

Bayesian Unconditional Quantile Regression: An analysis of recent expansions in wage structure and earnings inequality in the U.S. 1992-2009

Michel Lubrano* Abdoul Aziz Junior Ndoye†

January 2012

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 confidence intervals for the Oaxaca-Blinder decomposition. We analyse the recent changes in the U.S. 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 unionisation rate has a small impact, and that earnings inequality is rising more at the top end of the wage distribution.

Keywords: Bayesian inference, unconditional quantile regression, influence function, Oaxaca-Blinder decomposition.

JEL classification: C1, C11, C13, C14, C21.

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

*AMSE (GREQAM, CNRS), Centre de La Vieille Charité, 2 rue de la Charité, 13002 Marseille, France. Tel: +33 491 14 07 45, Fax:+33 491 90 02 27. email: michel.lubrano@univ-amu.fr

†AMSE (GREQAM), email: abdoul-aziz-junior.NDOYE@univ-amu.fr

1 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. They have been increasingly used in empirical labour 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 for the Bayesian side). Since 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) 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 is 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 et al. (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 paper, we develop a reliable Bayesian inference method for the RIF-regression model of Firpo et al. (2009) in which we estimate parametrically 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. Our first empirical results show that the presence of a heavy right-hand tail in the wage distribution makes unreliable the usual density kernel estimate used in the RIF-OLS method of Firpo et al. (2009) as it leads to an overestimation of the RIF-regression

coefficients for the corresponding quantile. Our parametric approach provides better estimates of the RIF-regression coefficients in the upper tail.

Standard errors obtained by classical RIF-OLS are slightly smaller than our Bayesian standard errors due to the fact our Bayesian approach takes a better account of parameter uncertainty of the density estimation and is pursued so as to propose valid confidence intervals for the Oaxaca-Blinder decomposition.

We illustrate our approach, analysing the recent trends in U.S. wage structure and earnings inequality. The recent rise in earnings dispersion in U.S. is more of a shock as it started quite a long time ago. 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 globalisation 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 labour market institutions, in wage setting norms including the decline in unionisation, 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 (1981)’s 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”. Our paper does not refute these explanations but aims simply at measuring the alternative role of some factors such as union, education, experience and gender for explaining the recent changes in the U.S. wage structure and in earnings inequality.

Most previous studies on changes in the structure of wages in the U.S. have used wage data from the March Current Population Survey (CPS) (see e.g. Buchinsky 1994 and others). We use instead the hourly wage data from the Outgoing Rotation Group (ORG) supplements of the CPS as Lemieux (2006) and Firpo et al. (2007). The ORG CPS files provide a better data set for measuring changes in hourly wage distribution than the March CPS as they give a better representation of the dispersion of wages for each and every hour worked in the labour market, regardless of who

is supplying this hour (Bernstein and Mishel 1997). Since January 1992, the CPS has changed the coding scheme of its education attainment question from completed years to degree acquired. Kominski and Siegel (1993) show that the new educational attainment item provides more relevant and useful data for current and future analysis. Over our period of analysis (1992-2009), unconditional quantile regressions show that earnings inequality is more rising in the top end of the wage distribution while an Oaxaca-Blinder decomposition shows that for a large part (around 77% on average), changes in the U.S. wage structure is explained by the wage structure effect (differences in yields of initial characteristics). If the decline in unionisation has an impact on the recent changes in wage structure, this effect is mainly operating for low wages. For higher wages the increase in returns to education and gender discrimination are the dominant factors.

The paper is organised as follows. In section 2, 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 3, 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 three normal densities. In section 4, 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-Blackwellisation. Section 5 illustrates the approach by using the CPS-ORG sample from 1992 through 2009. Section 6 concludes.

2 Conditional quantile regression models

Consider the usual linear regression model

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

where (y_i, x_i) , $i = 1, 2, \dots, n$ are independent observations, y_i being the response variable and $x_i' = (1, x_{i1}, \dots, x_{ik})$ being the $(k+1)$ known covariates. $\beta' = (\beta_0, \dots, \beta_k)$ represents the $(k+1)$ unknown regression parameters, and ϵ_i , $i = 1, \dots, n$ are the error terms which are supposed independent and identically distributed. The unbiased estimation of β in this regression model requires that $E(\epsilon_i | x_i) = 0$, whatever the distribution of ϵ_i .

A quantile regression model considers a similar linear regression as (2), 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:

$$y_i = x_i' \beta_\tau + \epsilon_{i\tau}, \quad (2)$$

where the parameter to be estimated are the $\beta'_\tau = (\beta_{0\tau}, \dots, \beta_{k\tau})$. 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 $f(\cdot)$ is the density of ϵ , this means that

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

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

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

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

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

where $\mathbf{1}(\cdot)$ is 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.

2.1 Using the asymmetric Laplace distribution

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

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

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

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

provides exactly the same value as that provided by the estimator proposed of Koenker and Bassett (1978) for β_τ . With however the same difficulties as the loss function $\rho_\tau(u)$ (5) is not differentiable at zero. A Bayesian approach does not lead to the same difficulties; the likelihood function (times the prior) has to be integrated and differentiability plays no role in integration.

2.2 Bayesian Inference for conditional quantile regression

To make inferences on the parameter of interest β_τ and σ_τ , given τ and the observations on (X, Y) , one has to specify a prior density for β . The posterior distribution of β_τ , $\pi(\beta|y)$ is proportional to

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

where $L(y|\beta)$ is the likelihood function given in (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. They choose an improper prior and no conjugate prior is available when the model is presented in this form.¹ 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.²

2.3 Oaxaca-Blinder decomposition and quantiles

The popular Oaxaca-Blinder decomposition (Oaxaca 1973; Blinder 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 $y_i = X_i\beta + e_i$, then $E(y_i) =$

¹The posterior density has to be integrated out by a simple random walk MCMC method. The method is available as package `bayesQR` in R. As noted in Kozumi and Kobayashi (2011), the random walk Metropolis maybe difficult to tune, because a different tuning parameter has to be chosen for every value of τ so as to get an acceptance rate of around 25%.

²The 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 hyper-parameters for each quantile. The Gibbs sampler has an important drawback compared to a direct Metropolis approach which is its extreme slowness due to the fact that one has to draw random numbers in an inverted generalised Gaussian for each observation and this is a slow operation which has to be done for each observation separately.

$E(X_i)\beta$. 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

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

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 secondly 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 $E(e_{i\tau}) \neq 0$. 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 of 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. But in the presence of heteroscedasticity, this method produces misleading results. 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 non-parametrically. Machado and Mata (2005) have proposed a simulation method to compute the wage structure sub-components 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 since 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 more consistent for estimating the detailed components of both the wage structure and the composition effects. This is

the method that we shall discuss in the next section and use as a basis for a Bayesian implementation.

3 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 $\nu(F)$, where F is a cumulative distribution function. The IF of the functional ν is defined as

$$IF(y, \nu, F) = \lim_{\epsilon \rightarrow 0} \frac{\nu(F_{\epsilon, \Delta_y}) - \nu(F)}{\epsilon} = \left. \frac{\partial \nu(F_{\epsilon, \Delta_y})}{\partial \epsilon} \right|_{\epsilon=0} \quad (8)$$

where $F_{\epsilon, \Delta_y} = (1 - \epsilon)F + \epsilon\Delta_y$ is a mixture model with a perturbation distribution Δ_y which puts a mass 1 at any point y . The expectation of IF is equal to 0.

Firpo et al. (2009) make use of (8) by considering the distributional statistics $\nu(\cdot)$ as being the quantile function ($\nu(F) = q_\tau$) in order 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 of the distribution of y is given by

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

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

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

with the immediate property that

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

The illuminating idea of Firpo et al. (2009) is to regress the RIF on covariates, so the 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:

$$E[RIF(y, q_\tau|X)] = X\beta. \quad (10)$$

They propose different estimation methods: a standard OLS regression (RIF-OLS), a logit regression (RIF-Logit) and a nonparametric logit regression. The estimates of the coefficients of the unconditional quantile regressions, $\hat{\beta}_\tau$ obtained by a simple Ordinary Least Square (OLS) regression (RIF-OLS) are as follows:

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

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

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

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 characterised by a heavy right-hand tail, the kernel density estimation may under-smooth the tail density estimates, leading to unreliable inference for the upper quantile regression coefficients. To overcome this problem, we propose a semi-parametric approach to estimate the distribution of log-wages using a mixture of normal densities.
- *Second* the classic RIF-OLS estimation does not take into account the uncertainty introduced by the use of a point estimate for $f(q_\tau)$. A Bayesian approach should help to remove this difficulty.

3.1 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 $\theta = (\theta_k)_{k=1, \dots, K}$, 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:

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

where

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

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. Details of the approach can be found in e.g. Robert and Casella (1999) or Frühwirth-Schnatter (2006).

The RIF for a quantile regression can then be reformulated as follows

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

where \hat{q}_τ remains the natural estimator of the τ^{th} quantile while θ is an unknown parameter. The quantile regression model becomes

$$y(\theta, \tau) = X\beta(\theta, \tau) + \epsilon. \quad (13)$$

where ϵ is now normal with zero mean and variance σ^2 . This 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 the given mixture of normal densities in (12). Given this estimation, we run the linear regression (13). More precisely, we first derive the posterior density of θ by a Gibbs sampler, store the posterior draws of θ and then treat model (13) given the former Gibbs output. Marginal moments of β are obtained by averaging over the draws of θ .

The conditional posterior density of β in (13) is Student with

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

where, if we suppose a non-informative prior for β and σ^2

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

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

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

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

Let us give a brief sketch of the procedure for estimating θ . The Bayesian approach for estimating mixture relies on a data augmentation representation. For each observation y_i of (y_1, \dots, y_n) from (12), we associate a missing variable z_i that indicates its component of origin. The conditional likelihood function of the sample is

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

where the sufficient statistics are

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

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 in (18), we obtain a conditional Student posterior density for μ_k , and an inverse gamma conditional posterior density for σ^2

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

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

where

$$\begin{aligned} \nu_{*k} &= \nu_0 + n_k, & s_{*k} &= s_0 + s_k^2(z) + \frac{n_0 n_k}{n_0 + n_k} (\mu_0 - \bar{x}_k(z))^2, \\ \mu_{*k} &= \frac{n_0 \mu_0 + n_k \bar{y}_k}{n_{*k}}, & n_{*k} &= n_0 + n_k, \end{aligned}$$

and where μ_0 , n_0 , s_0 and ν_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 $p^{(0)}$, $\mu^{(0)}$, $\sigma^{2^{(0)}}$, 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 $j = 1, \dots, m$
 - (a) Begin loop on $k = 1, \dots, K$
 - i. Generate $z_i^{(j)}$ from

$$\mathbb{P} \left(z_i^{(j)} = j | p_k^{(j-1)}, \mu_k^{(j-1)}, \sigma_k^{2^{(j-1)}}, y_i \right) \propto p_k^{(j-1)} f \left(y_i | \mu_k^{(j-1)}, \sigma_k^{2^{(j-1)}} \right)$$

for each observation i

- ii. Compute $n_k^{(j)} = \sum_{i=1}^n \mathbb{I}_{z_i^{(j)}=j}$, $s_k^{(j)} = \sum_{i=1}^n \mathbb{I}_{z_i^{(j)}=j} y_i$
 - iii. Generate $p_k^{(j)}$ from $D\left(\gamma_1 + n_1^{(j)}, \dots, \gamma_k + n_k^{(j)}\right)$,
 - 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 $y(\tau)^{(j)} = \hat{q}(\tau) + \frac{\tau - \mathbf{1}(y \leq \hat{q}(\tau))}{\sum_k p_k^{(j)} f(\hat{q}(\tau) | \mu_k^{(j)}, \sigma_k^{2(j)})}$
 - (d) Store $\beta_*^{(j)} = M X y(\tau)^{(j)}$
 - (e) Store $s_*^{(j)} = y(\tau)^{(j)'} y(\tau)^{(j)} - X y(\tau)^{(j)'} M X y(\tau)^{(j)}$
- 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 , $\varphi(\theta)$ that will be useful for the derivation of Oaxaca-Blinder decomposition.

3.2 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 by A and B the two different groups, the RIF-regressions for each group g , ($g = A, B$) are given by

$$y_g(\theta, \tau) = X_g \beta_g(\theta, \tau) + \epsilon_g, \quad g = A, B. \quad (21)$$

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

$$\underbrace{\mathbb{E}(y_B(\theta, \tau) | X_B) - \mathbb{E}(y_A(\theta, \tau) | X_A)}_{\Delta_O(\theta, \tau)} = \underbrace{\bar{X}_B(\beta_B(\theta, \tau) - \beta_A(\theta, \tau))}_{\Delta_\beta(\theta, \tau)} + \underbrace{(\bar{X}_B - \bar{X}_A)\beta_A(\theta, \tau)}_{\Delta_X(\theta, \tau)}. \quad (22)$$

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 (22) are conditional on θ which has now to be integrated out. Formally,

$$\begin{aligned} \int \Delta_O(\theta, \tau)\varphi(\theta)d\theta &= \bar{X}_B \int [\beta_B(\theta, \tau) - \beta_A(\theta, \tau)]\varphi(\theta)d\theta \\ &+ (\bar{X}_B - \bar{X}_A) \int \beta_A(\theta, \tau)\varphi(\theta)d\theta. \end{aligned} \quad (23)$$

We want to compute the posterior marginal expectation and posterior marginal variance of the two components of the Oaxaca-Blinder decomposition. With (23), 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 (23), which means replacing the parameter by its posterior conditional expectation. Following this way, the marginal expectation of the composition and wage structure effects can be evaluated in a straightforward way:

$$E[\Delta_\beta(\tau)] = \bar{X}_B \left(\frac{1}{m} \sum_{j=1}^m (\beta_{*j}^B - \beta_{*j}^A) \right) \quad (24)$$

$$E[\Delta_X(\tau)] = (\bar{X}_B - \bar{X}_A) \left(\frac{1}{m} \sum_{j=1}^m \beta_{*j}^A \right), \quad (25)$$

where $(\beta_{*j}^A; \beta_{*j}^B) = (\beta_{*j}^A(\tau, \theta^{(j)}); \beta_{*j}^B(\tau, \theta^{(j)}))$ are the draws of the RIF-regression coefficients obtained from the Gibbs output $\theta = (\theta^{(j)})_{j=1}^m$. The expectation of the total effect is just the sum of the two components expectations.

Remark:

We could also proceed in another way. Conditionally on a draw of θ , say θ_j , we can compute the hyperparameters in (15) and then using (14), we can get m posterior draws of β_g . As consequently m draws for $\Delta_O(\tau)$, $\Delta_\beta(\tau)$ and $\Delta_X(\tau)$. Once we have these m draws, we can compute the mean and variance of the Oaxaca decomposition. See Radchenko and Yun (2003) for a similar implementation in the framework of the usual linear regression. Note that the first method should produce more precise results as it corresponds to a Rao-Blackwellisation.

3.3 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 variance estimator for the conditional mean of the RIF is given by:

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

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 (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 (17):

$$\begin{aligned} V(\bar{X}_B(\beta_*^B - \beta_*^A)) &= \bar{X}'_B V(\beta_*^B - \beta_*^A) \bar{X}_B \\ &= \bar{X}'_B (V(\beta_*^B) + V(\beta_*^A)) \bar{X}_B \end{aligned} \quad (26)$$

$$V((\bar{X}_B - \bar{X}_A)\beta_*^A) = (\bar{X}_B - \bar{X}_A)' V(\beta_*^A) (\bar{X}_B - \bar{X}_A), \quad (27)$$

provided β_*^B and β_*^A are independent. Standard deviations reported in Tables of section 4.4 are obtained using this method.

4 Application: Trends in U.S. wage structure and earnings inequality 1992-2009

Over the past two decades, the U.S. experienced a sharp rise in wage inequality accompanied by large increase in wage differentials by skill groups. A large and growing empirical literature attempts at explaining the changes in the U.S. wage structure by using a variety of data sets. As stressed by Firpo et al. (2007), these various explanations can all be summarised in terms of the respective contributions of various sets of factors such as education, experience, unions and gender. This paper illustrates the approach developed above by measuring the contributions of some factors and explaining the recent changes in the U.S. wage structure and earnings inequality.

4.1 The data

This paper uses the Current Population Surveys (CPS),³ Outgoing Rotation Groups (ORG).⁴ We take the monthly earnings files for January 1992 through May 2009. We decide to focus our attention on three 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 in order to get an homogeneous definition of hourly wages.⁵ We deflate these wages by the annual average CPI which is respectively 140.2, 177.1 and 214.5 for these three 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 education⁶ which include the highest level of school completed or the highest degree received. 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 U.S. cover the period 1973-1989 in order to provide a comparison basis between the different papers. 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 non-parametric estimate of the wage density. We notice that the distribution of real wages is characterised by a heavy right tail in 2009.

³The CPS is the monthly household survey conducted by the Bureau of Labor Statistics to measure labor force participation and employment. 50-60,000 households per month are queried. This is not really a panel survey since households are not followed if they move. They include the March CPS file and the Outgoing Rotation Group (ORG) files.

⁴The 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.

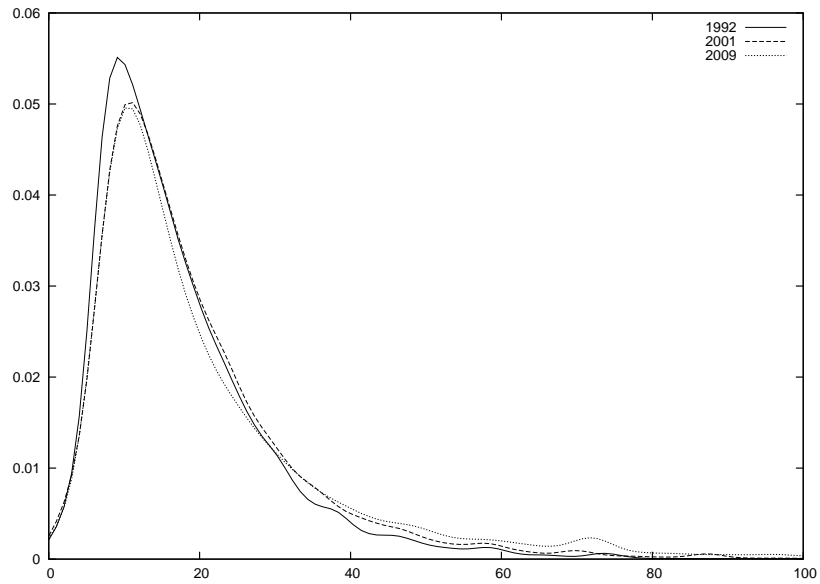
⁵The ORG files are often used because they include a direct observation of the hourly wage, which thus has not to be computed as the ratio between the weekly wage and the number of worked hours. However, many individuals did not answer to that question, so we prefer to compute a ratio in order to keep 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.

⁶The 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 U.S. 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



4.2 The model

The formulation we adopt is a standard Mincer equation:

$$\ln(y_i) = \beta_{0\tau} + \beta_{1\tau}Educ_i + \beta_{2\tau}Exp_i + \beta_{3\tau}Exp_i^2 + \beta_{4\tau}Union_i + \beta_{5\tau}Fem_i + \epsilon_{i\tau}, \quad (28)$$

where $(y_i, \quad i = 1, \dots, n)$ 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.

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 Table 2. As already explained, this estimator requires the use of a non-parametric estimation of the data density. With Figure 1, we see that it is quite difficult to obtain a smooth estimate for the right tail with a unique window size. Figure 2 indicates that a non-parametric density estimate of the log wages is also problematic. This lack of smoothness may disturb the classical RIF-OLS. The adjusted mixture of three normal densities (see also Figure 2) provides of course a much smoother picture. We now turn to Bayesian inference results, using a non informative prior for β and σ . Posterior means and standard deviations are reproduced in Table 3, using 10000 draws for each year and the same quantiles $\tau = 0.10, 0.50$ and 0.90 .

The comparison of Tables 2 and 3 motivates the following comments. First, the posterior means are very comparable to the classical estimates in the body of the log wage distribution (10^{th} and 50^{th} quantiles). However, there is a difference between the coefficient estimates of the covariates in the right tail (90^{th} quantile) for the years 2001, 2009 (but not for 1992) that we can explain by the difference in smoothness between the two different methods for estimating the log wage densities. 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 characterised by a heavy tail. This imply that the classical RIF-regression coefficients are overestimated in the upper tails (90^{th} quantile) of the wage distribution in 2009. This might have an impact on the results of the Oaxaca-Blinder decomposition using classical RIF-regression method of Firpo et al. (2009) (see differences in Tables 4 and 6).

Second, the posterior standard deviation 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 decomposition. Nevertheless, all the coefficients are well estimated with rather small standard errors.

Table 2: Classical unconditional Quantile regression coefficients
with kernel density estimation on log wages

	10th percentile	50th percentile	90th percentile
Cst			
1992	1.591 (0.0124)	2.058 (0.0143)	2.927 (0.0175)
2001	1.540 (0.0158)	1.990 (0.0138)	2.890 (0.0201)
2009	1.723 (0.0130)	1.883 (0.0173)	2.774 (0.0335)
Educ			
1992	0.021 (0.0010)	0.047 (0.0012)	0.047 (0.0014)
2001	0.031 (0.0013)	0.060 (0.0011)	0.065 (0.0017)
2009	0.021 (0.0011)	0.065 (0.0014)	0.101 (0.0028)
Exp			
1992	0.008 (0.0005)	0.010 (0.0006)	0.007 (0.0007)
2001	0.010 (0.0007)	0.010 (0.0006)	0.006 (0.0009)
2009	0.011 (0.0006)	0.016 (0.0007)	0.010 (0.0014)
Exp² * 100			
1992	-0.009 (0.0007)	-0.010 (0.0009)	-0.005 (0.0011)
2001	-0.011 (0.0010)	-0.009 (0.0008)	-0.003 (0.0012)
2009	-0.013 (0.0008)	-0.020 (0.0011)	-0.012 (0.0022)
Union			
1992	0.171 (0.0071)	0.409 (0.0081)	0.025 (0.0100)
2001	0.212 (0.0098)	0.310 (0.0085)	0.016 (0.0125)
2009	0.140 (0.0083)	0.394 (0.0110)	0.041 (0.0213)
Female			
1992	-0.024 (0.0055)	-0.127 (0.0063)	-0.117 (0.0077)
2001	-0.065 (0.0070)	-0.136 (0.0061)	-0.151 (0.0089)
2009	-0.060 (0.0057)	-0.160 (0.0076)	-0.245 (0.0146)

Figure 2: Fitting a mixture of three normal densities on real log-wages

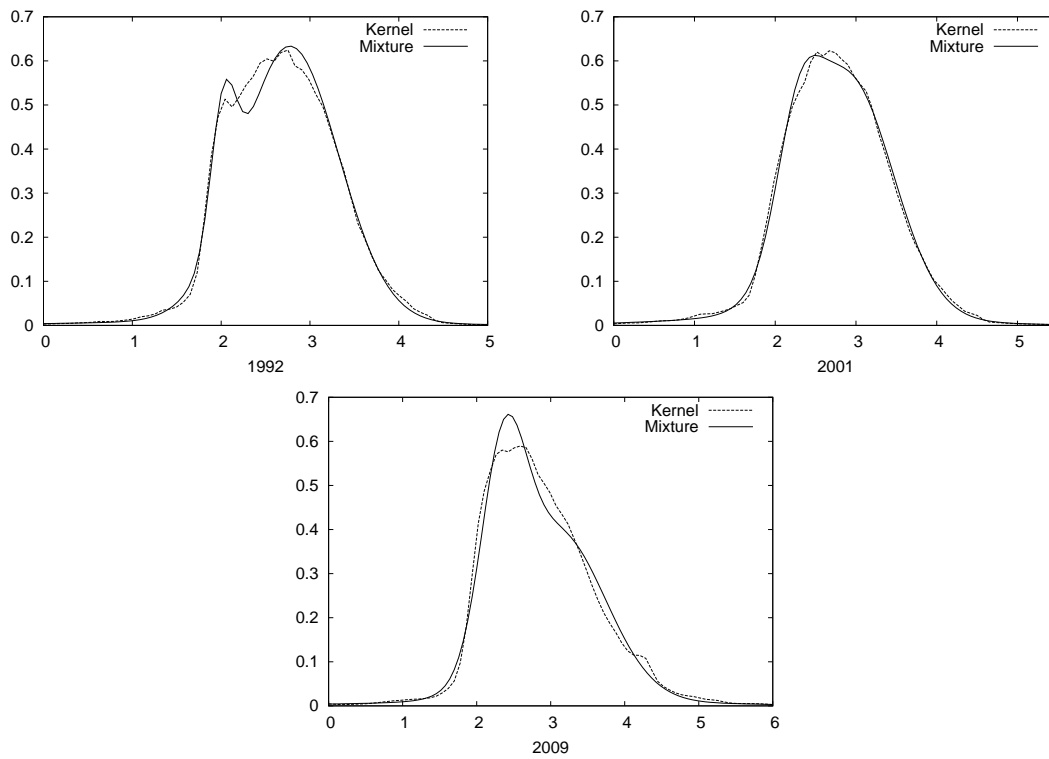


Table 3: Bayesian unconditional Quantile regression coefficients
with a mixture of normal densities on log wages

	10th percentile	50th percentile	90th percentile
Cst			
1992	1.592 (0.0124)	2.074 (0.0139)	2.917 (0.0178)
2001	1.511 (0.0167)	1.952 (0.0145)	2.911 (0.0196)
2009	1.659 (0.0154)	1.803 (0.0189)	2.900 (0.0295)
Educ			
1992	0.021 (0.0010)	0.046 (0.0011)	0.048 (0.0014)
2001	0.033 (0.0014)	0.063 (0.0012)	0.063 (0.0016)
2009	0.024 (0.0013)	0.071 (0.0016)	0.089 (0.0025)
Exp			
1992	0.008 (0.00053)	0.010 (0.00060)	0.007 (0.00077)
2001	0.011 (0.00073)	0.011 (0.00063)	0.006 (0.00085)
2009	0.013 (0.00067)	0.018 (0.00082)	0.009 (0.00129)
Exp² * 100			
1992	-0.0090 (0.00077)	-0.0097 (0.00087)	-0.0052 (0.00111)
2001	-0.0118 (0.00104)	-0.0095 (0.00090)	-0.0026 (0.00122)
2009	-0.0159 (0.00100)	-0.0220 (0.00123)	-0.0106 (0.00191)
Union			
1992	0.170 (0.0070)	0.399 (0.0079)	0.026 (0.0102)
2001	0.225 (0.0104)	0.326 (0.0089)	0.016 (0.0121)
2009	0.165 (0.0098)	0.431 (0.0120)	0.036 (0.0188)
Female			
1992	-0.024 (0.0055)	-0.124 (0.0061)	-0.119 (0.0079)
2001	-0.069 (0.0074)	-0.143 (0.0064)	-0.146 (0.0086)
2009	-0.071 (0.0067)	-0.175 (0.0083)	-0.216 (0.0129)

4.3 Economic interpretation

Let us now detail the economic interpretation of Table 3. 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%)⁷ and for higher wages (4.8% to 9%). This provides an explanation for 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 paradoxical. In 1992, this was very profitable for median wages to be a union member with a wage differential of 49%. The yield of being unionised 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 unionised 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, but this only in terms of intercept earnings. 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 labour market. The rise of the constant term in 2009 reflect nicely the next rise of the minimum wage.

4.4 Oaxaca-Blinder decomposition

The results of the Oaxaca-Blinder decomposition are given in Table 4 for the Bayesian approach and in Table 6 for the classical estimates as a point of comparison. These estimates are very comparable for the composition effect because β_{1992} is roughly the same with classical and Bayesian RIF. The most important change is concentrated on the wage structure effect because β_{2009} is much different with the two estimation methods, but these difference are concentrated in the tails (10th and 90th percentiles). From now on, we shall report only the Bayesian results.

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

⁷As 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 4: Oaxaca Blinder decomposition using Bayesian RIF-regression with a mixture of normal densities on log wages, CPS ORG 1992 – 2009

	10th percentile	50th percentile	90th percentile
Total effect	0.1485 (0.0043)	0.0697 (0.0051)	0.3856 (0.0075)
Wage structure			
Total	0.136 (0.0391)	<i>0.049 (0.0465)</i>	0.347 (0.0683)
Cst	0.067 (0.0198)	-0.271 (0.0235)	<i>-0.017 (0.0345)</i>
Educ	0.036 (0.0164)	0.258 (0.0195)	0.4149 (0.0287)
Exp	0.139 (0.0251)	0.227 (0.0298)	0.0621 (0.0438)
Exp ²	-0.081 (0.0148)	-0.143 (0.0176)	-0.0632 (0.0259)
Union	<i>-0.0006(0.00165)</i>	0.0044 (0.00197)	<i>0.0014(0.00292)</i>
Female	-0.024 (0.0045)	-0.026 (0.0054)	-0.050 (0.0079)
Composition			
Total	0.0121 (0.00103)	0.0206 (0.00115)	0.0380 (0.00148)
Educ	0.0140 (0.00068)	0.0307 (0.00076)	0.0322 (0.00098)
Exp	0.0094 (0.00062)	0.0116 (0.00070)	0.0082 (0.00089)
Exp ²	-0.0036 (0.00031)	-0.0038 (0.00034)	-0.0020 (0.00048)
Union	-0.0080 (0.00033)	-0.0188 (0.00037)	-0.0012 (0.00048)
Female	0.00017 (0.000038)	0.00085 (0.000042)	0.00082 (0.000054)

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

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 end 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 plays a weaker role, while being still significant. The decline in the unionisation rate is significant for all quantiles but cannot be regarded as a main explanation for wage inequality, contrary to what was a convincing explanation in a previous period (see DiNardo et al. 1996) because the rates of decline are rather small.

Most of the explanation about the evolution of wages inequality 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 extreme 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. Unionisation rate is not significant. The large wage

increase in the highest quantile is due to a much higher reward of education (40%), compensated by an 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 over the more recent period 2001-2009 of 9 years in order to see if there was an acceleration in the trends of wage inequality. The results are reported in Table 5.

Table 5: Oaxaca Blinder decomposition using Bayesian RIF-regression with a mixture of normal densities on log wages, CPS ORG 2001 – 2009

	10th percentile	50th percentile	90th percentile
Mean total	0.0718 (0.00499)	<i>0.0065 (0.00521)</i>	0.247 (0.00774)
Wage structure			
Total	<i>0.0630 (0.0450)</i>	<i>-0.0095 (0.0472)</i>	0.2286 (0.0702)
Cst	0.1487 (0.0227)	-0.1484 (0.0238)	<i>-0.0111 (0.0354)</i>
Edu	-0.0872 (0.0189)	0.0801 (0.0198)	0.2666 (0.0295)
Exp	0.0587 (0.0208)	0.2068 (0.0289)	0.0999 (0.0449)
Exp ²	-0.0483 (0.0170)	-0.1458 (0.0179)	-0.0931 (0.0266)
Union	-0.0081 (0.00195)	0.0143 (0.00205)	<i>0.0028 (0.00306)</i>
Female	<i>-0.00091 (0.00525)</i>	-0.0166 (0.00548)	-0.0365 (0.00813)
Composition			
Total	0.0088 (0.00046)	0.0161 (0.00040)	0.0189 (0.00054)
Educ	0.0093 (0.00039)	0.0177 (0.00033)	0.0175 (0.00045)
Exp	0.0031 (0.00021)	0.0031 (0.00018)	0.0016 (0.00024)
Exp ²	-0.00073 (0.000065)	-0.00059 (0.000056)	-0.00016 (0.000076)
Union	-0.00294 (0.00013)	-0.00425 (0.00012)	<i>-0.00021 (0.00016)</i>
Female	0.000059 (0.000006)	0.00012 (0.000005)	0.00012 (0.000007)

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

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 denote 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 6: Oaxaca Blinder decomposition using Classical RIF-regression with kernel density estimation on log wages, CPS ORG 1992 – 2009

	10th percentile	50th percentile	90th percentile
Total effect	0.1484 (0.00394)	0.0695 (0.00491)	0.3857 (0.00825)
Wage structure			
Total	0.136 (0.0356)	<i>0.048 (0.0444)</i>	0.348 (0.0749)
Cst	0.132 (0.0180)	-0.175 (0.0224)	-0.153 (0.0378)
Educ	-0.002 (0.0149)	0.186 (0.0186)	0.545 (0.0315)
Exp	0.081 (0.0228)	0.175 (0.0285)	0.102 (0.0480)
Exp ²	-0.052 (0.0134)	-0.119 (0.0168)	-0.081 (0.0285)
Union	-0.004 (0.0015)	-0.002 (0.0019)	0.002 (0.0032)
Female	-0.019 (0.0041)	-0.017 (0.0051)	-0.067 (0.0087)
Composition			
Total	0.0121 (0.0010)	0.0211 (0.00118)	0.0372 (0.00145)
Educ	0.0141 (0.00068)	0.0315 (0.00078)	0.0316 (0.00096)
Exp	0.0094 (0.00062)	0.0120 (0.00072)	0.0081 (0.00088)
Exp ²	-0.0036 (0.00031)	-0.0040 (0.00035)	-0.0020 (0.00043)
Union	-0.0080 (0.00033)	-0.0193 (0.00038)	-0.0012 (0.00047)
Female	0.00017 (0.000037)	0.00088 (0.000043)	0.00080 (0.000053)

5 Conclusion and summary

In this paper, we have proposed a reliable Bayesian inference 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 standard deviations.

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

We have illustrated our method on a Mincer equation for the U.S. covering the period 1992-2009 in order to analyse the most recent changes in the wage structure and the earnings inequality. Most of the evolution 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 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 non-linear framework, at the cost of making an Oaxaca-like decomposition more difficult.

References

- Atkinson, A. B. (2008). *The Changing Distribution of Earnings in OECD Countries*. Oxford University Press, Oxford.
- Bazen, S. (2011). *Econometric Methods for Labour Economics*. Oxford University Press, Oxford.
- Bernstein, J. and Mishel, L. (1997). Has wage inequality stopped growing? *Monthly Labor Review*, 120:3–16.
- Blinder, A. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human Resources*, 8:436–455.
- 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):371–92.
- Buchinsky, M. (1994). Changes in the U.S. wage structure 1963-1987: An application of quantile regression. *Econometrica*, 62:405–458.
- Chamberlain, G. (1994). Quantile regression censoring and the structure of wages. In Sims, C., editor, *Advances in Econometrics*. Cambridge University Press, Oxford.
- DiNardo, J., Fortin, N., and Lemieux, T. (1996). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica*, 64(5):1001–1044.
- 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):953–973.
- Firpo, S., Fortin, N. M., and Lemieux, T. (2011). Decomposition methods in economics. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 4, Part A of *Handbook of Labor Economics*, pages 1–102. Elsevier Science Publishers.
- Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer Series in Statistics. Springer, Heidelberg.
- Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 69(346):383–393.
- Jann, B. (2008). A Stata implementation of the Blinder-Oaxaca decomposition. *The Stata Journal*, 8(4):453–479.
- Juhn, C., Murphy, K., and Pierce, B. (1993). Wage inequality and the rise in returns to skill. *The Journal of Political Economy*, 101:410–442.
- Katz, L. F. and Autor, D. H. (1999). Changes in the wage structure and earnings inequality. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 3 of *Handbook of Labor Economics*, chapter 26, pages 1463–1555. Elsevier.
- Katz, L. F. and Murphy, K. M. (1992). Changes in relative wages, 1963–1987: Supply and demand factors. *The Quarterly Journal of Economics*, 107:35–78.
- Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46:33–50.
- Kominski, R. and Siegel, P. (1993). Measuring educational attainment in the current population survey. *Monthly Labor Review*, 116(9):34–38.
- Kozumi, H. and Kobayashi, G. (2011). Gibbs sampling methods for Bayesian quantile regression. *Journal of Statistical Computation and Simulation*, 81(11):1565–1578.
- Lemieux, T. (2006). Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill? *The American Economic Review*, 96(3):461–498.
- Machado, J. A. F. and Mata, J. (2001). Earning functions in Portugal 1982–1994: Evidence from quantile regressions. *Empirical Economics*, 26(1):115–134.

- Machado, J. A. F. and Mata, J. (2005). Counterfactual decompositions of changes in wage distributions using quantile regressions. *Journal of Applied Econometrics*, 20(4):445–465.
- Melly, B. (2005). Decomposition of differences in distribution using quantile regression. *Labour Economics*, 12(4):577–590.
- Mincer, J. (1993). Human capital, technology, and the wage structure: What do time series show? In Mincer, J., editor, *Studies in Human Capital*. E. Elgar Publishing, Brookfield, VT.
- Murphy, K. and Welch, F. (1992). The structure of wages. *Quarterly Journal of Economics*, 107:285–326.
- Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, 14:693–709.
- Oaxaca, R. L. and Ransom, M. (1998). Calculation of approximate variances for wage decomposition differentials. *Journal of Economic and Social Measurement*, 24:55–61.
- 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):431–436.
- Robert, C. P. and Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer, Heidelberg.
- Rosen, S. (1981). The economics of superstars. *The American Economic Review*, 71(5):845–858.
- Yu, K. and Moyeed, R. (2001). Bayesian quantile regression. *Statistics and Probability Letters*, 54:437–47.