53.85	−9.52	23.08
Union	−59.09	−11.85	25.00
Female	−57.90	−12.20	25.19



Sensitivity analysis

For estimating the log wage density, we have used a semiparametric approach with a mixture of normal densities where the number of components is sample determined. This is a direct competitor for a fully nonparametric approach and in particular for the classical kernel approach. However, if a distribution is particularly characterized by a fat tail, the semiparametric approach provides a better fit for the tails than the classical kernel density estimation (see Bahadur and Savage, 1956). This particular case is well illustrated in the previous section where we obtained two more components for the year 2009 motivated by the fatter tail of the wage distribution. Of course, when the number of components is arbitrarily fixed, we stay in the parametric situation which also may provide a bad approximation of the tails. As an illustration, if we decide to use a parametric form with a simple normal, we could well simulate the determinant influence of the density estimation procedure for inference on the highest quantiles of the regression. As a sensitivity analysis that we shall now illustrate for the year 2009.

In Table 8, we have displayed the error committed by using an inadequate modeling for the marginal density of the observations, in this case a simple normal density instead of mixture with eight components for the year 2009. The error is indicated in percentage. The parametric model underestimates the 10th quantile on average by 50% while it overestimates the 90th quantile by 25% on average. From the graph given in the right panel of Table 8, we see that the fit of the normal is very bad for the left tail, which explains the huge error committed for the 10th quantile. The fit of the normal for the right tail seems visually much better, but there is still an average error of estimation of 25% for the 90th quantile. This illustrates our point concerning the use of a sample determined mixture of normals. Of course, this result illustrates only a particular case (the year 2009). But it gives a clear indication on the potential errors.

V Conclusion and Summary

In this study, we have proposed a reliable Bayesian inference procedure for the *RIF* regression of Firpo *et al.* (2009) in which we have first estimated the log wage distribution using a mixture of normal densities and then provided marginal posterior densities for the quantile regression parameters. As a by-product, we were able to provide an Oaxaca–Blinder decomposition together with its small sample standard deviations.

Our first empirical results show that in the presence of a heavy right-hand tail in the wage distribution, the kernel estimation might lead to unwanted variability in the *RIF*-OLS method of Firpo *et al.* (2009) for the extreme quantiles. Our parametric approach, using a mixture of normal densities on log wages provides a smoother fit for the tails and provides better estimates for the extreme quantile regression coefficients. Bayesian standard errors are more realistic as they take into account the uncertainty of the first stage when estimating the marginal density.

We have illustrated our method on a Mincer equation for the US labor market covering the period 1992–2009 to analyze the most recent changes in the wage structure and the earnings inequality. Most of the evolutions of the period are concentrated on the extreme quantiles. The median wages do not experience very significant changes. The lowest wages have increased due to the yield of experience while the highest wages have experienced an enormous acceleration mainly due to an increase in the yield of education. The composition effects are rather low.

Writing the *RIF* as a linear conditional expectation provides a simple solution both for the quantile regression and the Oaxaca decomposition. However, it is only a local approximation. Bayesian exploration of this question should be continued using a nonlinear framework, at the cost of making an Oaxaca-like decomposition more difficult.

Acknowledgments

We are grateful to Stephen Bazen for pointing out several references and providing stimulating comments. We would also like to thank Tony Atkinson and Karim Abadir for their helpful suggestions. We thank a coeditor and two anonymous referees for constructive comments. Of course, the usual disclaimers apply.

References

Atkinson, A. B. (2008). *The Changing Distribution of Earnings in OECD Countries*. Oxford: Oxford University Press.

Bahadur, R. R. and Savage, L. J. (1956). The nonexistence of certain statistical procedures in nonparametric problems. *The Annals of Mathematical Statistics*, **27**, 4, pp. 1115–22.

Bauwens, L., Lubrano, M. and Richard, J.-F. (1999). *Bayesian Inference in Dynamic Econometric Models*. Oxford, Oxford University Press.

Bazen, S. (2011). *Econometric Methods for Labour Economics*. Oxford: Oxford University Press.

Benoit, D. F., Al-Hamzawi, R., Yu, K. and Van den Poel, D. (2012). bayesQR: Bayesian quantile regression. http://cran.rproject. org/web/packages/bayesQR/index.html.

Bernstein, J. and Mishel, L. (1997). Has wage inequality stopped growing? *Monthly Labor Review*, **120**, 12, pp. 3–16.

Blinder, A. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human Resources*, **8**, 4, pp. 436–55.

Bound, J. and Johnson, G. (1992). Changes in the structure of wages in the 1980's: an evaluation of alternative explanations. *American Economic Review*, **82**, 3, pp. 371–92.

Buchinsky, M. (1994). Changes in the US wage structure 1963–1987: an application of quantile regression. *Econometrica*, **62**, 2, pp. 405–58.

Chamberlain, G. (1994). Quantile regression censoring and the structure of wages. In C. Sims (ed.), *Advances in Econometrics, vol. 1*. Oxford: Cambridge University Press, pp. 171–209.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, **90**, 432, pp. 1313–21.

DiNardo, J., Fortin, N. and Lemieux, T. (1996). Labor market institutions and the distribution of wages, 1973–1992: a semiparametric approach. *Econometrica*, **64**, 5, pp. 1001–44.

Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, **90**, 430, pp. 577–88.

Fergusson, T. S. (1983). Bayesian density estimation by mixtures of normal distributions. In H. Rizvi and J. Rustagi (eds), *Recent Advances in Statistics*. New-York: Academic Press, pp. 287–302.

Firpo, S., Fortin, N. M. and Lemieux, T. (2007). Decomposing wage distributions using recentered influence function regressions. Technical Working Paper 339, NBER.

Firpo, S., Fortin, N. M. and Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica*, **77**, 3, pp. 953–73.

Firpo, S., Fortin, N. M. and Lemieux, T. (2011). Decomposition methods in economics. In O. Ashenfelter and D. Card (eds), *Handbook of Labor Economics, Part A of Handbook of Labor Economics*, vol. 4. Amsterdam: Elsevier Science Publishers, pp. 1–102.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models. Springer Series in Statistics*. Heidelberg: Springer.

Gelfand, A. E. and Dey, D. K. (1994). Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society, series* B, **56**, 3, pp. 501–14.

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association,* **69**, 346, pp. 383–93.

Jann, B. (2008). A Stata implementation of the Blinder–Oaxaca decomposition. *The Stata Journal,* **8**, 4, pp. 453–79.

Juhn, C., Murphy, K. and Pierce, B. (1993). Wage inequality and the rise in returns to skill. *Journal of Political Economy,* **101**, 3, pp. 410–42.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association,* **90**, 430, pp. 773–95.

Katz, L. F. and Autor, D. H. (1999). Changes in the wage structure and earnings inequality. In O. Ashenfelter and D. Card (eds), *Handbook of Labor Economics*, chapter 26, vol. **3**. Amsterdam: Elsevier, pp. 1463–555.

Katz, L. F. and Murphy, K. M. (1992). Changes in relative wages, 1963–1987: Supply and demand factors. *The Quarterly Journal of Economics*, **107**, pp. 35–78.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica,***46**, 1, pp. 33–50.

Kominski, R. and Siegel, P. (1993). Measuring educational attainment in the current population survey. *Monthly Labor Review*, **116**, 9, pp. 34–8

Kottas, A. and Gelfand, A. E. (2001). Bayesian semiparametric median regression modeling. *Journal of the American Statistical Association,* **96**, 456, pp. 1458–68.

Kottas, A. and Krnjajic, M. (2009). Bayesian semiparametric modelling in quantile regression. *Scandinavian Journal of Statistics,* **36**, 2, pp. 297–319.

Kozumi, H. and Kobayashi, G. (2011). Gibbs sampling methods for Bayesian quantile regression. *Journal of Statistical Computation and Simulation*, **81**, 11, pp. 1565–78.

Lancaster, T. and Jun, S. J. (2010). Bayesian quantile regression methods. *Journal of Applied Econometrics*, **25**, 2, pp. 287–307.

LEMIEUX, T. (2006). Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill? *The American Economic Review*, **96**, 3, pp. 461–98.

MACHADO, J. A. F. and MATA, J. (2001). Earning functions in Portugal 1982–1994: evidence from quantile regressions. *Empirical Economics,* **26**, 1, pp. 115–34.

MACHADO, J. A. F. and MATA, J. (2005). Counterfactual decompositions of changes in wage distributions using quantile regressions. *Journal of Applied Econometrics*, **20**, 4, pp. 445–65.

MELLY, B. (2005). Decomposition of differences in distribution using quantile regression. *Labour Economics*, **12**, 4, pp. 577–90.

MINCER, J. (1993). Human capital, technology, and the wage structure: What do time series show? In J. Mincer (ed.), *Studies in Human Capital*. Brookfield, VT: E. Elgar Publishing.

MURPHY, K. and WELCH, F. (1992). The structure of wages. *Quarterly Journal of Economics*, **107**, 1, pp. 285–326.

NEWTON, M. A. and RAFTERY, A. E. (1994). Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society, series B*, **56**, 1, pp. 3–48.

OAXACA, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, **14**, 3, pp. 693–709.

OAXACA, R. L. and RANSOM, M. (1998). Calculation of approximate variances for wage decomposition differentials. *Journal of Economic and Social Measurement*, **24**, 1, pp. 55–61.

POWELL, D. (2011). Unconditional quantile regression for exogenous or endogenous treatment variables. Technical report, RAND Labor and Population working paper series.

RADCHENKO, S. I. and YUN, M.-S. (2003). A Bayesian approach to decomposing wage differentials. *Economics Letters*, **78**, 3, pp. 431–6.

ROBERT, C. P. and CASELLA, G. (1999). *Monte Carlo Statistical Methods*. Heidelberg: Springer.

ROSEN, S. (1981). The economics of superstars. *The American Economic Review*, **71**, 5, pp. 845–58.

SCHENNACH, S. (2005). Bayesian exponentially tilted empirical likelihood. *Biometrika*., **92**, 1, pp. 31–46.

SCHWARZ, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, **6**, 2, pp. 461–4.

YU, K. and MOYEED, R. (2001). Bayesian quantile regression. *Statistics and Probability Letters*, **54**, 4, pp. 437–47.

Date of receipt of final manuscript: 22 March 2013