Working paper series

Discrimination and Monopsony Power

Mark Stelzner
Kate Bahn

July 2020

https://equitablegrowth.org/working-papers/discrimination-and-monopsony-power/

© 2020 by Mark Stelzner and Kate Bahn. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.
Discrimination and Monopsony Power

By Mark Stelzner and Kate Bahn

Abstract: Wage inequalities between identical workers of different race, ethnicity, and gender are a persistent feature of labor markets. However, most labor market models either ignore important empirical evidence or focus very narrowly on specific labor market dynamics. To better understand such wage differences, we create a labor market model that integrates firm competition for workers, employee movement between jobs in response to market signals, potential monetary frictions in the job transition process, and workers’ collective action which is a function of government support. Our model shows that because of gender and race specific characteristics, like the relatively lower household wealth of Black and Latino families and the increased household responsibilities of women, women and minority workers are more exploitable. Also, our model shows that the cumulative wage gap for non-white women is greater than the additive gaps of being non-male and non-white. Lastly, our model shows that a reduction in government support for workers in general enables employers to wield monopsony power more freely, independent of changes in employer concentration. Because certain groups are more exploitable, employers’ increased capability in wielding monopsony power means increased wage differentials replicating discriminatory biases against marginalized groups of workers.

1 Thanks to Nancy Folbre, Sydnee Caldwell, Terry-Ann Craigie, Todd Sorensen, and Douglas Webber for comments on the paper.
2 Mark Stelzner, Assistant Professor of Economics, Connecticut College, New London, CT USA. Email: mstelzn2@conncoll.edu.
3 Kate Bahn, Director of Labor Market Policy and Economist, Washington Center for Equitable Growth, Washington, D.C. USA. Email: kbahn@equitablegrowth.org.
Introduction

Disparate wage outcomes between like workers of different race, ethnicity, and gender remains one of the most persistent features of the labor market. For example, in 1979, after controlling for education, experience, and location, black women working full-time in the United States earned 69 percent of equivalent black men, and white women earned 62 percent of equivalent white men. While gender wage differences decreased between 1979 and the early 1990s, since then the female-male wage gap has not improved. In 2015, full-time black women still made only 84 percent of their black male counterparts, and white women made 78 percent of equivalent white males. In terms of wage differences between individuals of different races, in 1979 after controlling for education, experience, and location, a black male working full-time made, on average, 83 percent of an equivalent white male. Since then, wage difference between equivalent white and black workers have worsened. By 2015, the relative wage of a black male worker compared to an equivalent white male worker had decreased five percentage points (Daly, Hobijn, and Pedtke, 2017).

Similar results exist between Latino and non-Hispanic white workers. In 1979, after controlling for education, experience, and location of work, Latino men working full-time earned 83.5 percent of equivalent non-Hispanic white males. Since then, the wage gap between Latino male workers and non-Hispanic white male workers has not improved by much with Latino males making 85.1 percent of equivalent non-Hispanic white males in 2017 (Mora and Davila, 2018). In addition to these patterns, gender, race, and ethnicity interact to create intersectional wage gaps. For example, the wage gap faced by black and Latina women is greater than the sum of individual racial, ethnic, and gender wage gaps (Darity et.al., 2018; Bahn and McGrew, 2018).

How do we understand these wage differences? And how do we understand how they have changed over time? While a number of current labor market models highlight important features of persistent disparate pay levels, current theoretical literature only offers partial understandings. In this paper, we seek to provide a more complete answer to these questions through highlighting a number of important labor market dynamics and worker characteristics from empirical and theoretical literature and through constructing a labor market model which incorporates all of these elements. We expand upon the job search model pioneered by Burdett, Mortensen and Pissarides (Mortensen 1998) with literature on how racial and gender disparities in wealth impact labor market outcomes. As we will see, this model sheds light on the
mechanisms underlying racial, ethnic, and gender wage differences, how they change over time, and helps us better understand how to remedy them.

*The Literature*

Existing labor market models do not provide an adequate answer to the above questions. The predominant explanation in economics is human capital theory, which explains wage differences between workers as stemming from differences in productivity that can be measured by “capital” allotments such as years of education and years of work experience, as well as industry and occupation, which presumably have naturally differentiated levels of productivity. Thus, according to human capital theory, the wage gaps highlighted in the introduction are a result of differences in productivity not revealed in the data. If data on workers were better, there would be no differences in wages between like workers. Nobel prize-winning economist James Heckman (1998) makes this exact point: “most of the disparity in earnings between blacks and whites in the labor market of the 1990s is due to differences in skills they bring to the market, and not to discrimination within the labor market.” And, political scientist and historian, Abigail and Stephan Thernstrom (1997) editorialized the same point in the *Wall Street Journal*: “what may look like persistent employment discrimination is better described as employers rewarding workers with relatively strong cognitive skills.” Human capital theory explanations portray wage differentials based on presumed productivity differences as justified, since workers are paid equivalent to the value they produce, as moderated through competitive forces.

The “skills gap” narrative prevalent in policymaking, undergirded by the theory of skill-biased technical change (Autor, Katz, and Kearny, 2006), leads to the conclusion that solution for wage inequality should primarily be through increasing human capital with education and training. While the human capital theory explanation is pervasive in the economic profession and other spheres, it doesn’t seem to well fit with empirical data. For example, when controlling for skill level as measured by education, wage gaps grow for women and minority men (Barrow and Rouse, 2005; Gould, 2019). Likewise, those from low-income backgrounds have lower returns to education compared to those from higher-income backgrounds, resulting in lower lifetimes earnings on average (Bartik and Hershbein, 2018). Also, when decomposing wages of women by race and ethnicity compared to white men, the explanatory power of the human capital-based hedonic wage model declines for black and Latina women, with a larger portion “unexplained,”
or interpreted as the result of discrimination, compared to white women. This implies that increasing the human capital of non-white women will not be sufficient to bridge intersectional wage gaps.

Furthermore, numerous audit studies have shown that employer discrimination persists even after better controlling for productivity. For example, Pager (2003) and Pager, Bonikowski, and Western (2009) hired and trained black, white, and Latino testers and construct identities such to better control for productivity. The only exception to these identically constructed testers was that white testers were given a criminal record of serving 18 months for possession of cocaine with the intent to distribute. These testers were sent to hundreds of employers in Milwaukie and New York City; both studies found that minority applicants with no criminal record received positive responses from employers at essentially equivalent rates as whites just out of prison. Likewise, Gaddis (2015) performed a correspondence study by sending resumes with white and black sounding names for individuals with elite and less selective college degrees to jobs listed online. He found that the response rate for black individuals with an elite college degree was almost the same as the response rate for a white individual with a less selective degree, and of the black candidates that received calls, the salaries offers were ten percent lower on average than white candidates. And Bendick, Jackson, and Reinoso (1994) conducted a field experiment in which they sent out 149 job applicants for in-person interviews to observe employer responses for equally qualified candidates of different racial backgrounds, finding that treatment was less favorable for the African American applicants. Controlled field experiments have also demonstrated how employer racial biases persist and lead to discriminatory hiring outcomes.

Compensating wage differential theory is another popular explanation for wage differentials between like individuals of different gender groups. The explanation in this theory is that women of a given educational level are likely to choose jobs that have lower wages because of love for that line of work. In this situation, lower paid workers are compensated by higher utility from enjoyment of the work; thus, overall utility from work is the same between like workers although wages are different. While this is potentially an important factor, it doesn’t negate other dynamics, as we will show below. At the same time, the compensating wage differential explanation fails to explain wage differences between like workers inside a given industry – a prominent feature of labor markets.
A number of economists have sought to explain some part of these wage gaps through employer monopsony power, where workers are exploited by being paid less than the value of their marginal product (Manning, 2003). In the original application of monopsony, Robinson (1933; pp. 301 – 304) presents a wage discriminating monopsonist model where white and black workers have different supply elasticities, and the lower labor supply elasticity of black workers leads to a lower wage compared to white workers. In a similar line of research, Reich (1981; pp. 204 – 215) presents two models of white-black wage disparities. In the first model, employers pay white workers more than their marginal product in order to break up worker solidarity and pay black workers less than their marginal product. In the second model, parallel to Robinson, employers have monopsony power and white and black workers have different labor supply curves. Because the supply curve for black workers is less elastic, monopsonistic employers exploit them more. Paralleling Becker’s original analysis but adding a monopsonist twist, Black (1995) creates a monopsony model and assumes that some firms will not hire black workers. As a result, wage differentials both hurt discriminating firms and workers.

In terms of the gender pay gap and monopsony, Robinson explains lower wages for women compared to identical men as stemming from the higher rate of unionization of men compared to women. And most recently, Manning (2003) uses his generalized oligopsonist model to integrate the idea that differences in wages between men and women are due to the greater household constraints on the latter which reduces competition for female labor and increases the degree to which women are willing to trade-off wage growth for non-pecuniary benefits. Both dynamics increase the degree to which a monopsonist employer can exploit female workers.4

All of these monopsony models highlight an important dynamic: employers seem to have significant wage setting power. For example, there is evidence that monopsony power exists in nursing (Hurd, 1973; Link and Landon, 1975; Bruggink et. al., 1985; Sullivan, 1989; Staiger, Spetz, and Phibbs, 2010; Prager and Schmitt, 2019), in professional sports (Bodvarsson and Brastow, 1999; Bodvarsson and Pettman, 2002; Kahn and Shah, 2005), and in manufacturing (Benmelech et al., 2018). There is even evidence that employers can push the wage below the marginal product in jobs contracted through the internet that are completed remotely (Dube et al., 2018).

4 By exploitation, we refer to Robinsonian exploitation – i.e. where the wage is lower than the marginal product of the worker.
Wage setting power seems to be wielded by the largest employer in the United States, Walmart (Dube et al., 2007), but also seems common among companies that command less employees. Indeed, it even seems to be wielded by the government in the education sector (Landon and Baird, 1971; Ransom and Sims, 2010; Ransom and Lambson, 2011; Falch, 2011). 

At the same time, these models highlight a number of other important dynamics like the ability of employers to exploit the increased household obligations of women compared to men and the less elastic labor supply curves of black workers. In terms of the first point, Webber (2016) finds that women’s lower labor supply elasticity, across industries in the United States, leads to 3.3 percent lower earnings, all else equal. This is reinforced by industry studies, where research on K-12 teachers in Missouri finds that the gender wage gap is replicated in the education sector, despite rigid pay structures, due to men being more likely to sort into higher paying school districts as a result of their relatively higher labor supply elasticity (Ransom and Lambson, 2011).

In terms of the relative elasticities of the labor supply curves for black, Latino, and white workers, there is considerable empirical evidence that shows that black workers’ labor supply is more elastic than white workers’ labor supply. Given a monopsonistic employer that could wage discriminate, this would mean that white workers would be more exploited. However, Reich (1981) and Seltzer and Wrigley-Field (2020) explain that this empirical data is flawed because black workers are more likely to be fired when labor demand decreases. Thus, statistics on comparative elasticity of labor supply confound worker exit in search of higher wages or movement out of the labor force with employers firing workers because of a decrease in labor demand. Likewise, despite the passage of the Civil Rights Act of 1968, Darity and Mason (1998) survey the literature to demonstrate the persistence of discriminatory hiring practices along the lines of race and gender, which would lead to the intuitive conclusion that women and minority men would be less likely to receive job offers in their search compared to other workers, all else equal.

To be sure, there is significant reason to think many non-white workers might have less elastic labor supply curves that would impact job matching and result in discriminatory wage outcomes. For example, wealth is important in overcoming the many potential monetary obstacles that can confront a worker attempting to move between jobs. Potential monetary 

---

5 For a survey of current empirical work on the existence of monopsony power, see Naidu et al. (2018).
obstacles could be as small as temporarily forgone wages resulting from the need to take days off to interview or a delay in pay when one transitions between jobs – a cost which low-wealth families may not be able to bear. Obstacles could potentially be much larger, involving a period of unemployment because the initial transition to the new job didn’t go as planned. Thus, extreme wealth inequality between white, black, and Latino families, as well as between women and men, in the United States would mean that like workers from different racial and gender groups have different ease and thus ability to navigate labor markets. As a result, it is likely that women and minority men are less sensitive, and rightly so, to wage differences between their job and others when the cost of foregone wages is too high for themselves and their families.

In terms of wealth inequality, using data from the Survey of Consumer Finances, the Urban Institute calculates that the average financial wealth of a black family in the United States in 2016 was $139,523 – 84.8 percent less than that of the average white family. The median financial wealth for a black family in the same year was $17,409 – 89.89 percent less than that of the median white family.\(^6\) In certain places, the disparity between white and black family wealth is even larger. For example, Muñoz et al. (2015) find that in Boston the median net wealth of black families is only $8 – while the median net worth of white families is $247,500. While research clearly demonstrates the low median wealth of black families, Chiteji and Hamilton (2002) also find that middle class black families are not buffered against the impacts of poverty, and to a greater extent than middle class white families. Extended family networks with siblings and other relatives are in poverty reduce the ability of black households to accumulate wealth as well.

Similar patterns are replicated for Latino families, whose median net worth was $20,700 in 2017, compared to a median net worth of $171,000 for white families in the same year (Dettling et. al., 2017). Following the Great Recession, both Latino and black families’ net worth also took longer to recover from the financial shock. While white families began to see positive growth in their net worth between 2010 and 2013, Latino and black families continued to have decreasing net worth in this period, only beginning to recover between 2013 and 2017 (Solomon and Weller, 2018).

While the gender wealth gap is harder to empirically measure due to shared household assets in heterosexual couples, women also face significant obstacles in their financial security.

\(^6\) For example, see https://apps.urban.org/features/wealth-inequality-charts/.
that could make it more difficult to leverage wealth to search for a job. Although variations in stock market participation is often given as one explanation for women’s lower levels of wealth, women also face additional risk in the labor market and housing markets as well as through caregiving responsibilities, resulting in a greater propensity for negative shocks to their assets (Weller and Tolson, 2017). Similarly to the multiplicative effect of gender and race on the gender wage gap facing black women, research also suggests that they face multiplicative barriers to accumulating wealth by retirement years due to the compounding effects of labor market discrimination, state policies that exacerbate racial inequality, residential segregation, and disparities in health outcomes (Brown 2011). Reduced opportunities in the labor market further exacerbate women’s ability to save earnings and grow assets, which ultimately results in lower levels of economic well-being throughout the life cycle (Ruel and Hauser, 2013; Denton and Boos, 2007).

While Robinson (1933), Reich (1981), Black (199), and Manning (2003) highlight important dynamics in explaining racial and gender wage differences, they are also incomplete. For example, Robinson and Reich start from the firm level labor supply curves. Thus, differences in labor supply between individuals of different racial groups has to be abstractly explained instead of derived, and changes in firm level competition for labor also has to be abstractly attributed to the firm level labor supply curve instead of derived from competition in wages between firms. This level of abstraction reduces the potential theoretical understanding the model can provide.

In contrast, using a version of Burdett and Mortensen (1998), Manning (2003) explicitly models the job search process for employed and unemployed workers. While labor market frictions give employers monopsony power, the form used by Manning does so in a curious way. In the model, employers randomly send out job offers to a small group of workers. If the offer is greater than the worker’s current wage or reservation wage depending on if she is employed or unemployed, respectively, the worker accepts the offer. The tunnel vision imposed on workers through only seeing job offers from a few employers and the inability to apply themselves gives the employer monopsony power. This setup both seems unnecessary, except maybe in creating equilibrium wage dispersion, and unrealistic. Contrary to the dynamic presented in Manning, firms rarely, except maybe for very elite positions, seek out specific workers that have not applied for a job. Second, by not incorporating frictions in the job search and change process for
workers, the model presented by Manning is not able to explain differential monopsony power stemming from differences in worker’s ability to confront job market frictions as a result of different levels of wealth. Potentially, this is why Manning thought that monopsony power was less important for explaining black-white wage differences.

Also, the models in Robinson (1933), Reich (1981), Black (199), and Manning (2003) cannot explain why the actual wage gap for black and Latina women, as revealed by empirical work, is higher the additive wage gap between black and white men and white women and men. The reason for their inability to explain this empirical outcome is because none of the models integrate the unique socially salient identities that emerges at the intersection of gender and racial and ethnicity, resulting in dynamics that increase the exploitability of these groups.

Lastly, these models do not well explain the change in wage differences between equivalent workers over time. For example, from Robinson (1933) and Reich (1981), change in wage differences between workers of different race and ethnicity could be attributed to change in the relative elasticity of labor supply of black workers. However, like explained above, these changes would have to be exogenous changes as opposed to endogenously derived in the model. In terms of Manning (2003), changing gender wage differences could only come from a reduction in the relative mobility of women. To be sure, change in relative mobility is important. For example, following the Women Right’s Movement and associated policy and cultural shifts in the 1970s, including expanding access to contraception that delayed women’s childbearing, the gender wage gap converged to women earning 70 percent of equivalent men by the 1990s and just under 80 percent by 2000 (Bailey, Hershbein and Miller 2012). However, it doesn’t seem to be the only dynamic at play.

An important dynamic which is missing from all of these models is that workers can act as an important countervailing force to stymie employer monopsony power. However, workers’ ability to do so is dependent on institutional support for collective action. When institutional support for unions, strikes, and other forms of collective action doesn’t exist, employers are able to wield monopsony power more freely, independent of changes in employer concentration. Because certain groups are more exploitable due to decreased mobility resulting from greater household obligations or a decreased ability to weather job market shocks, employers’ increased capability in wielding monopsony power means increased wage differentials.
There is considerable evidence that worker power, in general, has decreased over the last four decades as a result of changes in judicial understanding and administration of labor laws. For example, Stelzner (2017) shows that the National Labor Relation Act (NLRA), the main federal labor law in the United States, has been reinterpreted to the detriment of workers by the courts and the National Labor Relations Board (NLRB), the body charged with administering the NLRA. Additionally, through decreases in funding, dramatic changes in the mindset of those heading the administration, and increased outside political pressure, the NLRB has significantly increased the time it takes to decide contested cases. For example, during the 1960s and 1970s, the median number of days the NLRB in D.C. took to respond to contested unfair labor practice cases averaged 131 days. In the 1980s, response times jumped to 258 days. And by the 2000s, the median number of days the NLRB in D.C. took to respond to contested unfair labor practice cases jumped to 486. Unfair labor practices are actions taken by an employer or union that are deemed illegal by the NLRA. For example, firing a worker for participating in an economic strike or participating in union activity is an unfair labor practice. Clearly, if it takes more than a year for NLRB to issue its final decision on an unfair labor practice, engaging in such actions is a very effective strategy for scaring workers in the short-term.

Additionally, there have been dramatic changes in state level labor laws. For example, a number of states have recently passed right-to-work laws — prohibitions of stipulations in union-employer-employee contracts that require workers to join the union, or at least pay some part of the union dues, in order to stay employed at the job. Right-to-work laws also represent “ideological onslaught of the first order” sending notice to workers that government supports employers’ prerogatives in the workplace (Licthenstein, 2013; pp. 117-18). Louisiana passed a right-to-work law in 1976; Idaho in 1985, Oklahoma in 2001, Indiana in 2012, Michigan in 2013, Wisconsin in 2015, West Virginia in 2016, and Kentucky and Missouri in 2017. And, the Supreme Court, in a recent decision, seems to have adjudicated a federal right-to-work for public sector unions for the entire United States.7

Sadly, the above exposition doesn’t exhaust the new role of government in labor markets. Indeed, there has been a number of other changes affecting worker power. For example, the federal real minimum wage has been allowed to decrease considerably from its high in the late

---

Franchise law has been remade to allow franchisors to exert more power on franchises and ultimately the workers at the ground level (Callaci, 2020), and antitrust laws has been increasing used against workers instead of businesses – even though firm concentration has increased dramatically over the last four decades and labor unions and strikes have plummeted (Vaheesan, 2018).

A number of studies have shown that these changes in the orientation of government have had considerable effects on workers. For example, many have found that right-to-work laws have a statistically significant negative effect on collective action (Carroll, 1983; Ellwood and Fine, 1987; Davis and Huston, 1993; Garofalo and Malhotra, 1992; Gould and Shierholz, 2011; Stelzner, Hoyt, and Ramchurn, 2019). And Stelzner, Hoyt, and Ramchurn (2019) show that changes in adjudication and administration of the NLRA and change in social norm around employers’ using permanent replacement workers during economic strikes explain much of the fall in worker power since 1980.

This reduction in worker power has enabled employers to wield monopsony power more freely, independent of changes in employer concentration (Stelzner and Paul, 2019). These changes in institutions which frame the wage setting process are important for understanding changes in the degree to which employers can discriminate. Because certain groups are more exploitable due to decreased mobility resulting from greater household obligations or a decreased ability to weather job market shocks, employers’ increased capability in wielding monopsony power means increased wage differentials. This same dynamic of the protective power of formal labor institutions is revealed in a recent paper on the extension of the minimum wage to agricultural, food service, and nursing home jobs in 1966. Because this extension of the minimum wage reduced firms’ ability to wield monopsony power and because almost one third of black workers were located at firms in such industries, the wage gap between black and white low-wage workers fell considerably in the late 1960s and 1970s (Derenoncourt and Montialoux, 2019).

In order to better understand wage differences between like workers of different race, ethnicity, and gender and how these differences change over time, in the following section, we create a job market model that incorporates all of the dynamics highlighted above. Firm level competition over labor is endogenized. Labor market dynamics pivot around firms’ posted wage and workers’ decisions to respond to market signals, and worker power, which is a function of
institutional support for collective action, can mitigate employer’s monopsony power. As we will see, integrating these dynamics yield a labor market model that can endogenously explain wage differences, how they change over time, and help us better understand how to remedy these persistent wage disparities.

The Model

Imagine there are \( L \) workers supplying their labor and \( N \) firms demanding it. Initially, workers are spread evenly between firms. Thus, each firm starts with \( \frac{L}{N} \) workers. However, workers can move between firms to take advantage of wage differentials. The probability that worker \( i \) will leave firm \( j \) to get a job at another firm is given by \( \theta_{ij} \):

\[
\theta_{ij} = \begin{cases} 
0, & w_j > \frac{1}{N} \sum_{k=1}^{N} w_k \\
\alpha_i \left( \frac{1}{N} \sum_{k=1}^{N} w_k - w_j \right), & \frac{1}{N} \sum_{k=1}^{N} w_k \geq w_j \geq \frac{1}{N} \sum_{k=1}^{N} w_k - \frac{1}{\alpha_i} \\
1, & w_j < \frac{1}{N} \sum_{k=1}^{N} w_k - \frac{1}{\alpha_i}
\end{cases}
\]  

(1)

\( w_j \) is the wage paid to worker \( i \) at firm \( j \), and \( \frac{1}{N} \sum_{k=1}^{N} w_k \) is the average wage for like workers at all \( N \) firms. \( \alpha_i \) is a measure of how sensitive worker \( i \) is to differences in wages. For a higher value of \( \alpha_i \), worker \( i \) is more likely to change jobs given even a small wage difference between his current wage and that paid to like workers at other firms. If \( \alpha_i \) is infinite, worker \( i \) would behave in the way assumed in perfect competitive – leaving his job even if the wage he was receiving is infinitesimally lower that of workers at other firms. As explained above, \( \alpha_i \) is a function of wealth – i.e. \( \alpha_i = f(\phi_i) \) where \( \phi_i \) represents the level of worker \( i \)’s wealth. Everything else equal, for higher levels of wealth, workers are better able to weather monetary shocks from transitioning between jobs. Thus, workers with greater wealth are more sensitive to wage differentials – i.e. \( \frac{\partial \alpha_i}{\partial \phi_i} \) is greater than zero.

Using the initial firm level labor supply explained above and the probability that a worker will leave the firm for another job, we can construct an expected labor supply curve for firm \( j \),
If we assume that all workers are equally sensitive to wage differentials, the expected labor supply curve facing firm $j$ takes the following form:

$$E[S_j] = \frac{L}{N} (1 - \theta_{ij}) = \frac{L}{N} \left( 1 - \alpha_i \left( \frac{1}{N} \sum_{k=1}^{N} w_k - w_j \right) \right) \quad (2)$$

Because of the symmetry, $1 - \theta_{ik}$ represents both the probability that a worker stays at firm $j$ and the probability that other workers will move to firm $j$ given its wage differentials. Thus, $\frac{L}{N} (1 - \theta_{ij})$ is the expected labor supply for a firm with an initial labor supply of $\frac{L}{N}$ workers.

By taking the derivative of $E[S_j]$ in terms of the wage, $w_j$, we can see that this setup yields a firm level labor supply curve that is upward sloping:

$$\frac{\partial E[S_j]}{\partial w_j} = \frac{\alpha_i (N - 1)L}{N^2} > 0 \quad (3)$$

If there is more than one firm (i.e. $N > 1$), when firm $j$ increases the wage it pays its workers a small amount, its expected labor supply increases. Likewise, when firm $j$ decreases the wage, its expected labor supply decreases. If there is only one firm, the labor supply to the firm is fixed at $L$, and no matter the number of firms, the overall labor supply across all $N$ firms is also always fixed at $L$.

Given the assumption that firms are rational and want to maximize profits, we can now determine firm hiring and production behavior. In order to focus on the wage dynamic, let’s assume that the firm faces a perfectly elastic demand curve for its output at the market price of one, and that each worker produces one unit of output. Thus, the profit function for firm $j$ takes the following form:

$$\pi_j = (1 - w_j)E[S_j] \quad (4)$$

The first order condition for the firm is the following:
\[
\frac{\partial \pi_j}{\partial w_j} = \frac{\alpha_i(N-1) L}{N^2} - \frac{\alpha_i(N-1) L w_j}{N^2} - \frac{L}{N} \left( 1 - \alpha_i \left( \frac{1}{N} \sum_{k=1}^{N} w_k - w_j \right) \right) = 0
\]

\(\frac{\alpha_i(N-1)L}{N^2}\) is the marginal benefit from increasing the wage – the increase in revenue from the increase in expected labor supply and thus production from an increase in the wage.

\(\frac{\alpha_i(N-1)L w_j}{N^2} + \frac{L}{N} \left( 1 - \alpha_i \left( \frac{1}{N} \sum_{k=1}^{N} w_k - w_j \right) \right)\) is the marginal cost – the increase in the wage bill from hiring a new worker and from the change in wages for all other workers. The firm is maximizing where the marginal benefit is equal to the marginal cost, i.e. where \(\frac{\partial \pi_j}{\partial w_j} = 0\).

The best response function for the firm takes the following form:

\[
w_j^* = \frac{1}{2} - \frac{N}{2 \alpha_i (N-1)} + \frac{1}{2} \frac{N}{(N-1)} \sum_{k=1, k \neq j}^{N} w_k
\]

Because all of the \(N\) firms’ best response functions take the form in equation (6) if their profit function is represented by equation (4), at the Nash equilibrium, the wage each of the \(N\) firms choose is the same. Thus, \(w_1\) is equal to \(w_2\) which is equal to \(w_N\). As a result, at the Nash equilibrium, the wage is the following:

\[
w^* = 1 - \frac{N}{\alpha_i (N-1)}
\]

As we can see, at the Nash equilibrium, the wage is normally less that the value of the marginal product for any given worker. Firms are able to exploit workers using their wage setting power. Only if \(\alpha_i\) goes to infinity is the wage setting power of firms eliminated and the Nash equilibrium wage goes to the value of the marginal product. Thus, the perfectly competitive case can be derived from this model but it is clearly an extreme situation.

The equilibrium wage share the following relationships with the degree to which workers react to wage differentials, \(\alpha_i\), and to the number of firms demanding labor, \(N\):
\[
\frac{\partial w^*}{\partial \alpha_i} = \frac{N}{(N - 1)\alpha_i^2} > 0
\] (8)

\[
\frac{\partial w^*}{\partial N} = \frac{1}{\alpha_i(N - 1)^2} > 0
\] (9)

Thus, an increase in the degree workers react to wage differentials increases the equilibrium wage, and an increase in the number of firms competing in a given labor market increases the equilibrium wage.

Consequently, equation (9) shows that if, on average, female workers are more likely to limit the geographic extent of their potential labor supply because of greater household responsibilities, the number of firms that compete for their labor, \( N \), will be less. If firms are able to wage discriminate and recognize that women, on average, have less demand for their labor, they would be able to exploit female workers more than male workers – creating a gender wage gap. Note, that this dynamic would reinforce misogyny; unless the higher level of exploitation is consciously recognized, female workers would seem to be less productive than male workers. Also, employers that arrive at \( w^* \) through trial and error as opposed to maximization of their explicit profit function could easily understand wage differences as stemming from differences in productivity of male and female workers.

From equation (8) and the relationship between wealth and sensitivity to wage differentials, we can see also that black and Latino labor would be more exploitable than white labor:

\[
\frac{\partial w^*}{\partial \phi_i} = \frac{N \frac{\partial \alpha_i}{\partial \phi_i}}{(N - 1)\alpha_i^2} > 0
\] (10)

As explained above, \( \frac{\partial \alpha_i}{\partial \phi_i} \) is greater than zero; more wealth makes workers more responsive to wage differentials because they are better able to weather potential monetary shocks in the job change process. Thus, if employers are able to wage discriminate and they recognize that black and Latino workers have much less wealth than white workers on average, they can more intensely exploit black and Latino workers – creating a wage gap between workers of different
racial and ethnic groups. Note, that this dynamic would reinforce racism; unless the higher level of exploitation is consciously recognized, black and Latino workers would seem to be less productive than white workers. Also, employers that arrive at $w^*$ through trial and error as opposed to maximization of their explicit profit function could easily understand wage differences as stemming from differences in productivity of black, Latino, and white workers.

From equations (7), (8), (9), and (10), we can analyze the cumulative wage gap for black and Latina female worker. As mentioned above, empirical evidence shows that the wage gap for individuals that are both non-male and non-white are greater than the sum of individual racial, ethnic, and gender wage gaps. Our model yields the same results. The additive and cumulative reduction in wage for a black female worker would be the following:

\[
\text{Additive: } \frac{\partial w^*_{wm}}{\partial N} + \frac{\partial w^*_{wm}}{\partial \alpha} = \frac{1}{\alpha_w (N_m - 1)^2} + \frac{N_m}{(N_m - 1)\alpha^2_w} \tag{11}
\]

\[
\text{Cumulative: } \frac{\partial w^*_{wm}}{\partial N} + \frac{\partial w^*_{wf}}{\partial \alpha} = \frac{1}{\alpha_w (N_m - 1)^2} + \frac{N_f}{(N_f - 1)\alpha^2_w} \tag{12}
\]

The subscript ‘w’ for the variables in equations (11) and (12) represents white, and the subscript ‘f’ and ‘m’ represent female and male, respectively. Thus, $\frac{\partial w^*_{wm}}{\partial N}$ represents the difference in wage between a white male and a white female because the derivative is in terms of $N$ – representing the change in geographic extent of labor markets for men and women because of different average household responsibilities. $\frac{\partial w^*_{wm}}{\partial \alpha}$ is the difference in the wage between a white and a black male because the derivative is in terms of $\alpha$ – representing differences in average household wealth between white and non-white households which affects individuals ability to weather potential negative labor market outcomes. And $\frac{\partial w^*_{wf}}{\partial \alpha}$ represents the difference in the wage between a white and black female.

The first value in equations (11) and (12), $\frac{1}{\alpha_w (N_m - 1)^2}$, are the same. Thus, the difference between the additive and cumulative effect pivot on the second value. From simple manipulation, we see that the cumulative effect is larger, because $\frac{\partial w^*_{wm}}{\partial \alpha}$ is greater than $\frac{\partial w^*_{wf}}{\partial \alpha}$, i.e.
The penalty for being non-white is larger for smaller values of $N$. Thus, if you are female and, as a result, household responsibilities limit the geographic extent of your labor market more than the average male, the penalty for also being non-white is larger than for a male. Likewise, the penalty for being non-male is larger for smaller values of $\alpha$.

To better conceptualize these differences, in Figure 1 below, we graph the Nash equilibrium wage, $w^*$, as a function of workers’ sensitivity to wage differentials, $\alpha$. $VMP$ stands for the value of a worker’s marginal product – the wage in perfect competition. $\frac{\Delta w^*_{wm}}{\Delta \alpha}$ and $\frac{\Delta w^*_{wf}}{\Delta \alpha}$ are $\frac{\partial w^*_{wm}}{\partial \alpha}$ and $\frac{\partial w^*_{wf}}{\partial \alpha}$ in discrete form. $w^*(N_m)$ and $w^*(N_f)$ represent the Nash equilibrium wages, i.e. equation (7), for male and female workers, respectively, as represented by the subscript on the number of firms competing for their labor.

**Figure 1: Visualizing wage penalties**

As show in equation (8) and depicted in Figure 1, increases in the $\alpha$ lead to increases in the Nash equilibrium wage; in order to retain workers better able to weather potential negative shocks in the job search process, employers are forced to offer higher wages. Likewise, in Figure 1, we can see that the cumulative effect is bigger than the additive; $\frac{\Delta w^*_{wf}}{\Delta \alpha}$ is greater than $\frac{\Delta w^*_{wm}}{\Delta \alpha}$. The wage penalty for being non-white, or female, is increasing as a function of other
aspects that create marginalization – i.e. being female or non-white, respectively. Thus, our model shows that a black or Latina female will see a greater wage gap than the sum of individual gaps between a black and white male and between a white female and male.

**Changing Wage Gap**

How has the wage gap changed over time? The above model can only explain a change in the wage gap from changing gender roles at the home, and increased wealth inequality between racial, ethnic, and gender groups. While those dynamics definitely offer some explanatory power, it seems like other factors are also at work. As explained above, decreased institutional support for collective action in the wage setting process has increased employers’ ability to wield their monopsony power, independent of changes in employer concentration. Because some workers are more exploitable for the reasons highlighted in the previous two sections, an increase in employers’ ability to wield their monopsony power would lead to an increase in the wage gap between white and black workers, white and Latino workers, and male and female workers.

In order to model this dynamic, imagine workers have the following utility function, $U$:

$$U = Aw + B(\mu - w)\sigma - \frac{\sigma^2}{2\varepsilon}$$  \hspace{1cm} (13)

$Aw$ is the positive utility felt from earning wage, $w$. $A$ is a constant which represents a worker’s preferences for wages. $\sigma$ is the intensity of collective action. $\sigma$ ranges from zero, no collective action, to one, the most intense level of collective action. $\mu$ is the worker’s perceived deserved wage. $B$ is a positive constant specific to the worker representing the degree to which the worker feels a positive utility from confronting a perceived wage injustice. Thus, $B(\mu - w)\sigma$ represents the positive utility felt from engaging in collective action when a worker is payed less than their perceived deserved wage, $\mu - w$.

$\frac{\sigma^2}{2\varepsilon}$ represents the disutility from engaging in collective action. This disutility stems from the many potential costs to engaging in collective action – ranging from the opportunity cost of one’s time to the psychological stress of confronting the individuals in charge of one’s source of income to other potentially very negative outcomes. $\varepsilon$ is the level of institutional support for
workers in the wage setting process with zero representing no institutional support for workers, and higher levels of $\varepsilon$ representing more support for workers. Higher levels of support shields workers from the potential costs to engaging in collective action. For example, a NLRB that more quickly decides contested unfair labor practice cases acts as a stronger deterrent to employers from committing unfair labor practices – like firing workers for supporting unionization or striking.

Thus, workers’ utility maximizing intensity of collective action takes the following form:

$$\sigma = B(\mu - w)\varepsilon$$

(14)

As we can see from equation (14), the worker’s utility maximizing intensity of collective action is increasing in institutional support for collective action in the wage setting process – i.e. $\frac{\partial \sigma}{\partial \varepsilon}$ is greater than zero. Increased support for workers shields them from more of the costs of engaging in collective action and thus increases the utility maximizing intensity of collective action, everything else constant. Also, the utility maximizing intensity of collective action is decreasing in the wage – i.e. $\frac{\partial \sigma}{\partial w}$ is less than zero. An increase in the wage decreases the perceived injustice and thus the benefit a worker feels from confronting an employer.8

Imagine a firm’s total output is proportionally reduced by one minus the level of collective action. Zero collective action would lead to the same amount produced as in the previous model. When collective action is non-zero, output is decreased. Indeed, the decrease in output is the leverage workers place on employers to better their situation. The profit function would take the following form:

$$\pi_j = (1 - \sigma)E[S_j] - w_jE[S_j]$$

(15)

As we can see from equation (15), collective action decreases output. If collective action is zero, equation (15) reduces to equation (4). Given equations (2) and (14) and the symmetry of firms explained above, the Nash equilibrium wage would take the following form:

8 Even though they define the costs and benefits to collective action differently, Stelzner and Paul (2019) find a best response for workers’ collective action with the same relationship with government support for workers, i.e. $\frac{\partial \sigma}{\partial \varepsilon} > 0$, and with the wage, i.e. $\frac{\partial \sigma}{\partial w} < 0$. 

19
\[ w^* = \frac{1 - \mu B \varepsilon}{1 - B \varepsilon} \frac{N}{\alpha_t(N - 1)} \]  (16)

When there is no institutional support for collective action in the wage setting process and thus workers engage in zero collective action, equation (16) reduce to equation (7). When there is institutional support, the equilibrium wage has the following relationship with \( \varepsilon \):

\[ \frac{\partial w^*}{\partial \varepsilon} = \frac{B(1 - \mu)}{(1 - \varepsilon)^2} > 0 \]  (17)

An increase in institutional support for collective action increases the utility maximizing level of collective action. As a result, employers increase the wage in order to limit the disruption to production. Thus, more institutional support for workers allows workers to act as a counter force to employers’ monopsony power. If institutional support for collective action is decreased, employers are freer to wield monopsony power. Although modeled differently, this is the same result found by Robinson (1933), Galbraith (1968), Stelzner and Paul (2019), and many others.

As a result of the dynamic highlighted in equation (17), a decrease in the level of institutional support for workers would increase inequality between equally productive workers:

\[ \frac{\partial}{\partial \varepsilon} \frac{w^*_b}{w^*_w} = \frac{B(1 - \mu)}{w^*_w(1 - \varepsilon)^2} \left(1 - \frac{w^*_b}{w^*_w}\right) > 0 \]  (18)

\( \frac{w^*_b}{w^*_w} \) is the wage for black workers divided by the wage for white workers. Where the only differences in determination of the wage is \( \alpha_b \) for black workers is less than \( \alpha_w \) for white workers as a result of relatively lower household wealth, on average, for the former group of workers. As we can see from equations (17) and (18), a decrease in institutional support for workers would decrease the degree to which workers contest the monopsony power of employers. Because certain workers are more exploitable than others, for example black workers because of relatively lower household wealth on average, an increased ability to wield monopsony power would mean an increased wage gap between workers from groups that are
susceptible to different levels of exploitation. Thus, we can see that decrease in support for workers highlighted above can explain theoretically the increase in the wage gap between black and white workers.

In terms of the wage gap between women and men, a decrease in institutional support for collective action would have the same effect as with the racial wage gap:

\[
\frac{\partial w_f^*}{\partial \varepsilon} = \frac{B(1 - \mu)}{w_m^*(1 - \varepsilon)^2} \left(1 - \frac{w_f^*}{w_m^*}\right) > 0
\] (19)

\(\frac{w_f^*}{w_m^*}\) is the wage for female workers divided by the wage for male workers. Where the only differences in determination of the wage is \(N\) for female workers is less than \(N\) for male workers as a result of greater household responsibilities. As we can see, a decrease in institutional support for collective action, as highlighted above, would increase the ability of employers to wield monopsony power and thus increase the wage difference between female and male workers as a result of the former group being more exploitable because of greater household constraints.

As noted above, the female-male wage gap decreased between 1979 and the early 1990s and then has remained nearly unchanged since. Given the evidence on the decline in worker power from the 1980s on and the effect of the Women Right’s Movement Civil Rights Movement and associated policy and cultural shifts in the 1970s, this model would infer that the changes from the latter outweighed the former between the 1980s and early 1990s. However, since the early 1990s, this model would infer that the changes in worker power have neutralized any positive effects from policy changes and cultural shifts around gender roles at the home and at work.

**Conclusion**

As we have shown above, wage differences between like workers of different race, ethnicity, and gender, and the change in these wage differences overtime, can be explained theoretically by explicitly modeling firm competition for workers through wage posting, employee movement between jobs in response to market signals, differential household
responsibilities for male and female workers, differential response to potential monetary frictions in the job transition process due to racial and ethnic wealth inequality, and workers’ collective action which is a function of institutional support for collective action. These gender and race specific labor market friction give employers more power over women and black and Latino workers and thus create racial, ethnic, gender, and intersectional wage gaps. Additionally, the reduction in support for labor over time has made it increasingly possible for employers to push down the wage of these more exploitable groups.

In terms of remedying wage gaps between like workers of different race, ethnicity, and gender, the model shows that individual strategies, such as those implied by the human capital model, would most likely be ineffective. As explained above, the dynamics that create these wage differences make misogyny and racism profit maximizing because offering women, black, and Latino workers lower wages because of believed lower productivity would parallel the rational profit maximizing strategy given the increased exploitability of these groups. Thus, the market would not eliminate discrimination as argued by Becker (1957) and others. Instead, misogyny and racism would thrive in a monopsonist labor market. For the individual worker trying to combat her wage difference, increasing the geographic scope of potential jobs or increasing the degree to which she risks potential monetary shocks from the job market would not affect the treatment received by misogynistic and racist employers.

Also, for workers facing employers following the rational, profit maximizing strategy laid out above, it would be difficult for a worker to convey their individual difference from the group in terms of exploitability. And if the worker could only convey it to her actual employer, the employer might, consciously or subconsciously, realize that other employers are most likely unaware of the differences in exploitability of this individual. Thus, it might still be maximizing for the employer to offer the worker the lower wage – forcing her to accept the wage difference or quit and weather the job market where other employers will assume she is more exploitable.

In contrast, the model outlined above shows that policies which combat wage discrimination, wage setting power, gender household roles, and household wealth inequality would be effective in reducing wage differentials between like workers of different race, ethnicity, and gender. In terms of combating wage discrimination, such policies would reduce employers’ ability to profit from differential levels of exploitability through paying more exploitable workers lower wages. Indeed, we effectively followed this policy in the United
States with the Civil Rights Act of 1964 (Card and Krueger, 1993). However, since the 1980s we have moved away from actively combating wage discrimination. More recently, the Trump Administration halted an Obama Administration executive order that would increase the ability of the Equal Employment Opportunity Commission to gather more detailed and long-term firm-level data on wage levels by race and gender.

In terms of combating wage setting power more generally, as explained above, this can be done through increasing institutional support for workers. The countervailing power created by workers’ collective action reduces employers’ monopsony power and thus their ability to exploit workers generally and to more intensely push down the wage of more exploitable workers. For example, looking narrowly at the nursing occupation where women are significant majority of the field and black women are over-represented among low-wage nursing occupations, McGregor (2013) finds that the black-white wage differential between registered nurses all but disappears among unionized RNs. However, as shown above, since the 1980s, we have moved in the opposite direction in terms of supporting workers in the United States, with deleterious effects on the racial, gender, and intersectional wage gaps. Policies that increase the ability of unions to organize and bargain, such as sectoral bargaining and the repeal of Right-to-Work laws (Labor and Worklife Program, Harvard Law, 2020), would provide a greater countervailing power against exploitation.

At least a portion of the search frictions experienced by women are a result increased caregiving responsibility within families and the gendering of paid caregiving work as feminine. A suite of family economic security policies, including universal access to paid family and medical leave, paid sick leave, affordable childcare, improving transportation, would all have a disproportionate effect on women workers who bear the brunt of caregiving responsibilities within families. Improving paid care work through improving the pay of direct public sector employment such as in the K-12 education sector or more generous subsidies to quasi-public industries such as the healthcare sector would likewise have a disproportionate positive impact on women workers in caring labor. Helping families balance household care with market work and improving opportunities for paid care workers both serve to decrease women’s search frictions vis-à-vis misogynist employers.

In terms of reducing wealth inequality, a number of scholars have called for such actions. For example, Hamilton and Darity (2010) call for the implementation of ‘baby bonds’ to reduce
household wealth differences. Under their plan, every American would receive a bond at birth from the government funded through progressive taxation. The value of the bond would vary inversely with family wealth reaching $50,000 to $60,00 for children from families in the lowest wealth quartile. Such a strategy would reduce the wage gap through equalizing the relative degree to which white, black, and Latino workers respond to market signals given potential monetary shocks in the job search and transition process.

Understanding the underlying mechanisms that contribute to the persistence of racial, gender and intersectional wage gaps, and the systemic discrimination that supports them, is critical to developing a policy agenda that fosters equity. Thus, the model presented in this paper outlines a number of options to address persistent wage inequalities between like workers of different race, ethnicities, and gender.
Bibliography


Derenoncourt, E., & Montialoux, C. (2019). Minimum Wages and Racial Inequality. Working paper at the Washington Center for Equitable Growth. [https://equitablegrowth.org/working-papers/minimum-wages-and-racial-inequality/?mkt tok=eyJpIjoiWlRFM05HUTVOemt6TXpRNSIsInQiOiJPeUdxRE9RSEJNcTIWeUo3YWtwSWZoaVEzRW4vb0M3K1dKS3FBcU9XSHVVMWQ0b0Z3OVVP0VRTHZVY2NSbm1vZlwvR1U5MDVDUTNKN3Vpa1lMZm9SeCt2Qnl0Q29JUTJ2TEEExSzhOalYyTFwvY0RWZ3NqcHiDdFVwYWoYb3N5WDl6In0%3D](https://equitablegrowth.org/working-papers/minimum-wages-and-racial-inequality/?mkt_tok=eyJpIjoiWlRFM05HUTVOemt6TXpRNSIsInQiOiJPeUdxRE9RSEJNcTIWeUo3YWtwSWZoaVEzRW4vb0M3K1dKS3FBcU9XSHVVMWQ0b0Z3OVVP0VRTHZVY2NSbm1vZlwvR1U5MDVDUTNKN3Vpa1lMZm9SeCt2Qnl0Q29JUTJ2TEEExSzhOalYyTFwvY0RWZ3NqcHiDdFVwYWoYb3N5WDl6In0%3D).


Robinson, J. (1933). *Economics of imperfect competition*.


