The Racial Wage Gap and the Great Recession

Maxime Gueuder[*]

August 2017

Abstract

I study the evolution of the difference of median log-annual earnings between Blacks and Whites in the US, using decennial Census data from 1960 to 2010 and the American Community Survey from 2006 to 2015. I focus on the median log-wage of men between 25 and 55. I find that 1) the unconditional median racial log-wage gap increases from 0.55 before the crisis to a peak of 0.75 in 2010 and 2011, which is higher that its value pre-*civil rights act*, 2) controlling for age and education, the conditional racial wage gap increases from 0.42 before to crisis to 0.59 afterwards, 3) the unexplained share of the unconditional racial wage gap increases of 3 points between 2006 and 2012, but remains lower than before the *Civil Rights Act*, 4) adding State fixed-effects and controlling for occupation changes the level of both the conditional and unconditional racial wage-gap, but the dynamics of the latter does not change through specifications.

---

[*]Aix-Marseille School of Economics, maxime.gueuder@univ-amu.fr

1 Introduction

Does a macroeconomic shock as large as the Great Recession have different effects on the labour market depending on workers' characteristics? In particular, has the racial wage gap - the gap of wages offered to Blacks and Whites in the US - widened during the Great Recession?

I track the evolution of the racial wage gap from 2006 to 2015, and use decennial census from 1960 to 2000 as reference points. Naive estimators of this wage gap, computed using the Current Population Survey, do not take into account selective withdrawal from the labour market by incarceration into correctional institution. I propose an estimator of the racial wage gap which corrects for this bias, and control this difference in wages by age and educational achievement. I find that both unconditional and conditional racial wage gaps significantly increase between 2007 and 2010, to a value higher than the one computed using the 1960 Census. I decompose this difference in median log-wages into wage structure and composition effects, to find that the wage structure effect significantly increased since 2006, to reach an all-time high in 2012.

Literature review Blacks and Whites are not equal toward the rise of mass incarceration [Miller (2011)]. As inmates are drawn out of the general population, they are also drawn out of the working population and the population of interest of many surveys, including the *Current Population Survey*, the main source of statistics of the labour force. Chandra [Chandra (2003), Chandra (2000)] first took into account this withdrawal from the labour market using Census data from 1960 to 2000, to assess whether the convergence of wages between Blacks and Whites held with a larger population than just individuals on the labour market. The present paper continues Chandra's work, with two improvements: I use the more recent and yearly *American Community Survey* from IPUMS [Ruggles, Genadek, Goeken, Grover, and Sobek (2015)], which includes population in group quarters in its sample. As Winters and Hirsch (2012), I use a linear regression framework to isolate the ethnicity effect on wage from covariates, but I extend the analysis of the unconditional median racial wage gap to *unconditional* quantile regressions. In particular, I use the strategy developed in Firpo, Fortin, and Lemieux (2009) to decompose the unconditional difference of median log-wages between Blacks and Whites in a wage structure and a composition effect.

Selection and exclusion-restriction An abundant literature in econometrics, starting with the seminal work of Heckman (1979), studies selection biases for several statistics. A similar approach to the Heckman selection model in quantile regression has recently been developed by Arellano and Bonhomme (2017). Any model of selection relies heavily on the *exclusion restriction* hypothesis, that is that some vector of variables Z plays a role in the selection process, but not in the determination of the variable of interest. In our case, the use of one of the above selection model requires an available variable Z that plays a role in the selection, mainly the probability of going to jail, but not on the level of wages offered on the market. In that case, a part of the selection process relies on *observable* variables.

In the case of selection through detention in institutions, it is difficult to find such a vector Z of variables that plays a role in being in institution and no role in the determination of offered wages. As most males between 25 and 55 in institutions are actually in jail, the study of the selection process is congruent to the study of the determinants of incarceration. As a complex phenomena, crime is determined by

a series of socio-economic variables, such as poverty, social exclusion, wage inequality, cultural and family background, geography, age, sex etc.[1] Most of these variables are either taken into consideration in my analysis (age, gender, education), the subject of this work (poverty, wage inequality), or not a valid instrument (geography). A different issue arises when studying the determinants of incarceration, once the individual has been arrested: Western and Pettit (2002) note that both the probability of being arrested and of being incarcerated are highly ethnicity dependant.

Population of interest   I focus on males between 25 and 55 years old, living on the US soil, whether it is in their households or in group quarters. In 2016, the United States was the second country with the highest incarceration rate in the world[2]: therefore a non-negligible number of individuals drop-out of the labour market due to imprisonment. As I do not model endogenous labour market participation, the purpose of data restriction here is to restrict my sample of interest to individuals who are the most likely to participate to the labour market.

I restrict the analysis to men to minimize the number of individuals who might voluntarily choose to drop out of the labour market to favour in-house activities such as children care. As the gender wage gap is non-negligible and in favour of men, women are more prone to exit the workforce voluntarily.

Below 25, the participation rate of young males is highly correlated with their education level, and whether or not they are still attending school. At 25, the share of males still at school is low, while above 55 the participation rate starts to decline. According to the *Bureau of Labor Statistics*[3], in 2014 the participation rate of 16 to 24 years old was 55%, while only 40% of people aged 55 and older took part in the labour market. The participation rate of individuals between 25 and 54 was 80.9%.

Data

The ACS and Group Quarters   The American Community Survey is the second largest survey realised by the U.S. Census Bureau, with the explicit aim to track the socio-demographic characteristics of the American population regularly between decennial Censuses. As with the Census, the ACS targets individual living in regular households as well as those living in Group Quarters. As opposed to the Censuses, the ACS does not provide the details on the group quarter in which the individual lives, whether it is a correctional or a mental institution. The ACS provides with three general categories of group quarters: standard household, correctional and non-correctional group quarters. I exclude the last category from the analysis: the number of individuals concerned is often small, and stands for individuals at the university, whom are not in the scope of this analysis.

I use the 1-Year ACS surveys between 2006 and 2016, for which questionnaires are passed during the entire calendar year. The sample size of each 1-Year ACS increases from 2.9 millions housing units in 2006 to 3.5 millions in 2014. Once the data is trimmed to my population of interest, each 1-Year ACS is reduced to datasets of between 410,000 and 430,000 observations.

Labour income   In the ACS, respondents report their labour income by filling the item "Wages, salary, commissions, bonuses, or tips from all jobs" in the past 12 months. Note that, as the ACS is collected

---

[1] See Buonanno (2003) for a review of the studies on the determinants of crime in several fields.

[2] According to the World Prison Brief data : http://www.prisonstudies.org/country/united-states-america

[3] See : `https://www.bls.gov/emp/ep_table_303.htm`

throughout the calendar year, responses to this item do not necessarily stands for the individual income of the year of the survey. An individual filling the form in March 2015 would declare his perceived labour income from March 2014 to date. I use the adjustment variable (ADJUST) made available by the Census Bureau to imperfectly control for this discrepancy[4].

I use the log-yearly labour income as my variable of interest. Yearly labour income does not allow to identity productivity from the quantity of labour supplied, but takes into account periods of unemployment.

As for the distributional statistic of interest, Winters and Hirsch (2012) follow the evolution of the racial wage gap at every decile of the distribution. In this paper, I focus on the median, as high incomes are subject to top-coding in the ACS, and the differences in unemployment and incarceration between Blacks and Whites yield very high wage gaps in the lower deciles of the distribution. For these same reasons, I favour the median log-wages over the average. Overall, I take advantage of the robustness of the median to extreme values in the income distribution.

Sampling design and variance computation   Considering the ACS, the sample selection is done through a complex two-stage sampling based on the last decennial Census. From 2005 onwards, the ACS has been provided with person replicate weights, to allow estimation of sampling variability using Balanced Repeated Replications, with 80 sets of replications for each year. I follow the Census Bureau instructions and compute the variance of every estimators with Fay's method[5].

Let's have θ the statistic of interest and $\hat{\theta}$ its estimator using the weighted full sample. The BRR technique consists in sampling 80 independent half samples -one for each set of replication weights in the ACS- and computing the estimator $\hat{\theta}_p$ for each of them. The unbiased estimator of the sampling variance is the average of $(\hat{\theta}_p - \hat{\theta})^2$.

The Fay's method with parameter k ($k = 0.5$ in the ACS) can be seen as a generalisation of the BRR: for each replication, one half-sample is weighted by k while the other half is weighted by $2 - k$. The sampling variance estimate is therefore the average of $(\hat{\theta}_i - \hat{\theta})^2$, as above, divided by $(1 - k)^2$.

2   The Racial Wage Gap Before, During and After the Crisis

2.1   Unconditional Racial Wage Gap

Unconditional racial wage gap   Consider a population of interest constituted with at least two groups H and B. Write W_l the yearly labour income of an individual belonging to the group l, with $l = H, B$. The observed and unconditional average racial wage gap is

$$\Delta_o^\mu = \bar{W}_H - \bar{W}_B \tag{2.1}$$

where $\bar{W}_l$ stands for the average of W_l. It is usually more convenient to work with w the natural log of the annual labour income W:

$$\delta_o^\mu = \bar{w}_H - \bar{w}_B \tag{2.2}$$

---

[4]See the Census Bureau documentation at `https://usa.ipums.org/usa/acsincadj.shtml`

[5]As recommended in the Census Bureau documentation of the ACS: `https://www.census.gov/content/dam/Census/library/publications/2010/acs/acs_design_methodology_ch12.pdf`

and a similar definition stands for the τth quantile of the racial wage gap, thought of as the difference between each τth quantile of the distribution of the log-wage for each group:

$$\delta_o^\tau = Q_\tau(w_H) - Q_\tau(w_B) \tag{2.3}$$

In this paper, I name *unconditional racial wage gap* the difference in the natural log of median annual labour income between Blacks and Whites. This observed quantity is written $\delta_o^{0.5}$ following Equation 2.3.

Evolution of the Racial Wage Gap in the mid-term   Using Census data, the *American Community Survey* and the *Current Population Survey*, I compute this unconditional (median) racial wage gap for my population of interest using the above formula in Equation 2.3.

Figure 1 shows the evolution of this unconditional racial wage gap for my population of interest minus those living in institutional group quarters, between 2006 and 2015. This restriction allows me to make comparisons between the *ACS* and the *CPS*, as the later is most commonly used in the computation of labour statistics. Between 2006 and 2015, the median racial log-wage gap has significantly increased from a point estimate of 0.43 to 0.61, using the *ACS* as the source of this estimation. The same source indicates a decrease from this high estimate to 0.51 in 2015.

There is little to no statistical differences in the computation of the unconditional racial wage gap between those two surveys. As Figure 2 shows, the 95% confidence interval of the median racial wage computed with both surveys overlap every year from 2006 to 2015, but in 2006, 2014 and 2015. This first approach indicates that the Great Recession is contemporaneous with a statistically significant increase of the unconditional racial wage gap. How does this high level of racial inequality compares through time? And what does it become once we take into account selective withdrawal from the labour market?

The Racial Wage Gap and the long-term   Figure 1 shows the evolution of the unconditional racial wage gap every ten years between 1960 and 2000 using *Census* data, and every year from 2006 to 2015 using the *ACS*.

The median racial wage gap has decreased drastically between 1960 and 1970, of about 0.17 log-points. This improvement of the economic living conditions of Blacks after the Civil Rights acts are well documented, for its immediate aftermath to the end of the 1980s in Smith and Welch reference paper [Smith and Welch (1989)]. From 1970 to 2008, the level of the median racial log-wage gap remained between 0.43 and 0.48. The largest increase in the racial wage gap arises in 2009 and 2010, during the Great Recession. From a point estimate of 0.48 in 2008, the median racial log-wage gap increases to 0.59 in 2010. This high value for the unconditional racial wage gap is only second to the 0.64 estimate of 1960, and decreases in 2014 and 2015. This last estimate remains statistically different and higher to the pre-crisis level of the unconditional racial wage gap.

Correcting for selection by imprisonment   As important the increase of the racial wage gap during the Great Recession appears to be considering the previous estimations, it only imperfectly accounts for the actual racial wage gap. Since the late 1980s, the American society has experienced a dramatic
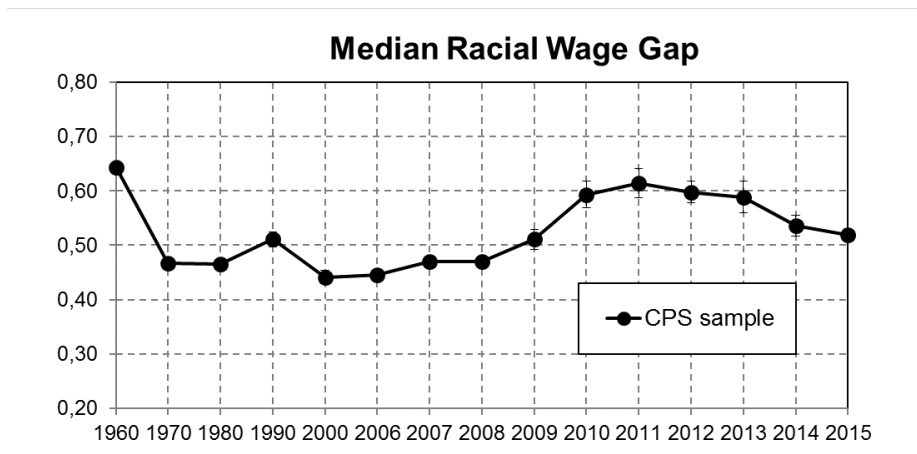
Figure 1: Unconditional Racial Wage Gap computed with the American Community Survey and excluding individuals leaving in Group Quarters
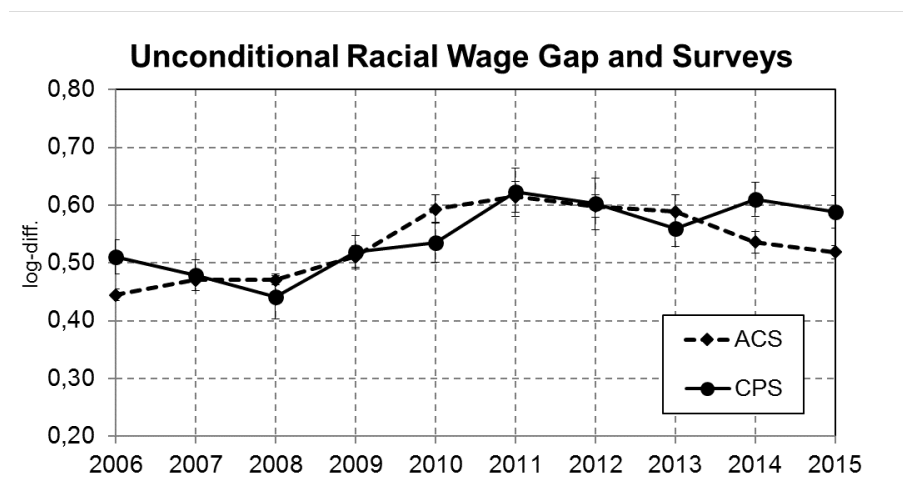

Figure 2: Unconditional Racial Wage Gap computed using the *American Community Survey* and the *Current Population Survey*

1990, 2000 Census and ACS	1960, 1970 and 1980 Census
Non-group quarter households	Non-group quarter households
Institution	
	Correctional institutions
	Mental institutions
	Institutions for the elderly, handicapped and poor
Non-institutional group quarters	
	Military
	College dormitory
	Rooming house
Other non-institutional GQ and unknown	Other non-institutional GQ and unknown

Table 1: Group quarters variable and surveys

rise in its incarceration rate, which has been multiplied by five from 1985 to 2015. As Neal and Rick (2014) mention, incarceration is not ethnicity neutral: individuals of colour are far more likely to be incarcerated than Whites. As those individuals get incarcerated, they exit the sampling universe of the *Current Population Survey*, thus yielding a selection bias in racial wage gaps computed with this survey. They do remain in the *American Community Survey* and the *Census*, which therefore must be used to compute the unconditional racial wage gap.

While the Census from 1960 to 1980 gives us the detailed type of group quarters in which the respondent lives, the 1990, 2000 Census and the ACS as well are not as precise. Table 1 shows the differences in the details of the group quarters variables available for *Census* waves and the *ACS*. While the earlier waves of the Census allows for identification of the evolution of the population in correctional institutions, from 1990 onwards it is impossible to distinguish correctional from non-correctional institutions using IPUMS. In the rest of this paper, I use the four modalities versions of the group quarters variable, thus striking the differences between correctional and mental institutions from my identification strategy. As such, there are reasons to trust that detention in mental institutions is not unrelated to ethnicity[6], it is unclear whether the decrease of individuals in mental institutions that we observe in the data is unrelated to the increase of individuals in jails (see Torrey, Kennard, Eslinger, Lamb, and Pavle (2010) on this issue). Furthermore, by selecting grown-up men, I minimize the number of individuals that would be subject to being placed into the third type of institutions, which are "Institutions for the elderly, handicapped and poor".

Institutions and ethnicity The stronger the asymmetry in the distribution of individuals in institutions by ethnicity, the more an estimator of the racial wage gap taking into account institutionalized individuals will be different from the naive estimator. Table 3 shows the repartition of Blacks and Whites in institutions relative to my population of interest across years. The most striking element of this repartition is the dramatic increase in the relative share of Blacks in institutions between 1960 and 2000. I

---

[6]

    Racial and ethnic minorities have less access to mental health services than do whites. They are less likely to receive needed care.

in Office of the Surgeon General (US), Center for Mental Health Services (US), and National Institute of Mental Health (US) (2001)
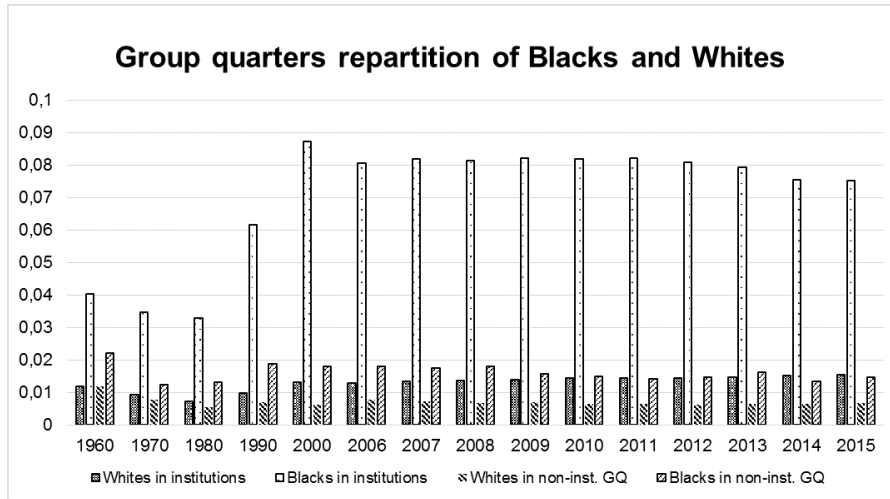
Figure 3: Share of total Blacks and Whites of our population of interest who live in GQ and their repartition

Figure 4: Average yearly income (1999 dollars) of individuals in institutions with 95% confidence interval

estimate that, in 1960, 4% of the individuals of my population of interest declaring to be Black were in institutions, while this share has more than doubled in 2000, with 8.8% of the total number of Blacks being in institutions[7]. I expect this important asymmetry to increase the observed unconditional racial wage gap.

Labour income in institutions   Although individuals in institutions vanish from the population of interest of the *Current Population Survey*, they remain on the sample of the *American Community Survey*. As such, interviewees answer to the question related to labour income in the last 12 months. Figure 4 shows their average annual labour income, in 2017 US dollars, which is capped to 5,000 USD but for the Census years of 1990 and 2000. In the case of inmates, existing evidence on the wage profiles of

---

[7]Those striking numbers can appear to be at odd with the estimate of 0.7% of the general population imprisoned in 2000, even though Institutions here have a larger definitions that only jails. One must keep in mind that my population of interest is a fraction of the general population, but is over-represented in correctional institutions. To put this result in the right context, in 2000 the US was populated of 282 millions individuals, with 2 millions in jail. Blacks and Whites between the ages of 21 and 64 represent 56% of the total population, and almost 80% of the inmates. This back-of-the-envelop calculus, assuming total independence between age, ethnicity and imprisonment, yields an aggregate detention rate of 1% for my trimmed data.
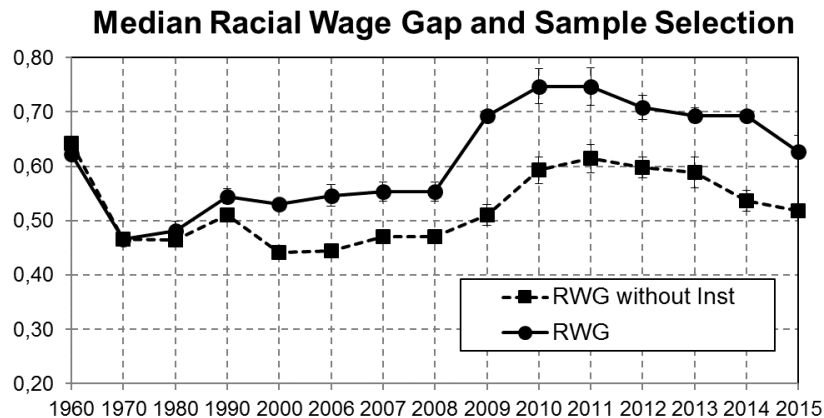
Median Racial Wage Gap and Sample Selection

Figure 5: Unconditional racial wage gap computed with Census data (1960 to 2000) and the *ACS*.

inmates show that their labour income are far lower than the general population, both before and after incarceration (Rabuy and Kopf (2015) for the former, and Visher, Debus, and Yahner (2008) for the later). While the "institutions" category is larger than inmates only, finding low annual labour income is consistent with these previous evidence. Note that, using the *ACS*, there are no means to know for how long respondents have been in institutions, nor whenever they will be able to get out. Finally, a large part of individuals in institutions are non-earners. How does the racial asymmetry in detention in institutions gets translated to the racial wage gap when we extend the sample to this population?

The racial wage gap and group quarters   As yearly labour income is available for individuals in institutions, I compute the median racial log-wage gap for this extended sample of my population of interest, and show the results of those estimations in Figure 5. I find that taking into account individuals in institutions dramatically increases the unconditional racial wage gap, as the difference between the value computed without and with individuals in institutions are statistically significant from 2000 onwards, and by more than ten points. Both series increase with the Great Recession, although this increase is even larger when I take into account individuals in institutions. In 2010, the level of the racial wage gap for this sample reaches a high at a point-wise estimate of 0.75.

The unconditional racial wage gap does not take into account the differences between each group in the distribution of variables which can explain wages. If Whites are on average more educated than Blacks, then a share of the unconditional racial wage gap is explained by the differences in educational attainment, and not by the difference in ethnicity. On one hand, computing the *conditional racial wage gap* will allow me to measure the evolution of the wage difference between Blacks and Whites through time while controlling for structural differences in age or educational attainment. On the other hand, *unconditional quantile regression* will allow me to measure the share of the *unconditional* racial wage gap that is explained by those factors.

2.2 Conditional Racial Wage Gap

Conditional Racial Wage Gap   I compute the log-difference of the median yearly labour income of Blacks and Whites, conditional on education and age:

$$Q_\tau[\ln(W)|X] = Xb(\tau) + u \tag{2.4}$$

where the covariate X are the ethnicity, the age and the educational level, and $\tau = 0.5$ the quantile of interest of the distribution of the log-labour income. The educational attainment is decomposed in three modalities: whether the individual is a high school drop-out (HS-), graduated from high school (HS), or earned a degree after high school (HS+). The age is a proxy for labour experience.
I name *conditional median racial wage gap* the quantity

$$\delta_{cond}^{0.5} = Q_{0.5}[w|H = Whites, \tilde{X}] - Q_{0.5}[w|H = Blacks, \tilde{X}] \tag{2.5}$$

where $\tilde{X} = age, educational\ attainment$. Running the following quantile regression on my population of interest, an estimator of $\delta_{cond}^{0.5}$ is the quantity $\hat{\beta}_1$:

$$Q_{0.5}[w] = \alpha + \beta_1 * \mathbb{1}_{White_i} + \beta_2 * age_i + \beta_3 * \mathbb{1}_{HS_i^+} + \beta_4 * \mathbb{1}_{HS_i} + u_i \tag{2.6}$$

Assume that $\hat{\beta}_1 > 0$. One must remain careful in the interpretation of the conditional racial wage gap, as it precisely say that for a given value of age and education, the median log-wages of Whites is higher of $\hat{\beta}_1$ compared to the one of Blacks. We can not say that, (*on average*) the median log-wage of Whites is higher of $\hat{\beta}_1$ to the log-wage of Blacks. Nonetheless, as it controls for differences in age and education, the conditional racial wage gap is useful as a first approach.

The evolution of the conditional racial wage gap is similar to the evolution of its *unconditional* counterpart (Figure 6). As in the unconditional case, its increase is larger during the Great Recession for the sample including people in institutions than with those living in households only. For the former, the conditional racial wage gap goes from a point estimate of 0.42 in 2007 to 0.59 in 2010, remaining constant from 2010 to 2014. For the household-only sample, the increase between 2007 and 2010 is of an order of magnitude of 0.10, reaching a similar *plateau* for four years afterwards. Note that the level of the conditional racial wage gap for the sample with institutionalised individuals is significantly higher in the 2010-2013 time period than in 1960. Finally, as in the unconditional case, the conditional racial wage gap significantly decreases in 2015 for both samples, but remains far from its pre-crisis level.

In this section, I first computed the unconditional difference between the log-annual wages of working-age White and Black males, which I named unconditional racial wage gap. Using the *American Community Survey*, I showed that this quantity has significantly increased between 2008 and 2010, that is during the Great Recession. Once I account for the withdrawal of individuals from the labour market into institutions, I find that this unconditional racial wage gap reaches a higher level in 2010 than in 1960, that is before the Civil Rights Act. Controlling for age and education yields a similar result: the *conditional* racial wage gap in 2010 is higher than the one computed in 1960, once individuals in institutions are taken into account.
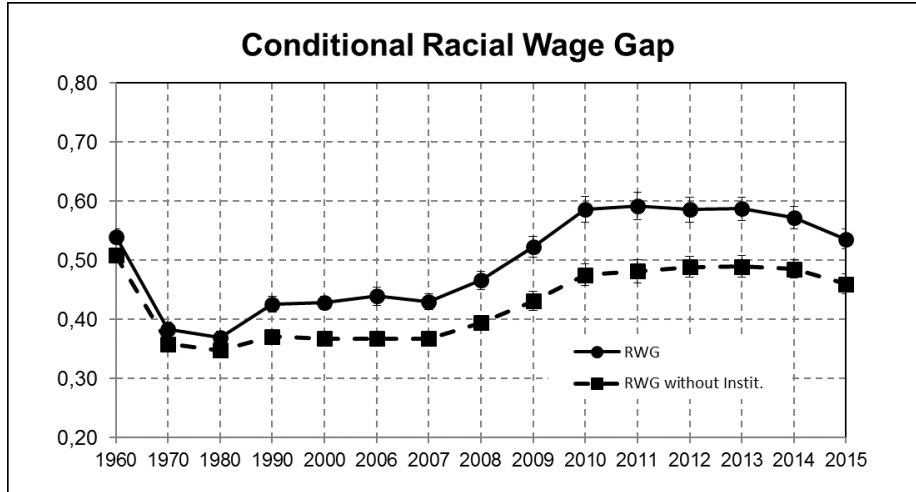
Figure 6: Conditional median racial wage gap

3 Decomposition of the Racial Wage Gap

The *conditional* median racial wage gap measures the differences in the conditional quantiles of the distribution of the log-wages for Blacks and Whites. The previous (conditional) estimators answer to the question: "for a given age and level of education, what is the corresponding observed median wage for Blacks ? for Whites ?". As conditional quantiles do not average out to their unconditional counterpart, the conditional racial wage gap does not tell us anything on the *unconditional* racial wage gap. We can not write that "on average, conditional on age and education, the median yearly log-wages of Blacks is smaller of $\delta_{cond}^{0.5}$ to the median yearly log-wages of Whites". This makes it impossible to build a simple counter-factual such as "what would be the median of log-annual earnings of Blacks would they have the same age and education structure as Whites?". In this section, I use the Recentered Influence Function regression of Firpo, Fortin, and Lemieux (2009) to build such a counter-factual, and study the evolution of the "unexplained" part of the unconditional racial wage gap through time.

Blinder-Oaxaca decomposition Assume that the unconditional wages of Blacks and Whites are a function of observable and unobservable variables, such that $w_l = g_l(X_l, \varepsilon_l)$ where $l = H, B$. As such, both the distribution of the X, ε and the function g differ between each group. Differences in the distributions of the X between group named the *composition* effect, while differences in g are called *wage structure* effects. A simple counter-factual one can think of for an individual i belonging to the class B is $w_c = g_W(X_{i,B}, \varepsilon_i)$, that is his wage would his human capital X be paid the same price as if he was to belong to the class W.

Such a decomposition is standard to get in the case of differences in averages between groups, using the Blinder-Oaxaca decomposition. Assume that one is to compare the average outcome $\bar{w}_l$ between two groups $l = H, B$. The unconditional gap is therefore $\Delta_o^\mu = \bar{w}_H - \bar{w}_B$. Assume that both w_l can be explained by the same set of covariates X in a linear regression, such that for each group we have $w_{li} = X'_{li} \beta l + \varepsilon_{il}$ with conditionally independent errors. Estimating the models to get the $\hat{\beta}_l$ for each group, we can build a counter-factual such as the wages of the individuals of the group B (with the X_{Bi}) would they have the wage structure of group H. This counter-factual is $X'_{Bi}\hat{\beta}_H$. Then the overall wage

gap can be decomposed as

$$\Delta_o^\mu = X_H'(\hat{\beta}_H - \hat{\beta}_B) + (X_H' - X_B')\hat{\beta}_B \tag{3.1}$$

where the first term on the right hand side of the equation is the *unexplained* part of the wage gap, that is the wage structure effect, and the second term the *explained* part of the wage gap, that is the composition effect. Moreover, this equation can be decomposed even further by developing each product $X'\beta$ to compute the contribution of each variable in the total (unconditional) wage gap. Note that in the decomposition of Equation 3.1, the wage structure that serves as a reference is the wage structure - the $\hat{\beta}$ - of group B. Another Blinder-Oaxaca decomposition can be done with the same data, using the wage structure of group H as the reference group. To compute each decomposition, one simply has to retrieve the $\hat{\beta}_l$ of each linear regression of the log-wage on each set of covariates, as well as the average of those covariates per group.

3.1 The RIF regression and the decomposition of the racial wage gap

The overall issue is to decompose the difference in a statistic of interest of the distribution of wage (here the median) in two parts: the part due to explained characteristics - the composition effect - and the part due to unexplained characteristics - the wage structure effect. In the case of the difference between two averages, one can use the Blinder-Oaxaca decomposition. Because quantiles regressions do not average out to unconditional quantiles, we can not use the Blinder-Oaxaca method to operate this distinction between an explained and an unexplained effect.

DiNardo, Fortin and Lemieux (1996) submit a strategy to decomposed the observed gap in a quantile of interest into an "explained" and an "unexplained" part. From the conditional distributions of the log-wages of each group

$$F_{w_l}(w) = \int F_{w_l|X_l}(w|X)dF_{X_l}(X) \tag{3.2}$$

they use a reweighing factor $\Psi(X) = \frac{dF_{X_H}(X)}{dF_{X_B}(X)}$ to build the counter-factual distribution

$$F_{w_B^C}(w) = \int F_{w_B|X_B}(w|X)\Psi(X)dF_{X_B}(X) \tag{3.3}$$

This counter-factual distribution is precisely the distribution of w would the individuals of group B had the observed characteristics of group H. Remark that this decomposition implies to have a reference group: here, the reference group is group H. The counter-factual individuals have the observed characteristics of group H, but the wage structure of group B.

This strategy allows to distinguish the share of the unconditional difference in quantiles which is due to the Xs, but does not allow to go further in the decomposition, and especially to find the role of each observable variable in the observable difference as with the Oaxaca-Blinder decomposition. This limitation is overpassed by the Recentered Influence Function Regression.

The strategy pursued by Firpo, Fortin and Lemieux [Firpo, Fortin, and Lemieux (2009)] is to work on the influence function of the statistic of interest to conduct a Oaxaca-Blinder decomposition. The influence function of a statistic ν - here the median - measures its behaviour when the distribution from which it is applied changes infinitesimally toward another distribution of the same class, that is for which

11

the statistic is also defined and bounded. Formally, when considering a statistic out of the distribution of some random variable Y, and using the notations of the authors:

$$IF(y; v, F_Y) = \left.\frac{\partial v(F_{Y,t.\Delta y})}{\partial t}\right|_{t=0} \tag{3.4}$$

which integers to 0 over the support of F_Y. Note that in the case of a quantile τ, the influence function is given by $(\tau - \mathbb{1}\{w \le Q_\tau\})/f_w(Q_\tau)$. Then, the recentered influence function is defined as

$$RIF(y; v, F_Y) = v(F_Y) + IF(y; v, F_Y) \tag{3.5}$$

which integers to the statistic v over F_Y, but which can also be used with covariates:

$$v(F_Y) = \int RIF(y; v, F_Y) dF_y \tag{3.6}$$

$$= \int \int RIF(y; v, F_Y).dF_{Y|X}(y|X = x).dF_X(x) \tag{3.7}$$

$$= \int \mathbf{E}[RIF(Y; v, F_Y)|X = x].dF_X(x) \tag{3.8}$$

In the case of (conditional) quantile regressions, the Blinder-Oaxaca is impossible, as $\mathbf{E}_X Q_\tau(Y|X) \ne Q_\tau(Y)$. By running an OLS on the *Recentered Influence Function*, that is $RIF(Y; \tau; F_Y) = Q_\tau(Y) + IF(Y; \tau; F_Y)$, one gets the unconditional quantile $Q_\tau(Y)$, and can run an Oaxaca-Blinder decomposition.

Assuming that the conditional expectation of the $RIF(Y; \tau; F_Y)$ on some set of covariates X can be modelled as a linear regression $\mathbf{E}[RIF(Y; \tau; F_Y)|X] = X\beta$, then the unconditional median log-wage gap can be decomposed as

$$\delta_o^{0.5} = Q_\tau(w_H) - Q_\tau(w_B) = \mathbf{E}\left[RIF(w_H, F_{w_H})\right] - \mathbf{E}\left[RIF(w_B, F_{w_B})\right] \tag{3.9}$$

$$= \mathbf{E}\left[RIF(w, F_w)|l = H\right] - \mathbf{E}\left[RIF(w, F_w)|l = B\right] \tag{3.10}$$

$$= \mathbf{E}\left[\mathbf{E}_X[RIF(w, F_w)|X_H, l = H]\right] - \mathbf{E}\left[\mathbf{E}_X[RIF(w, F_w)|X_B, l = B]\right] \tag{3.11}$$

$$\delta_o^{0.5} = \bar{X}_H \hat{\beta}_H^{RIF} - \bar{X}_B \hat{\beta}_B^{RIF} \tag{3.12}$$

which allows us to introduce a counterfactual $\bar{X}_B \hat{\beta}_H^{RIF}$, to decompose the unconditional racial wage gap into an "explained" and an "unexplained" component:

$$\delta_o^{0.5} = (\bar{X}_H - \bar{X}_B)\hat{\beta}_H^{RIF} + \bar{X}_B(\hat{\beta}_H^{RIF} - \hat{\beta}_B^{RIF}) \tag{3.13}$$

The decomposition obtained in Equation 3.13 is very similar to the above-mentioned Blinder-Oaxaca decomposition (Equation 3.1). In the same fashion, one must choose a wage structure of reference to measure the share of the unconditional difference between quantiles that is explained by the wage structure and the composition of the groups. In Equation 3.13, I choose the wage structure of group H as a reference group, as the evidence of the previous section tells us that the racial wage gap is mostly in favour of this social group.

The main benefit of the RIF regression, compared to the reweighing of the sample, is to allow to push further the Blinder-Oaxaca decomposition. Take the "explained" part of the unconditional racial wage gap $(\bar{X}_H - \bar{X}_B)\hat{\beta}_H^{RIF}$. One can develop this product to obtain the contribution of each covariate to the composition effect:

$$(\bar{X}_H - \bar{X}_B)\hat{\beta}_H^{RIF} = \hat{\beta}_{1,H}^{RIF}(\bar{X}_{1,H} - \bar{X}_{1,B}) + ... + \hat{\beta}_{p,H}^{RIF}(\bar{X}_{p,H} - \bar{X}_{p,B}) \tag{3.14}$$

and the same development can be applied for the wage-structure component of the unconditional racial wage gap.

3.2 Baseline Specification

In this subsection, I analyse the evolution of the "explained" and "unexplained" shares of the unconditional racial median log-wages gap between Blacks and Whites, with the latter group as reference. I show that most of the racial wage gap is unexplained, if we limit ourselves to age and education as main drivers of wages. The unexplained share of the racial wage gap decreased strongly between 1960 and 2000, and remained stable from 2006 onwards. Differences in educational achievement appear as the main factor behind the explained share of the racial wage gap.

I run the RIF-regression and subsequent Blinder-Oaxaca decomposition with the same covariates as above in the treatment of the *conditional* racial wage gap: the age of individuals and their educational achievement, whether they have only graduated from high school (*HS*), reached a higher level of education (*HS+*) or none of that. I use the entire sample of my population of interest, including individuals in institutions. I assume a linear form for the RIF-regression, such that for each group $l = B$ (Blacks) and $l = H$ (Whites), the regression is

$$\mathbf{E}[RIF(w_l; \tau = 0.5)|X_l] = \beta_{0,l}^{RIF} + \beta_{1,l}^{RIF} * age_l + \beta_{2,l}^{RIF} * \mathbb{1}_{HS_l^+} + \beta_{3,l}^{RIF} * \mathbb{1}_{HS_l} \tag{3.15}$$

which yields two sets of estimates, $\hat{\beta}_B^{RIF}$ and $\hat{\beta}_B^{RIF}$. Note that the dummy *HS*, designing high school returns, take as reference position the fact of being a high school drop-out (HS^-), which is left out of the equation to avoid perfect collinearity between regressors.

As the previous work on the unconditional racial wage gap has shown, Whites have the highest median log-wages of the two groups. I therefore choose them as the reference group, as it is common in the discrimination literature to choose the highest earners as the reference. The decomposition is therefore the same as in Equation 3.13:

$$\delta_o^{0.5} = (\bar{X}_H - \bar{X}_B)\hat{\beta}_H^{RIF} + \bar{X}_B(\hat{\beta}_H^{RIF} - \hat{\beta}_B^{RIF}) \tag{3.16}$$

I use Nicole Fortin's **rifreg** package for Stata to run the RIF-regression, and Ben Jann's **oaxaca8** routine to run the decomposition with the β^{RIF} from the RIF-regression and $\bar{X}$ from the data.

3.2.1 The "unexplained" share of the racial wage gap

Following the above decomposition of the unconditional median racial wage gap (Equation 3.13), I write its "unexplained" share $\bar{X}_B(\hat{\beta}_H^{RIF} - \hat{\beta}_B^{RIF})$. This "unexplained" share is the expected difference in
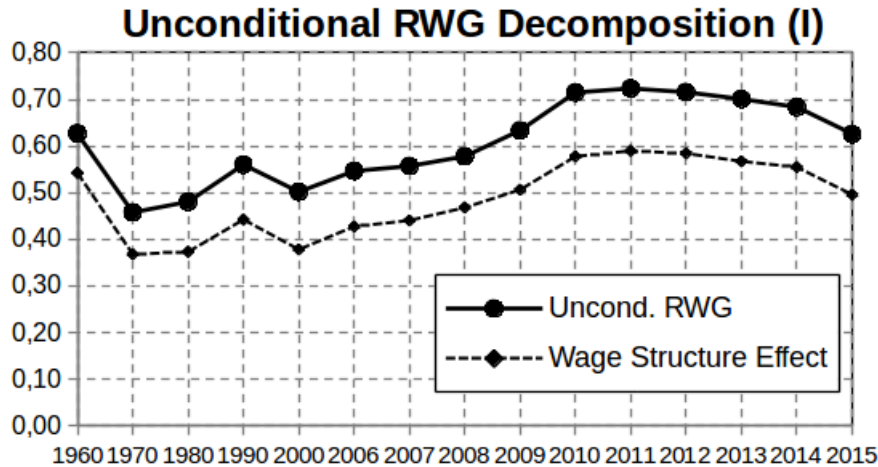
Figure 7: Unconditional Racial Wage Gap Decomposition (I) across Time

Reading note: In 2011, the unconditional racial wage gap (RWG) was 0.72. Differences in wage structures between Blacks and Whites, with the latter as the reference group, accounted for 0.58 of this difference, so 85.5% of the total difference.

the median log-wages of Blacks and Whites, if Whites had the characteristics of Blacks X_B. In studies in wage inequality, it is often described as the difference in the wage structure between groups, as one can think of the β as the rate of return on socio-economic characteristics X.

Using the RIF-regression (Equation 3.15) and the Blinder-Oaxaca decomposition (Equation 3.13), I find that the effect of the differences in wage structure are a large part of the unconditional racial wage gap between Blacks and Whites. Figure 7 shows the evolution of both the unconditional racial wage gap and the difference of wage structures for the entire sample. The "unexplained" component of the unconditional racial wage gap follows closely the evolution of the later. Nonetheless, in the long-run, it is worth mentioning that the unexplained share of the racial wage gap decreases from 1960 to 2000 by 11 points, as shows Figure 8, preceding a rise of 5.6 points between 2000 and 2008, without any significant movement in one direction or the other ever since. Although the unconditional racial wage gap has increased during the Great Recession to an all-time high according to our calculations, its unexplained share has not.

3.2.2 Relative importance of covariates in the RWG

An important asset of the RIF-regression is to allow to decompose both "explained" and "unexplained" parts of the unconditional racial wage gap in its covariates, and to measure the influence of each variable in the results obtained.

Decomposition of the explained share of the RWG Using only the age and a dummy for educational achievement (Equation 3.15) as covariates, it appears that most of the explained share of the racial wage gap is due to differences in endowments in the latter [Figure 9]. Age differences are not significant in this decomposition, which should not be too surprising as the data has been trimmed for this variable
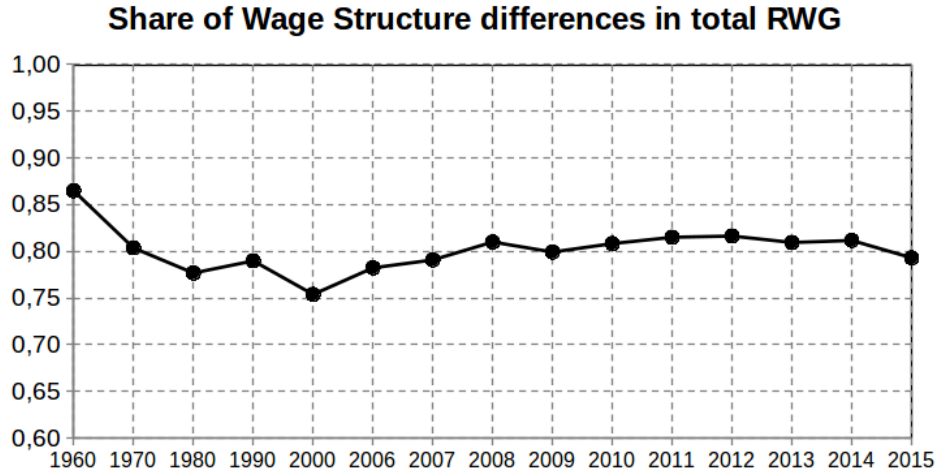
Share of Wage Structure differences in total RWG

Figure 8: Wage structure share of the unconditional racial wage gap.

Reading note: In 2011, the wage structure effect accounted for 81% of the unconditional racial wage gap.
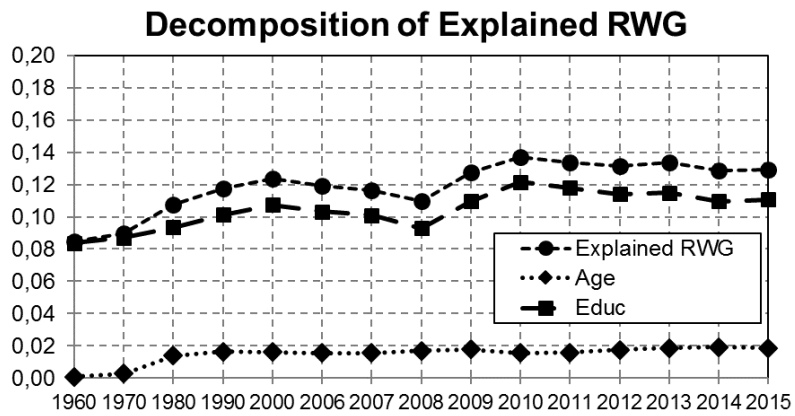

Decomposition of Explained RWG

Figure 9: Decomposition of the explained share of the unconditional Racial Wage Gap

Reading note: In 2011, the explained racial wage gap in median log-annual wages was 0.132.

before the analysis. I detect no change in the relative share of age and educational achievement between 2000 and 2015.

Decomposition of the unexplained share of the RWG The Blinder-Oaxaca decomposition allows to go further into the analysis of the unexplained share of the RWG, which can be developed as:

$$\bar{X}_B(\hat{\beta}_H^{RIF} - \hat{\beta}_B^{RIF}) = \beta_{0,H}^{RIF} - \beta_{0,B}^{RIF} + (\hat{\beta}_{1,H}^{RIF} - \hat{\beta}_{1,B}^{RIF})\bar{age}_B + (\hat{\beta}_{2,H}^{RIF} - \hat{\beta}_{2,B}^{RIF})\mathbb{1}_{\overline{HS_B^+}} + (\hat{\beta}_{3,H}^{RIF} - \hat{\beta}_{3,B}^{RIF})\mathbb{1}_{\overline{HS_B}}$$

(3.17)

Figure 10 shows the evolution of this decomposition, joining together the last two members of the above equation describing educational achievement. The most striking fact in the decomposition of this unexplained share of the racial wage gap is the negative contribution of the educational achievement, which gets regularly stronger from 1960 to 2010: the returns from education *compared to being a high*
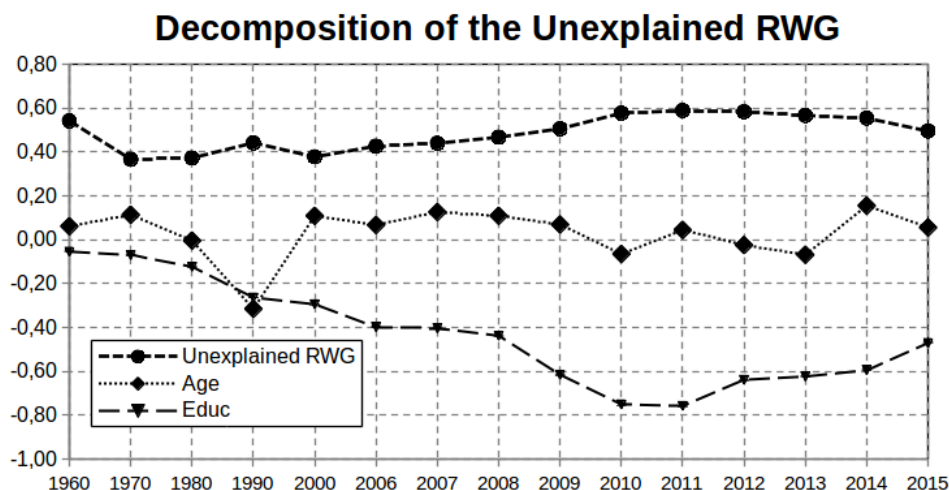
Decomposition of the Unexplained RWG

Figure 10: Decomposition of the Unexplained Racial Wage Gap

In 2011, the composition effect accounted for 40% of the overall difference in log-median wages between races.

school drop-out are estimated to be much higher for Blacks than for Whites. As the share of the age in this unexplained racial wage gap is non significant, a large share of the wage structure effect resides in the $\beta_{0,H}^{RIF} - \beta_{0,B}^{RIF}$ element of the decomposition, that is the unexplained part of the wage structure effect.

In this subsection, I show that despite an increasing racial wage gap in log-yearly wages, the share of this increasing difference explained by the age and qualifications of individuals remains almost constant during the Great Recession. Moreover, the marginal return of getting a high-school diploma or reaching a post-high-school educational attainment for Blacks compared to Whites increases between 1960 and 2010. In the following subsection, I study the robustness of those results when adding situational covariates such as the State of residence and the Occupation.

3.3 Fixed-effects Specifications

In this subsection, I enrich the previous baseline specifications with additional variables to control for regional and occupation differences between individuals. I use a set of State dummy variables to control for institutional and economical differences between States[8], and the unified 1990 Census Bureau occupational classification scheme (*OCC1990*), designed to provide a set of occupations comparable through time, from 1960 onwards [Meyer and Osborne (2005)]. I recode the 389 categories of this variable into 7 dummies figuring in Table 2, and in line with the Census Bureau instructions.

Conditional Racial Wage Gap with State and Occupation Controls   Adding explanatory variable to any statistical model changes the interpretation of the model results. Adding a set of State and Occu-

---

[8]Average GDP per capita can differ from a 1-to-5 factor between States (see Bureau of Economic Analysis: `https://www.bea.gov/scb/pdf/2017/08-August/0817-real-per-capita-personal-income.pdf`).

Categories	Corresponding Values
Managerial and Professional	000-200
Technical, Sales, and Administrative	201-400
Service	401-470
Farming, Forestry, and Fishing	471-500
Precision Production, Craft, and Repairers	501-700
Operatives and Labourers	701-900
Non-occupational responses	900-999

Table 2: Occupation Codes (*OCC1990*)

pations variables nests the conditional quantile regression of Equation 2.6 in Equation 3.18.

$$
\begin{aligned}
Q_{0.5}[w] = \alpha &+ \beta_1 * \mathbb{1}_{White_i} + \beta_2 * age_i + \beta_3 * \mathbb{1}_{HS_i^+} + \beta_4 * \mathbb{1}_{HS_i} \\
&+ \rho_1 * \mathbb{1}_{Alabama_i} + \dots + \rho_{51} * \mathbb{1}_{Wyoming_i} \\
&+ \gamma_1 * \mathbb{1}_{Managerial\ and\ Professional_i} + \dots + \gamma_6 * \mathbb{1}_{Operatives\ and\ Labourers_i} \\
&+ v_i
\end{aligned}
\tag{3.18}
$$

The previously estimated conditional racial wage gap assumed that $\forall i, j \ 0 = \rho_i = \gamma_j$. With both sets of State and occupational variables, the model has 62 explanatory variables. $\hat{\beta}_1$ is now an estimator of the difference between Blacks and Whites median log-wages for a given age/educational attainment/State/Occupation profile.

Adding conditioning variables reduces the scope of the analysis. $\hat{\beta}_1$ as the result of the estimation of Equation 3.18 gives us an estimator of the difference in the medians of two groups defined by Blacks and Whites, close to a given age, for a given educational attainment, in a given State, within a given occupation. To put it clearly in an example, $\hat{\beta}_1$ is an estimator of the difference between the median of the log-yearly wage for the group formed of Black males close to 30 years old who have not graduated from high-school, working in Alabama as Operatives and labourers, and the median of their close White colleagues[9]. We are therefore comparing the medians of two groups which are more homogeneous, maybe more comparable, the drawback being the education variable loosing interest: one goes to college not to become a higher paid individual *for a given State + Occupation combination*, but precisely to move to another higher-earning State to occupy higher-earning occupation. Beside from yielding less precise estimates, controlling for state and occupation precisely make us look at the conditional racial wage gap within a given occupation and State profile.

Estimation of the Unconditional RWG with several specifications   I estimate the conditional quantile regression 3.18 for several specifications: the baseline specification (with the age and the educational attainment variable as covariates only), the State specification (adding the States dummies to the baseline specification), the "Occ." specification (adding the Occupation dummies to the baseline specification) and the full specification. I work on the full sample, that is including individuals in institutions.

The conditional racial wage gap varies greatly depending on the covariates added (Figure 11). When controlling for occupation, the conditional racial wage gap is much smaller than in the baseline specifi-

---

[9]It is also the estimator of the difference in medians of any group of individuals having the same age/educational attainment/State/Occupation profile but different ethnicities, in this model.
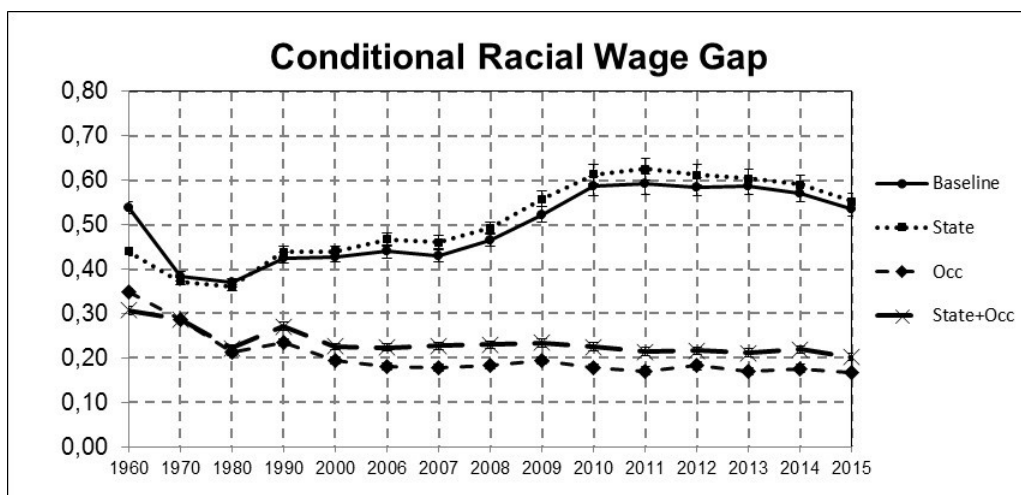
Figure 11: Conditional Racial Wage Gap and Specifications

In 2008, the conditional Racial Wage Gap was of 0.45 for the baseline specification, 0.49 when adding the State as covariate, and 0.19 when controlling for differences in occupations.

cation, below 0.20 from 2006 onwards. Moreover, after a significant decrease between 1960 and 1980 similar to the case of the baseline specification, the conditional racial wage gap remains stable during the Great Recession, for both specifications using the Occupation as covariates. We can deduce that at the occupation level, the wage gap between Blacks and Whites is smaller than at the level of the general population, but remains significant, although no movement has been detected during the Great Recession. It is therefore the structure of the relative distribution of Blacks compared to Whites in the education variable and *between* occupations that will explain the unconditional racial wage gap.

Note that adding our State variable has little effects on the estimator of the conditional racial wage gap, as opposed to adding the Occupation variable. The most notably element arising from taking into account the institutional and historical differences between States lies at the beginning of the sample, in 1960: pre-*Civil Rights Act*, taking States into accounts lowers the racial wage gap, suggesting high hetero-geneity of treatments across States from Blacks relative to Whites. Otherwise, there is a non-significant difference between the specifications with and without State dummies.

Augmented Unconditional RWG Decomposition An increase in the number of explanatory variables is expected to lead to an increase of the goodness-of-fit of any statistical model, without consideration for parsimony. Figure 12 shows the evolution of the unexplained shares of the *unconditional* racial wage gap depending on model specifications: specifications with *Occupation* variables lowers the unexplained share of the RWG, while its dynamics during the crisis remains similar to the one of the baseline model, that is an increase of around 4 points between 2000 and 2012.

4   Conclusion

I have studied the distance between the medians of the log-annual wages of Blacks and Whites during the Great Recession, using *American Community Survey* data to correct for selection by incarceration. I showed that both the corrected and uncorrected estimator of the conditional racial wage gap reach
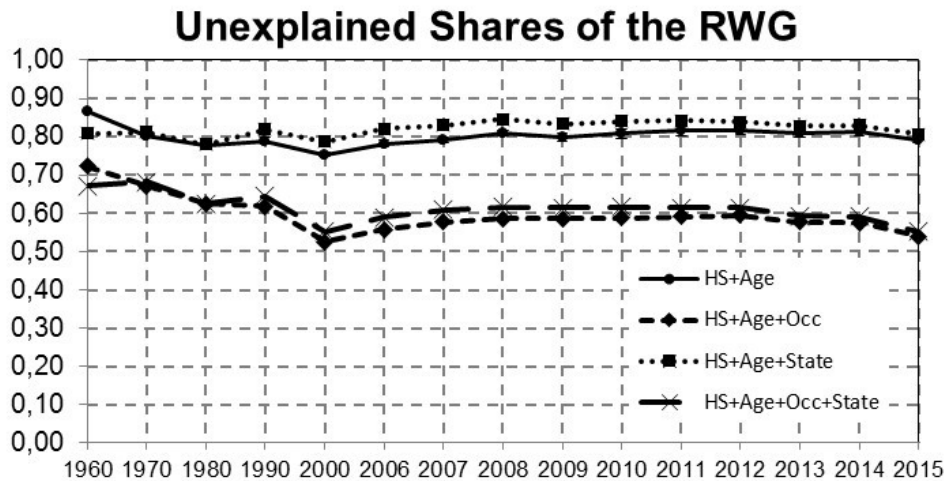
Figure 12: Unconditional Racial Wage Gap and Specifications

In 2008, the unexplained share of the unconditional racial wage gap was of 0.60 when computed with the full model.

a high in 2010, compared to the estimators obtained with Census data and the ACS between 1960 and 2015. When decomposing this unconditional racial wage gap, I showed that most part of this difference was due to the wage structure effect, which is the unexplained part of the racial wage gap. This unexplained share has increased slightly during the Great Recession. I obtain a similar result using three other specifications, taking into account geographical effects through State dummies and labour quality through occupation dummy variables. While the unexplained share of the racial wage gap is smaller with the use of occupation variables, its dynamics during the crisis remain unchanged compared to the baseline specification. On the contrary, the conditional racial wage gap is highly affected by a change in the model specification, yielding lower estimates with the use of occupation variables, and flat dynamics: no increase in the conditional racial wage gap occurs during the Great Recession using the individuals occupations as covariates.

References

Arellano, Manuel and Bonhomme, Stfffdfffdphane (2017). "Quantile Selection Models with an Application to Understanding Changes in Wage Inequality." *Econometrica* 85.1, pp. 1–28. URL: http://onlinelibrary.wiley.com/doi/10.3982/ECTA14030/full (visited on 05/16/2017).

Buonanno, Paolo (2003). *The Socioeconomic Determinants of Crime. A Review of the Literature*. Working Paper 63. University of Milano-Bicocca, Department of Economics. URL: https://ideas.repec.org/p/mib/wpaper/63.html (visited on 05/17/2017).

Chandra, Amitabh (2000). "Labor-Market Dropouts and the Racial Wage Gap: 1940-1990." *American Economic Review Papers and Proceedings* 90.2, pp. 333–338.

Chandra, Amitabh (2003). *Is the convergence of the racial wage gap illusory?* Tech. rep. National Bureau of Economic Research. URL: http://www.nber.org/papers/w9476 (visited on 09/29/2016).

Firpo, Sergio, Fortin, Nicole M., and Lemieux, Thomas (2009). "Unconditional Quantile Regressions." en. *Econometrica* 77.3, pp. 953–973. ISSN: 1468-0262. DOI: 10.3982/ECTA6822. URL: http://onlinelibrary.wiley.com/doi/10.3982/ECTA6822/abstract (visited on 05/09/2017).

Heckman, James J. (1979). "Sample Selection Bias as a Specification Error." *Econometrica* 47.1, pp. 153–161. ISSN: 0012-9682. DOI: 10.2307/1912352. URL: http://www.jstor.org/stable/1912352 (visited on 05/16/2017).

Meyer, Peter B. and Osborne, Anastasiya M. (2005). *Proposed category system for 1960-2000 census occupations*. US Department of Labor, Bureau of Labor Statistics, Office of Productivity and Technology.

Miller, Jerome G. (2011). *Search and destroy: African-American males in the criminal justice system.* 2nd ed. OCLC: ocn656771872. Cambridge ; New York: Cambridge University Press. ISBN: 978-0-521-76779-8 978-0-521-74381-5.

Neal, Derek and Rick, Armin (2014). *The Prison Boom and the Lack of Black Progress after Smith and Welch*. Tech. rep. National Bureau of Economic Research. URL: http://www.nber.org/papers/w20283 (visited on 09/29/2016).

Office of the Surgeon General (US), Center for Mental Health Services (US), and National Institute of Mental Health (US) (2001). *Mental Health: Culture, Race, and Ethnicity: A Supplement to Mental Health: A Report of the Surgeon General*. eng. Publications and Reports of the Surgeon General. Rockville (MD): Substance Abuse and Mental Health Services Administration (US). URL: http://www.ncbi.nlm.nih.gov/books/NBK44243/ (visited on 08/28/2017).

Rabuy, Bernadette and Kopf, Daniel (2015). *Prisons of Poverty: Uncovering the pre-incarceration incomes of the imprisoned*. URL: https://www.prisonpolicy.org/reports/income.html (visited on 05/20/2017).

Ruggles, Steven, Genadek, Katie, Goeken, Ronald, Grover, Josiah, and Sobek, Matthew (2015). *Integrated Public Use Microdata Series: Version 6.0*. en.

Smith, James P. and Welch, Finis R. (1989). "Black Economic Progress After Myrdal." *Journal of Economic Literature* 27.2, pp. 519–564. ISSN: 00220515. DOI: 10.2307/2726688. URL: http://www.jstor.org/stable/2726688.

Torrey, E. Fuller, Kennard, Aaron D., Eslinger, Don, Lamb, Richard, and Pavle, James (2010). "More mentally ill persons are in jails and prisons than hospitals: A survey of the states." *Arlington, VA: Treatment Advocacy Center*. URL: http://tulare.networkofcare.org/library/final_jails_v_hospitals_study1.pdf (visited on 08/28/2017).

Visher, Christy Ann, Debus, Sara, and Yahner, Jennifer (2008). *Employment after prison: A longitudinal study of releases in three states*. Urban Institute, Justice Policy Center. URL: http://www.

canatx.org/CAN-Research/Reports/2008/411778_employment_after_prison.pdf (visited on 09/29/2016).

Western, Bruce and Pettit, Becky (2002). "beyond crime and punishment: prisons and inequality." en. *Contexts* 1.3, pp. 37–43. ISSN: 1536-5042, 1537-6052. DOI: 10.1525/ctx.2002.1.3.37. URL: http://ctx.sagepub.com/lookup/doi/10.1525/ctx.2002.1.3.37 (visited on 05/17/2017).

Winters, John V. and Hirsch, Barry T. (2012). "An Anatomy of Racial and Ethnic Trends in Male Earnings." *Andrew Young School of Policy Studies Research Paper Series* 12-18. URL: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2139024 (visited on 09/29/2016).

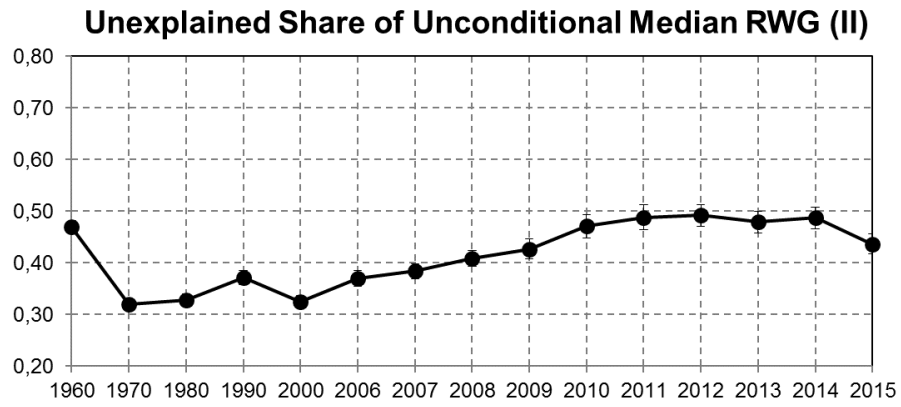**Unexplained Share of Unconditional Median RWG (II)**

Figure 13: In 2011, the unexplained share of the unconditional median Racial Wage Gap represented 59% of the total difference in median log-wages between Blacks and Whites.

5 Appendix