11-2023

The Use of Regularization to Detect Racial Inequities in Pay Equity Studies: An Empirical Study and Reflections on Regulation Methods

Christopher M. Peña

The Use of Regularization to Detect Racial Inequities in Pay Equity Studies: An Empirical Study and Reflections on Regulation Methods

Abstract

Since the late 1970s, multiple linear regression has been the preferred method for identifying discrimination in pay. An empirical study on this topic was conducted using quantitative critical methods. A literature review first examined conflicting views on using multiple linear regression in pay equity studies. The review found that multiple linear regression is used so prevalently in pay equity studies because the courts and practitioners have widely accepted it and because of its simplicity and ability to parse multiple sources of variance simultaneously. Commentaries in the literature cautioned about errors in model specification, the use of tainted variables, and the lack of causal explanations. An empirical study comparing multiple linear regression, ridge regression, and LASSO regression models on a university employment data set was conducted next, focusing on racial inequity and methods informed by the literature review. The study results showed that while multiple linear regression yielded the highest coefficient of determination and the lowest mean squared error, LASSO regression yielded the highest predictive accuracy as measured by the standard error of the estimate. The study discovered the presence of racial inequity for Hispanic and Black employees at the university, including racial inequity for women in these groups.

Document Type

Dissertation

Degree Name

Ph.D.

First Advisor

Duan Zhang

Second Advisor

Michele Hanna

Third Advisor

P. Bruce Uhrmacher

Keywords

LASSO regression, Pay equity, QuantCRiT, Regression, Ridge regression

Subject Categories

Applied Statistics | Other Statistics and Probability | Physical Sciences and Mathematics | Quantitative, Qualitative, Comparative, and Historical Methodologies | Social and Behavioral Sciences | Statistical Models | Statistics and Probability

The Use of Regularization to Detect Racial Inequities in Pay Equity Studies: An Empirical Study and Reflections on Regulation Methods

---

A Dissertation

Presented to

the Faculty of the Morgridge College of Education

University of Denver

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Christopher M. Peña

November 2023

Advisor: Dr. Duan Zhang

Author: Christopher M. Peña
Title: The Use of Regularization to Detect Racial Inequities in Pay Equity Studies: An Empirical Study and Reflections on Regulation Methods
Advisor: Dr. Duan Zhang
Degree Date: November 2023

ABSTRACT

Since the late 1970s, multiple linear regression has been the preferred method for identifying discrimination in pay. An empirical study on this topic was conducted using quantitative critical methods. A literature review first examined conflicting views on using multiple linear regression in pay equity studies. The review found that multiple linear regression is used so prevalently in pay equity studies because the courts and practitioners have widely accepted it and because of its simplicity and ability to parse multiple sources of variance simultaneously. Commentaries in the literature cautioned about errors in model specification, the use of tainted variables, and the lack of causal explanations. An empirical study comparing multiple linear regression, ridge regression, and LASSO regression models on a university employment data set was conducted next, focusing on racial inequity and methods informed by the literature review. The study results showed that while multiple linear regression yielded the highest coefficient of determination and the lowest mean squared error, LASSO regression yielded the highest predictive accuracy as measured by the standard error of the estimate. The study discovered the presence of racial inequity for Hispanic and Black employees at the university, including racial inequity for women in these groups.
Keywords: pay equity, QuantCRiT, regression, ridge regression, LASSO regression

ACKNOWLEDGEMENTS

I would like to thank the faculty of the Research Methods and Statistics program for preparing me for this work, particularly Dr. Kathy Green, who helped me see and believe in my potential as a scholar.

Thank you to the members of my dissertation committee: Dr. Duan Zhang, Dr. Lolita Tabron, Dr. Bruce Uhrmacher, and Dr. Michele Hanna. Your patience, guidance, and feedback helped me grow significantly as a researcher through the dissertation process.

Thank you to my colleagues in the Department of Business Information and Analytics, especially Dr. Stephen Haag, Dr. Tamara Hannaway, and Dr. Kellie Keeling. Your mentorship and support helped me manage my new role as a professor while working toward completing my dissertation.

Finally, I would like to offer my heartfelt thanks to my friends Scott Hennelly and Dr. Helen Chao for their support and encouragement throughout my doctoral journey. Scott, you have been a true and loyal friend, and having you in my life gave me the strength and endurance I needed to complete this work. Helen, you have been a constant cheerleader and role model for me. Thank you both for your friendship and faith in me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

The use of multiple linear regression models to explore pay equity was first suggested in a student note in the *Harvard Law Review* in 1975 (Finkelstein, 1980). By 1979, complex statistical analyses dominated Title VII discrimination cases, with one federal judge commenting that they had become "contests between college professor statisticians who revel in discoursing about advanced statistical theory" (Otero v. Mesa County, 1979). Forty years later, notions of statistics and chance variability remain deeply embedded in the legal framework for demonstrating pay discrimination, with multiple linear regression still being the preferred analytic technique of plaintiffs, defendants, and the courts in discrimination suits.

This dissertation examines the use of multiple linear regression in pay equity studies through the lens of quantitative critical race theory (QuantCRiT), focusing on racial identities. In Chapter 1, the investigation of this topic begins with a discussion of the history of equal pay protection and the legal framework for identifying pay discrimination. In Chapter 2, a review of the literature is presented to synthesize different perspectives of the use of multiple linear regression in pay equity studies. The results of this review informed the methods investigated in the subsequent empirical study, described in Chapter 3, which examined the application of ridge regression and LASSO regression compared to multiple linear regression to address multicollinearity, overfitting, and the tension between inference and prediction. In Chapter 4, the results of this study are synthesized, followed by a discussion of findings and implications in Chapter 5.

Background

In FY 2021, the Department of Labor reported that People of Color nationally earn an average of $0.76 per dollar relative to White workers (DOL, 2022). That same year, the Equal Employment Opportunity Commission (EEOC) received 20,908 charges of racial employment discrimination, including discrimination in pay, representing 34.1% of all complaints received by the agency. The EEOC found cause to investigate 30.3% of the racial discrimination cases it received, resulting in $99.3 million in awards to claimants through conciliation and litigation. Despite legal protections afforded under the law, equal pay and equal employment opportunity remain elusive and costly ideals in the United States.

Early Protections

The origins of racial employment discrimination trace to the mid-nineteenth century following the Civil War. The thirteenth Amendment to the Constitution abolished slavery in the United States, but the fight for civil rights continued over the following decades through a series of legislative and executive actions. The Civil Rights Act of 1866 guaranteed all citizens the same rights as those held by White people, and while its primary purpose was to protect individuals from discrimination by the federal and state governments, the courts subsequently interpreted the Act to apply to discrimination by private employers as well. A decade later, the Civil Rights Act of 1875 guaranteed all men equality under the law, regardless of race, color, nationality, religion, or political affiliation. Women continued to be excluded from such protections for nearly a century afterward.

Reconstruction

Reconstruction was a time of upheaval in the United States as the country tried to rebuild after the Civil War and redefine its relationship with formerly enslaved Black Americans. While Reconstruction saw some progress in legal protections for Black Americans, it also created new forms of systemic racism that would persist for decades (Shlomowitz, 1979).

Sharecropping functioned as a contractual agreement between landowners and laborers in which Black Americans – mostly formerly enslaved – were provided with access to land, tools, and essential supplies by landowners in exchange for a share of the crops they produced (Shlomowitz, 1979). The sharecropping system directly resulted from the collapse of the antebellum plantation economy and the emancipation of enslaved Black Americans. The system was designed to address the labor shortage that arose when formerly enslaved Black Americans sought greater economic independence and mobility. Although seemingly advantageous, this arrangement was inherently skewed in favor of the landowners, who often held significantly more power and influence. The contracts offered to Black sharecroppers were frequently characterized by ambiguous terms and exploitative clauses, such as exorbitant interest rates for loans, exactions for using tools, and predetermined crop prices (Fusfield & Bates, 2005). These contractual discrepancies allowed landowners to maintain control over the means of production and secure a disproportionate share of the profits, leaving sharecroppers trapped in cycles of debt and dependence. However, establishing this system effectively tied Black farmers to their White landowners in a system of debt bondage (Fusfield & Bates, 2005).

Most landowners during Reconstruction were White, while sharecroppers were predominantly Black (Fusfield & Bates, 2005). Accordingly, the social dynamics of the system perpetuated racial discrimination and violence, further marginalizing Black Americans and hindering their pursuit of economic growth. Discriminatory practices, such as the imposition of harsh labor contracts, violent reprisals for perceived infractions, and denial of access to legal protection, created a climate of fear and intimidation, stifling any attempts by Black Americans to challenge the existing system (Shlomowitz, 1979). Additionally, the sharecropping system ensured that Black labor remained primarily confined to the agricultural sector, limiting opportunities for skill development, education, and social mobility.

Post-Reconstruction

Post-Reconstruction saw a continuation of many of the same racial inequities established during Reconstruction. While the end of slavery had eliminated the most egregious form of exploitation, Black workers still faced discrimination and mistreatment in many forms. Many White employers, for example, refused to pay Black workers the same wages as White workers for the same work, and many Black workers were relegated to low-paying jobs with few opportunities for advancement (Guerin-Gonzales, 1994).

Hispanic workers faced various forms of employment discrimination during this time, similarly fueled by systemic racism and economic interests. Hispanic workers, particularly Mexican laborers, were often subject to unfair labor practices and exploited for cheap and sometimes forced labor. Many worked in agriculture, mining, and railroad construction, enduring long hours, low wages, and hazardous working conditions (Guerin-Gonzales, 1994). In the post-Reconstruction era, Hispanic workers also faced segregation in the labor market as the United States moved toward industrialization. Like Black workers, Hispanic workers were often restricted to low-paying manual jobs and excluded from better-paying opportunities available to White workers.

While Reconstruction brought some progress in civil rights for Black workers, there were limited legal protections against employment discrimination for Hispanic workers. State and local laws failed to address or prohibit discriminatory practices at the time based on race and ethnicity. Federal and state governments also discriminated against Hispanic workers in public sector jobs, openly discriminating against them in hiring, promotion, and wages. These actions reflected an anti-Mexican sentiment pervasive in the United States at the time, particularly in California and Texas, contributing to hostility against Hispanic workers and exclusion from mainstream opportunities.

Early Twentieth Century

The early twentieth century saw the rise of Jim Crow laws and the institutionalization of racial segregation in many parts of the country. These laws profoundly impacted Black Americans, as they were denied access to many public services and continued to be restricted to low-paying jobs. In addition, many White employers actively sought to exclude Black workers from higher-paying jobs, reinforcing the racial inequities established during Reconstruction and the decades after (Beeby & Nieman, 2002).

The convict leasing system that emerged in the early twentieth century was another key factor in creating racial inequities in pay. Under this system, Black men were often arrested on fabricated charges and forced to work for White landowners under conditions that were arguably little better than slavery (Muller, 2018). Because these workers were often unable to pay their fines, they were effectively trapped in a system of debt bondage that kept them working for years without any prospect of improving their situation.

The Great Migration of Black Americans from the South to the North in the early twentieth century resulted in further racial inequities in pay. While many Black workers found better-paying jobs in the North, they were often relegated to lower-paying, menial jobs with no opportunity for advancement. Moreover, many White workers resented the influx of Black workers, which increased racial tensions and discrimination in many parts of the country (Tolnay, 2003).

The New Deal policies of the 1930s were intended to help lift the country out of the Great Depression, but they also had unintended consequences for Black and Hispanic Americans. Many of the programs established under the New Deal were administered at the state and local level, which meant that People of Color in many parts of the country were effectively excluded from them (Kennedy, 2009). Many of the jobs created under the New Deal were also low-paying and temporary, which did little to address the root causes of poverty and inequality for People of Color (Skalroff, 2009).

The post-World War II era saw the emergence of a new form of system racism in the United States, as White Americans increasingly moved to the suburbs and left behind the urban centers that had been home to many Black Americans. This created a situation in which Black Americans were effectively trapped in urban areas with few job opportunities and little chance for upward mobility (Tolnay, 2003). Likewise, many of the jobs that were available to Black Americans and other People of Color were low-paying and precarious, which perpetuated the cycle of poverty and inequality (Anderson, 1993).

Equal Pay Act of 1963

Legislation passed in the mid-twentieth century served as a first step in addressing the historical pay and employment discrimination addressed by People of Color. The Equal Pay Act of 1963 (EPA) was the first piece of legislation to deal with equal pay and was passed as an amendment to the Fair Labor Standards Act (FLSA) signed by President Kennedy. The EPA states that “no employer shall discriminate … between employees based on sex by paying wages to employees in such establishment at a rate less than the rate at which he pays wages to employees of the opposite sex … for equal work." The EPA further advises that wage inequalities between men and women should be reviewed to determine whether they indicate a pattern of sex discrimination. The EPA identifies differences that cannot be justified based on a bona fide seniority system, merit, quantity or quality of work, or a factor other than sex as an equal pay violation. While the EPA was initially intended to protect women from being paid less than men, the courts have since extended its protections to men.

Title VII of the Civil Rights Act of 1964 Pay Protection

In addition to the EPA, equal pay is also safeguarded under the Civil Rights Act of 1964, which protects individuals from employment discrimination. Title VII of the Act was more far-reaching than previous statutory and case law, incorporating unions and private companies, but it still did not apply to public sector organizations. Notably, however, Title VII extended protections against pay discrimination to individuals based on race, color, religion, sex, or national origin, turning employment discrimination into a legal offense under federal law.

Under Title VII, an employee may show pay discrimination through the legal principles of disparate treatment and disparate impact. To show *disparate treatment*, the complainant must demonstrate they received less money for "performing substantially similar work due to an intent to discriminate on the part of the employer" (Powell Jr., 1973). Since direct evidence of discriminatory intent rarely exists, pay equity cases typically rely on *disparate impact* instead. Disparate impact relies on circumstantial evidence to demonstrate discriminatory impact, often class wide. Under this legal principle, a facially neutral employment practice that disproportionately affects members of a protected class may be considered unlawful, even without proof of intent. Complainants often employ statistical analysis in disparate impact charges to demonstrate the discriminatory consequence of an employer's policies, practices, and actions (Roth et al., 2006).

Title VII incorporates four affirmative defenses established under the EPA: seniority, merit, quantity or quality of work, or any factor other than sex. However, the United States Supreme Court still needs to provide specific guidance on what is required to prove wage discrimination, and it has limited the scope of disparate impact as a judicial principle on several occasions, expressing concern for it being extended to constitutional review.

Protections Under Title VII and the EPA

Although there is significant overlap in the scope and protections provided by the EPA and Title VII, the two laws also differ in their treatment of the specific elements of discrimination cases.

Suing. In 1964, Congress created the Equal Employment Opportunity Commission (EEOC), which is charged with enforcing Title VII of the Civil Rights Act of 1964. This independent federal agency has been charged with enforcement of a broad range of employment-related legislation as well, including the EPA, the Age

Discrimination in Employment Act of 1967 (ADEA), section 501 of the Rehabilitation Act of 1973, Title I of the Americans with Disabilities Act of 1990 (ADA), and the Civil Rights Act of 1991. For all laws enforced by the EEOC – except the EPA – complainants pursuing a charge of employment discrimination must first bring their case to the agency before bringing it to the Court. Once a charge is filed with the EEOC, the agency can choose to investigate or settle. The EEOC prefers pursuing a conciliation agreement with the employer to remedy their discrimination findings, but if one cannot be reached, it can take the case to court. However, a charge brought under the EPA may be taken directly to court.

Scope. The most significant difference between the EPA and Title VII is the scope of who is protected under them. The EPA limits protections only to cases of pay discrimination based on sex. Historically, the courts interpreted the EPA to apply to women only, but it has since extended its protections to men. The EPA is also limited to wage discrimination and is not concerned with other employment practices, even those whose consequences are related to compensation. Title VII provides broader protections, extending remedies to individuals based on sex, race, color, national origin, or religion in compensation and employment practices.

Prima Facie. To first establish pay discrimination, the prima facie burdens under the EPA require the complainant to show that they were receiving a lower wage than employees of the opposite sex in the same establishment for doing work of equal skill, effort, and responsibility under similar working conditions. Under Title VII, making a prima facie case can be made under disparate treatment or disparate impact. If arguing under disparate treatment, the complainant must be a member of a protected class and provide evidence of an intent to discriminate against members of that class. If arguing under disparate impact, the complainant must be a member of a protected class and show that the treatment received was worse than that by similarly situated employees who are not in the protected class.

Intent. The EPA applies a strict liability standard: any difference in pay is potentially actionable, and intent to discriminate does not need to be proved. The only exceptions provided under the law are a seniority system, a merit system, a pay system that is based on quantity or quality of production, or any factor other than sex, such as experience or education. By contrast, Title VII requires the complainant to demonstrate pay discrimination through disparate treatment, which carries a presumption of intent, or disparate impact, which focuses on circumstantial evidence. The courts established explicit guidelines for showing discrimination under these two judicial principles through *Albermarle Paper Company v. Moody*, *Griggs v. Duke Power*, *United States v. Hazelwood*, and *United States v. Teamsters*.

Equal Work Standard. While the EPA requires claimants to show they are being paid less than employees of the opposite sex for equal work, Title VII does not have such a standard. The EPA requires that the work be of substantially similar skill, effort, and responsibility, and that it be performed under similar conditions, but it does not specify how these factors are to be determined or what evidence supports an assertion about them.

Title VII is broader and simply prohibits discrimination based on sex, race, color, national origin, or religion. The Bennett Amendment incorporated the EPA's affirmative defenses into Title VII, but it remains unclear to what extent the equal work standard has been incorporated as well. Since the passage of Title VII, the lower courts have offered different interpretations of whether the equal work standard applies to cases brought under the Civil Rights Act of 1964. In certain cases, a Title VII claim was not precluded simply because the two jobs were not considered equal under the EPA. However, in 1981, the Supreme Court ruled that the Bennett Amendment did not incorporate the equal work standard along with the four affirmative defenses and that plaintiffs suing for sex discrimination under Title VII did not need to meet the equal work standard of the EPA. The case in question was remanded to the lower court for further litigation.

Use of Multiple Linear Regression in Pay Equity Studies

The current accepted analytic technique for analyzing pay equity under statutory and case law is multiple linear regression (Altonji et al., 2013; Connolly, 1991; Dempster, 1988; Follett & Welch, 1983; Peterson, 1999; Stillson, 2002). To infer disparate impact or treatment, plaintiffs first develop a multiple linear regression model that identifies membership in a protected class as a statistically significant factor in predicting pay as an outcome. These models typically include additional variables that reflect an organization's compensation philosophy and affirmative defenses under Title VII.

Alternatively, plaintiffs may convert predicted pay from a multiple linear regression model into z-scores to identify individual outliers among the population of employees studied to determine whether the proportion of outliers who are members of protected classes differs significantly from the expected frequency. Conversely, defendants in a disparate impact or treatment case may use multiple linear regression to demonstrate that membership in a protected class is not a statistically significant factor in predicting pay or that the proportion of those individuals in a protected class who are identified as outliers does not differ significantly from the expected frequency.

Statement of the Problem

Federal law prohibits discrimination in pay based on sex, race, religion, national origin, or disability status. All fifty states align with federal law on prohibiting pay discrimination by sex, but only ten states have statutes that specifically address pay equity based on race or ethnicity (UC Irvine, 2021). Claiming discrimination therefore often requires navigating a difficult process at the federal level for many People of Color. If successful, these cases can result in heavy financial penalties against the offending organization. In the case of discrimination by public officials, the cost of these fines and awards is passed on to the taxpayers, drawing funds from other projects where they may be needed.

Pay discrimination for People of Color likewise has a cascading effect within the workforce when left unchecked. Organizational studies provide ample evidence that companies that employ a diverse workforce and practice pay equity are better positioned to recruit and retain high-performing individuals and respond to an ever-changing marketplace nimbly (Beyer, 2019; Carter, 2015). Pay equity and transparency of compensation philosophies have likewise been correlated with higher performance and employee morale, further supporting recruitment and retention (Beyer, 2019). Ensuring pay equity not only insulates an organization from heavy financial losses from litigation, but also supports its growth and advancement (McDermott, 2000).

Numerous studies have been conducted on pay equity for women at both the macro- and micro-data level (Altonji et al., 2013; Barbezat, 2002; Barrett et al., 1986; Barrett & Doverspike, 1989; Best et al., 2011; BIllard, 2017; Connell & Mantoan, 2017). However, few studies exist in the literature on pay equity for People of Color, and even fewer for intersectional identities, such as women of color. Understanding pay equity from the perspective of minoritized communities is necessary to design effective pay equity studies and to determine progress toward pay equity on a larger scale in the country. Currently, there is no effective mechanism to gauge such progress or guidelines for how best to measure it.

Since the 1970s, multiple linear regression has been adopted by the courts as the most comprehensive analytic technique available to determine the degree of influence membership in a protected class has on pay outcomes. While the literature contains many criticisms of this approach, no work has yet synthesized these findings into a single study. Further, despite the criticisms of multiple linear regression to analyze pay equity in the literature, very few studies have recommended or explored alternative analytic techniques on this topic.

Purpose and Significance of the Study

The primary purpose of the research was to provide a deeper understanding of the use of multiple linear regression and regularization methods to detect racial pay inequities within organizations. The current research sought to achieve this goal and contribute to the literature through an empirical study.

A literature review was first conducted to synthesize the support and criticisms of multiple linear regression to detect pay discrimination. This review contributes to the literature with a substantive discussion of the tension between inference and prediction that is inherent in multiple linear regression models, particularly those in which coefficients are not constrained.

In the empirical study, anonymized empirical data obtained from a four-year research university in the Western United States were analyzed with multiple linear regression, ridge regression, and LASSO regression to study pay equity. The goal was to determine whether regularization yielded better model fit and higher predictive accuracy than multiple linear regression estimated by ordinary least squares (OLS) when applied to the context of a pay equity study. To date, no such advanced analytic technique has been recommended or explored in the pay equity literature.

Research Questions

Research Question 1: Which of the three models yields the highest coefficient of determination (R^2), the lowest standard error of the estimate (S), and the lowest mean squared error (MSE)

Research Question 2: To what extent does racial disparity in compensation exist within the university's workforce, and what are the measurable factors contributing to this inequity?

Theoretical Framework

This research examined the use of multiple linear regression in pay equity studies through two distinct lenses: human capital theory and quantitative critical race theory. Together, these approaches provided a framework for identifying how multiple linear regression may obscure or reproduce racial inequity in pay equity studies.

Human Capital Theory

Human capital theory is a widely used economic framework for understanding how individuals invest in their education, training, and work experience, which in turn impacts their productivity and earning potential. This theory suggests that differences in pay between individuals can be attributed to differences in their level of human capital: individuals who have invested more in their education and training are likely to have a higher earning potential than those who have not invested as much in these areas. Human capital theory is accordingly often used to explain pay differences between groups of individuals with different levels of education and training.

Concerning pay equity studies, human capital theory is often used to examine pay differences between men and women, as well as between different racial and ethnic groups. Proponents of human capital theory argue that these pay differences can be explained by differences in human capital investments, such as education and work experience. For example, certain employees may have lower earnings on average than others because they are less likely to pursue high-paying occupations and more likely to take time off from work to care for children. Similarly, racial and ethnic pay gaps may be explained by differences in educational attainment, work experience, and other human capital factors.

Critics of human capital theory argue that it fails to account for other factors that may contribute to pay differences between groups, such as discrimination and bias in the labor market (Fogel, 1986). They suggest that even when controlling for differences in human capital, significant pay gaps persist between groups, which suggests that other factors must be at play. Many pay equity studies therefore use a combination of human capital theory and other frameworks to better understand and address pay disparities.

In alignment with EEOC guidance on lawful discrimination and affirmative defenses, this study employed human capital theory to predict pay using education, years of experience, and whether the employee had a graduate degree, with a particular focus on People of Color.

Quantitative Critical Race Theory

Quantitative critical race theory (QuantCRiT) has emerged in recent years in response to post-positivist claims about the objectivity of quantitative research methods (Tabron & Thomas, 2023). QuantCRiT posits that, like qualitative and mixed methods approaches, quantitative methods are likewise influenced by subjectivity and failing to recognize this truth weakens the credibility and applicability of quantitative research findings. This issue is even more problematic when it occurs in the context of diversity, equity, and inclusion, where quantitative analysis has historically been used to silence and marginalize already minoritized communities. Rather than being an offshoot of critical race theory (CRT), QuantCRiT offers a framework that applies "CRT understandings and insights whenever quantitative data is used in research and/or encountered in policy and practice" (Gillborn et al. 2018).

Racism Is Endemic. QuantCRiT posits that racism is a complex and deeply rooted aspect of our society that is not "readily amenable to quantification" or statistical inquiry (Gillborn et al. 2017). It holds that racism is a "complex, fluid and changing characteristic of society" and that the absence of a critical race-conscious perspective can mask or perpetuate existing racial inequities. This tenet suggests that regardless of whether a particular instance of discrimination occurred against an individual or class of employees, racism nonetheless exists within complex systems such as employment and human resources management.

Numbers Are Not Neutral. QuantCRiT holds that numbers are not neutral and should be interrogated carefully to determine how they reflect the "interests, assumptions, and perceptions of White elites" (Gillborn et al., 2017). This approach encourages researchers to consider how racist logic – overt and covert – has informed the collection and analysis of data and interpretation of findings. Gillborn et al. (2017) noted that statisticians may sometimes assert that "their view is the only true or legitimate understanding of the world." They warn of the dangers and implications of this fallacy, even when researchers approach a study with good intentions.

Categories Are Not Natural or Given. Gillborn et al. (2017) argued that categories and groups are not “natural or given.” This tenet directly addresses how demographic variables are encoded in quantitative studies, which they note can have “fundamental consequences for the re/presentation of race inequity” (Gillborn et al. 2017). They explained that historically, analysts have asserted that race/ethnicity has not been a statistically significant factor in a model after having compared White individuals against everyone else as a “non-White composite” (Gillborn et al., 2017). This practice not only relies on too few demographic categories, it also perpetuates the practice of centering whiteness in quantitative studies. They argued further that the concept of race “always carries the inherent threat of racist assumptions and actions” and readers should be prompted to view race “critically as a social construct that historically separates and oppresses particular groups” (Gillborn et al. 2017).

Data Do Not Speak for Themselves. QuantCRiT holds that numbers are a social construct and are likely to “embody the dominant (racist) assumptions that shape contemporary society” (Gillborn et al., 2017). Gillborn et al. (2017) noted that regression analyses have been particularly problematic because the analysts running them fail to recognize that racism “operates through and between many factors simultaneously.” Studies that purport to segment sources of variance therefore not only deny the reality of intersectional identity and oppression in the real world, but they also fallaciously assume these sources can be isolated and studied independently. This tenet suggests that quantitative findings must be interpreted in context and provided faithfully to readers to fully understand the analysis process and implications of the results.

Statistical Research Is Not Value-Free. QuantCRiT states that statistical research is not "value-free or politically neutral" (Gillborn et al., 2017). This approach seeks to support social justice goals and advance equity by adopting a position of "principled ambivalence, neither rejecting numbers out of hand nor falling into the trap of imagining that numeric data have any kind of enhanced status" (Gillborn et al, 2017). QuantCRiT holds that like CRT, practitioners should work to resist and eliminate racism by challenging conventional wisdom about statistics and highlighting uncomfortable truths that come to light when critically conscious approaches are employed. These goals stand in contrast to the purported objectivity and neutrality of quantitative studies, which ostensibly seek to quantify, study, and understand the natural world and the human condition. Instead, the act of research is inherently political and value-biased because researchers must make numerous choices in the analysis and reporting process that are subjective and often of significant consequence to individuals and communities.

Positionality Statement

Speaking on critical quantitative research and quantitative criticalism, Tabron & Thomas (2023) noted that "it is vital to share how we locate ourselves within [our] work." In the spirit of self-reflexivity, I acknowledge that my view and experience of the world is informed by my racial, intersectional, and professional identities. I am a cis-gender, Hispanic man who is a member of several minoritized communities, and I have had the privilege of access to education. I have worked in higher education for twenty years, with twelve of them in institutional research; my time as a practitioner has led me to focus on applied research with a quantitative focus. Throughout my career, I have been an avid participant in justice, diversity, equity, and inclusion spaces, and I gravitate towards QuantCRiT and CRT as frames for understanding systems, institutions, and the world around me.

Definition of Terms

Coefficient of Determination (R^2) – A measure that assesses the strength of the linear relationship between a dependent variable and one or more predictors.

Lambda (λ) – A tuning parameter that controls the strength of the penalty term in ridge regression and LASSO regression.

LASSO Regression – A method of regularization in regression in which some of the coefficients in the model are constrained to zero.

Mean Squared Error (MSE) – The average squared difference between predicted and observed values.

Multicollinearity – The statistical association between two or more independent variables in a regression model.

Regularization – A method of regression in which bias is introduced into the model to reduce variance and increase predictive accuracy.

Ridge Regression – A method of regression in which coefficients in the model are constrained close to zero.

Standard Error of the Estimate (*S*) – A measure of the average deviation of the errors in a model.

CHAPTER 2: LITERATURE REVIEW

In this chapter, a literature review is presented to identify methodological and technical support for and critiques of the use of multiple linear regression and other regularization methods in pay equity studies. The purpose of the review was to 1) synthesize different perspectives from commentaries in the relevant academic, professional, and legal literature and 2) inform the methods selected for the empirical study in this dissertation.

Detecting Pay Inequity and Discrimination

Early Criticism

Weisberg & Tomberlin (1983) raised some of the earliest concerns about technical and conceptual issues related to the use of statistical analyses in discrimination cases. They began by explaining that multiple linear regression allows for the testing of differences in salary distributions between groups, "conditional on some measures of qualifications" (Weisberg & Tomberlin, 1983). The authors listed the choice of method as the primary concern, and the verification of model assumptions – including linearity, normality, and homogeneity of variance – as related areas for consideration. The conceptual issues they raised include applying statistical analyses to "prove" discrimination in lawsuits. While Weisberg & Tomberlin did not discuss the requirements for establishing causality in their critique, they noted that it is difficult to state with confidence that a certain estimate effect in a regression model is attributable to employment discrimination. In noting an early proposal to measure discrimination by comparing the qualifications for different groups receiving a specified outcome, they also addressed what would be described as a counterfactual today, often cited as supplemental criteria for establishing a causal relationship (Guo & Fraser, 2014). The authors noted that it is possible for various alternatives to produce different and even conflicting conclusions. Although they did not name selection bias as a concern, they explained that information about certain minoritized groups may be lacking due to self-selection and limited public data for comparison.

Manifestation of Discrimination

In analyzing the early response of the courts to the use of multiple linear regression, Fogel (1986) explained that the use of statistical analyses cannot be appropriately applied in discrimination cases without first establishing what discrimination is and how it manifests in organizations. He discussed the Equal Pay Act, in which men and women are required to be paid equally for substantially equal work, and noted that cases at the time offered little guidance on how the courts could authorize such claims. Fogel described the typical process for developing a multiple linear regression model and emphasized the importance of model specification because “if any important influence on salary … is omitted from the regression, a confident assessment of the influence [of factors] cannot be made” (Fogel, 1986). Fogel alluded to interaction effects, explaining that the influence of one variable on salary as an outcome may be better explained in the presence of one or more other variables in the regression model.

He focused on model specification through his critique and referenced the human capital model of employment, in which individuals believe their pay should be determined by their qualifications, not the position they hold. This position aligns with the affirmative defenses for organizations to rebut discrimination claims under Title VII and the Equal Pay Act.

Like Weisberg & Tomberlin, Barnett discussed the then-growing emphasis placed on statistical proof of discrimination in salary studies as leading to "greater sophistication in the techniques used to prepare evidence in court" (Barnett, 1986). However, he explained that while regression analysis can in principle determine whether differences across groups reflect unlawful discrimination in pay, it can also be abused. He noted that the courts have appropriately rejected the results of multiple linear regression studies that "unaccountably ignored more influences on the process under study or employed irrelevant or unreliable data" (Barnett, 1986). His primary criticism of the use of multiple linear regression was that the courts had at the time offered little guidance on the functional form of modeling required to establish discrimination in pay equity cases. Barnett further explained that the accuracy of regression results is compromised when analysts fail to explain *how* the variables exert their influence on salary as an outcome, indirectly referencing the need for establishing a causal relationship.

Model Specification

Barrett et al. (1986) discussed the advocacy of the use of multiple linear regression to investigate pay discrimination across the statistical, social science, legal, and personnel management literature. They explained that multiple linear regression is frequently used by the courts because it allows them to predict compensation by evaluating independent variables, such as education and experience. The authors noted that while the use of multiple linear regression is widely accepted in discrimination cases, criticisms were paradoxically increasing. They critiqued an early study by Seberhegen (1979), which inappropriately attributed residuals in a regression model to sex discrimination. Their criticism focused on model specification, highlighting the problematic nature of identifying relevant predictors in the regression equation. This observation mirrors the one made by Barnett that same year. They explained that certain variables used in regression studies are ambiguous, poorly defined, or assume equality where none exists. The authors discussed education as a primary example of poor model specification, in which researchers assume the equality of all degrees of the same level, such as bachelor's degrees, when their relationship to a position may vary by discipline. They also cited the number of people supervised as a similarly problematic variable, explaining that supervising a greater number of people may not necessarily be more difficult depending on the nature of the work and the positions being managed. In their final analysis, Barrett et al. questioned whether individuals identified as being underpaid in the model see themselves as victims of discrimination. They noted that the lack of congruence between the statistical models proposed to investigate discrimination and the perception of discrimination indicates yet another problematic aspect of multiple linear regression.

Ananda & Gilmartin (1991) continued the conversation begun by Barnett and Barrett et al. on model specification with a comprehensive review of the issues relating to the inclusion of potentially tainted variables in regression analysis in the context of employment discrimination cases. The authors began by describing the utility of multiple linear regression in such cases, explaining that the technique allows for the simultaneous inclusion of potentially explanatory variables in predicting salary as an outcome. They further noted that multiple linear regression is the most frequently used technique to "analyze the combined effects of various independent variables" on salary (Ananda & Gilmartin, 1991). They also argued that the courts have not sufficiently distinguished between measures of employees' current job levels, initial job levels, and job levels with previous employers, and they have evaluated regression models without reference to specific allegations of pay discrimination. The authors stated that even when a variable is tainted, its inclusion in regression analysis has still been used by the courts to evaluate the disparate treatment and disparate impact of employees with respect to pay. They explained that explanatory variables commonly used in regression, such as education, previous work experience, and seniority, may be unavailable, inaccurate, or unreliable. Variables under control of the employer, they noted, are particularly suspect. Ananda & Gilmartin (1991) argued that a variable should be considered tainted when "there is a belief that the employer can shape its definition or measurement to the disadvantage of [a] protected group." However, the courts do not follow such a definition, they explained, and guidance offered by case law is often in conflict. They offered job level and performance appraisal ratings as evidence of such inconsistencies, in which the variables are considered affirmative defenses under Title VII and the Equal Pay Act but are subject to the potentially “discriminatory influence of the employer.”

Sarkar & Haverstick (2010) explained that a proactive pay equity audit is “one of the best ways for employers to reduce their risk” in discrimination cases. Following this logic, they noted that multiple linear regression allows organizations to model neutral job-related and productivity-related characteristics that may legitimately affect pay. These factors may include direct and indirect measures, which multiple linear regression is well-suited to model. However, they also explained that there is no one-size-fits-all approach to selecting variables for a pay equity regression, and that predictors should reflect the structure, policies, and processes in effect at an organization. The authors explained that the modeling process can only confirm the choice of variable by checking that each variable – when added to the model – has a statistical influence on the rate of pay. This observation again reiterates the importance of model specification begun by Barnett. Sarkar & Haverstick emphasized the importance of proper technique in developing a regression model, stating that methodological flaws can create the impression of pay inequities when none exist. They explained that such problematic use of multiple linear regression can ironically lead to new claims of discrimination, in addition to unnecessary and costly organizational changes. The authors highlighted the use of the coefficient of determination in determining the strength of statistical significance, and they discussed how improper aggregation inadvertently can lead to bias. The authors argued that multiple separate regression models may be warranted when it would be improper to combine all employees into a study, such as when different departments employ different human resources policies and practices, or when collecting bargaining agreements affect the pay of some employees but not others. They also argued that a single regression model may still be appropriate, provided that the model accounts and controls for the "differences between the types of employees that play a legitimate role in determining the compensation rate." These variables may be considered neutral when they affect pay similarly among the different groups being analyzed. If the pay equity regression fails to account for one or more key variables, they explained, the outcome of the regression may be subject to omitted variable bias.

Alternative Approaches

In discussing his preference for matching and stratification models, Gastwirth (1988) criticized the use of multiple linear regression in isolation in analyzing pay equity. He stated that these models fail to capture the relative importance of different productivity characteristics as related to the job itself. He further noted that based on his review of cases at the time, the use of multiple linear regression to establish a constant or consistent shift toward discrimination is unrealistic, and that the information concerning discrimination is likely better to be found through discordant pairs in a matched student. He explained that in most organizations, highly qualified employees will be promoted or earn more pay, while lower qualified employees will not. Gastwirth stated that the middle range of qualification levels is where discrimination is most likely to appear, which is equally the most difficult to model appropriately using regression analysis. He cited the need to model interactions as Fogel did, as well as non-orthogonal relationships in regression studies of pay equity. He further discussed the problematic feature of high variability of estimated coefficients with coefficients of determination that appear strong.

Bura et al. (2010) discussed several alternatives to standard regression methods, such as including binary variables to identify individuals belonging to a protected class. They noted these methods have been accepted by the courts and therefore are applicable in pay discrimination cases. Such models inappropriately aggregate individuals in those classes, they argued, implicitly assuming a discriminatory policy has the same effect on every employee belonging to the group. The authors explored Peters-Belson regression using both parametric and non-parametric approaches on the *EEOC v. Shelby County* equal pay case. This type of modeling is a form of statistical matching in which a regression is fitted to one group of employees to predict pay and then applied to another group using the same linear model. The difference, if any, between the actual and predicted pay for members of the second group estimates their under-compensation. Peters-Belson regression may help estimate the counterfactual in employment discrimination cases, but it is not used widely outside the area of measuring minority disparities (Pearl, 2000; Heckman, 2005). The authors explained that regression and matching procedures lose statistical power and efficiency when the distribution of predictors varies considerably between the two groups under study. Bura et al. concluded that while other statistical tests, such as the Chow-Rao test, are too stringent in employment discrimination cases, further research was still needed, particularly to address tests for a lack of interaction effects.

Evolution of Discrimination Cases

Bielby & Coukos (2007) summarized several issues relevant for the courts to interpret multiple linear regression models in analyzing pay equity. They provided a substantive discussion of pay discrimination cases, which focus on the applicability of the current legal framework in the context of a shifting nature of bias and inequity. Bielby & Coukos explained that multiple linear regression allows the incorporation of subjective decision-making, albeit with caveats. The authors noted that aggregation is a point of contention in many cases and argued that because of the way case law has evolved, "courts place more weight on 'statistical significance' than the magnitude of disparities between groups" (Bielby & Coukos, 2007). Particularly, Bielby & Coukos noted the importance of assessing disparities both organization-wide and within subunits, which may be particularly relevant when a company employs a decentralized personnel management structure. In such cases, a unique employment system may exist within individual units, which regression models must reflect. Bielby & Coukos described "second-generation" discrimination suits, where organizations have made good faith efforts to address inequities in their pay and leadership structures, despite claims to the contrary. These cases have led organizations to claim that discrimination may be the result of "individual deviance", which has resulted in the courts interpreting "excessive subjectivity" with suspicion, treating potentially discriminatory practices as neutral until proven otherwise (Bielby & Coukos, 2007). Bielby & Coukous explained that the courts continue to reach substantively different conclusions about the role of subunit differences in discrimination cases, with some courts giving little credence to claims that differences in one part of the organization constitute institutional policies or practices that demonstrate disparate impact. They noted that one reason subunit differences have not been fully evaluated in regression models in discrimination cases is they may lack sufficient statistical power to substantiate claims by both plaintiffs and defendants.

Rouen (2017) argued that one potential reason that plaintiffs and defendants reach different conclusions in analyzing pay disparities is that they fail to account for how compensation is determined. The author noted that without controlling for various determinants of compensation, analysts fail to distinguish between pay equity and equality. Rouen explained that pay equity is the notion that differences are or are perceived to be unfair, while pay inequality reflects an actual difference in compensation between groups. Rouen's doctoral research examined how compensation based on measurable attributes related to factors of performance and the labor market creates pay disparity without necessarily resulting in a perception of inequity. By contrast, pay disparity created by factors unrelated to economics, such as favoritism, may result in a perception of inequity. The author noted that the increase in pay-for-performance compensation programs complicates this problem. These systems are designed to distribute compensation based on effort and performance, but they also necessarily increase pay disparity within an organization "as some workers outperform others" (Lemieux, 2008; Trevor et al., 2012). Rouen explained that these systems are consistent with equity theory models since the distribution of pay will be perceived as fair and equitable because it reflects relevant economic and performance-based favors. Rouen concluded that pay disparity resulting from individual performance is positively related to team performance, while the relation between pay disparity that is unexplained by individual performance results in significantly or marginally negative team performance. The author's research supplemented commentary in the literature that focuses on employment discrimination by noting the relationship between pay equity and the perception of fairness to team and organizational performance. These factors may warrant additional scrutiny by organizations seeking to conduct a pay equity study.

Multiple Linear Regression and Regularization

Multiple Linear Regression

Linear regression is a supervised learning technique for modeling the relationship between a dependent variable y and an independent variable x by fitting a linear equation to the observed data. If a stable relationship can be defined, the value of x can be used to predict the value of y. This technique assumes an approximately linear relationship between the two variables and can be expressed as:

$$y = \beta_0 + \beta_1 x \tag{1}$$

In the linear model, β_0 and β_1 are two unknown parameters and represent the intercept and slope respectively. Perfect prediction of y is possible when the model is a function, but researchers can often only define a statistical relationship that provides a good fit to the data (Vasu, 1979). Inexact parameter specification yields bias in the linear model, which must be considered when evaluating its fit to the data.

The goal of linear regression is to obtain estimates for β_0 and β_1 that produce a line that is as close as possible to the observed data. The most common approach is to minimize the least squares criterion (Fox, 1997). The *squares* in the ordinary least squares (OLS) approach are the squared values of the difference between the actual and predicted value for each observation. This residual is represented as *e* in the linear model when predicting the value for *y*:

$$\hat{y}_i = \beta_0 + \beta_1 x_i + e_i \tag{2}$$

Simple linear regression considers the linear relationship between a dependent variable and one independent variable. To predict the value of *y* using multiple predictors, it is possible to develop regression equations separately for each independent variable. However, this approach has the limitation of making it impossible to determine the degree of influence of one predictor on the dependent variable in the presence of one or more other predictors. The preferred approach is multiple linear regression, which extends the simple linear regression model to accommodate multiple predictors in the same equation:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_i x_i + e \tag{3}$$

In this model, each independent variable has a distinct slope, or coefficient, which represents the degree of change in the dependent variable *y* given a one-unit change in *x* while holding all other variables constant.

The residual sum of squares (RSS) is the sum of the squares of all residual terms in the data set. The residuals are squared when calculating the RSS to convert negative differences to a positive value before summation. The least squares approach selects values for β_0 and β_1 that will minimize the RSS in the model:

$$RSS = e_1^2 + e_2^2 + \cdots + e_i^2 \tag{4}$$

Model Considerations

Variable Selection

Arguably the most important aspect of developing a linear model is the selection of variables to use as predictors. In a regression model, these variables may be selected *a priori* based on the literature, content expertise of the dependent variable, or specified research questions.

Two considerations must be assessed in selecting variables for inclusion in the model – quality and explanatory power. When the variable is not reliable or was collected for purposes other than those specified in the linear model, it could be considered tainted and inappropriate to use for analysis. Similarly, when the variable does not have a logical or statistically significant relationship to the response variable, including it in the model can introduce bias in the estimates for other parameters. Strong knowledge of the data and the relationship being modeled are required to appropriately address both these concerns.

Categorical Variables

Linear regression cannot accept non-numeric inputs in the linear model, so categorical variables must be transformed prior to analysis (Starkweather, 2010). The simplest approach to address this issue is to assign unique numerical values to each category. However, this form of coding can make the interpretation of the values problematic. Contrast coding may be used instead to compare different levels of a categorical variable, particularly when group sizes are unbalanced (Davis, 2010). When the variable has more than two levels, it can be recoded into multiple independent variables, each of which represents a planned contrast between the different levels. The sum of the values for the contrast variables should sum to zero.

Interaction Effects

An interaction occurs when the effect of one variable depends on the value of another variable in the linear model. In many cases, this relationship represents a moderating interaction, in which the value of one variable is a function of the level of a second variable. This effect can be represented in the linear model as the product of two or more independent variables, with a unique slope:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 \ldots \beta_i x_i + e \tag{5}$$

This type of effect makes the linear model more complex, but it may reflect a real-world interaction that should be represented mathematically. When incorporating interaction effects into a linear model, it is recommended to retain lower order terms, or main effects (James et al., 2017).

Because interaction effects are directly correlated with the main effects they are derived from, it may be difficult to independently estimate the relationship between each independent variable and the dependent variable. Instead, the interaction effects and main effects change in unison, limiting the ability to interpret the linear model while holding certain variables constant.

Threats to the Model

Linear regression assumes several conditions required to interpret the linear model and test inferences from it. While regression is robust to certain violations of these assumptions, others require special attention in the modeling process.

Non-Linear Relationships

Linear regression assumes a linear relationship between the dependent variable and each independent variable. In some cases, this relationship may be non-linear, such as when there is a quadratic relationship between the two variables. When non-linear relationships occur in the model, a polynomial expression may be used while preserving the linear model itself:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \cdots \beta_i x_i + e \tag{6}$$

When incorporating higher-order polynomial terms into a regression equation, it is recommended to retain main effects when building the model (Stevens et al., 2017).

Non-Normal Distribution of Residuals

Linear regression assumes a normal distribution of the residuals in the linear model. Unlike in other parametric techniques, the assumption of normality applies only to the residuals, not the independent variables. Linear regression is robust against moderate violations of normality, but when present at higher levels, it is recommended to remove outliers, transform the data, or apply non-parametric techniques (Fox, 1997).

Heteroscedasticity

Heteroscedasticity occurs when the variance of the residuals is not equal over a range of observed values. While heteroscedasticity does not cause bias in the coefficient estimates, it does make them less precise. Lower precision in the model increases the likelihood that the coefficient estimates are further from the correct value. Moreover, heteroscedasticity tends to produce *p* values that are smaller than they should be when testing for statistical significance. This effect occurs because heteroscedasticity increases the variance of the coefficient estimates, but multiple linear regression does not detect this change. Consequently, multiple linear regression calculates the critical test statistics using an underestimated amount of variance, which can lead to the conclusion that the model is statistically significant when it is not.

Heteroscedasticity often occurs in data sets with a large range of values, or when an incorrect number of independent variables have been selected (Godfrey et al. 2006). When heteroscedasticity is present in a model, it is recommended to redefine the independent variables, transform the dependent variable using a concave or log transformation, or employ weighted regression. However, these approaches can make interpretation of the model more challenging (Stevens et al., 2017).

High Leverage Points

High leverage points in the data can pivot the fitted regression lines toward them. These points have the potential to influence model fit and must be identified relative to other values in the data set (Hadi & Chatterjee, 1986). The DFFITS statistic measures the change in predictions of the independent variable *y* when a given observation for *x* is removed from a data set. An absolute value for a DFFITS statistic that is greater than twice the square root of the product of the number of parameters and the sample size indicates a high leverage point (Williams, 1981). Alternatively, Cook's D can identify data points with unusual leverage, defined as those with a value greater than 1 (Chatterjee, 1986). When high leverage points occur in a data set, it is recommended to remove them prior to regression analysis (Stevens et al., 2017).

Outliers

Outliers fall outside the expected range given the observed data and can have a strong influence over the fitted slope and intercept in a linear model, resulting in poor model fit. Outliers tend to occur when observations are from the same data set but one that is non-normally distributed (James et al., 2017). While outliers can be removed from the data set, doing so reduces the number of observations, which in turn affects measures of dispersion that are sensitive to sample size. Outliers can be identified as values with a studentized residual greater than 2.0. Neter et al. (1996) recommend that outliers that are clearly not erroneous should be discarded rarely, when "the model is not intended to cover the special circumstances related to the outlying cases."

Overfitting

With many predictors and a limited sample size, random sampling fluctuations will allow some linear combination of the predictors to predict the value of the independent variable y. Overfitting typically occurs when the number of predictors is large or greater than the number of observations, but it can also be considered a function of high variance in the model. Regularization can introduce a small amount of bias into the linear model, leveraging the bias-variance trade-off to reduce the variance and mitigate the complications associated with overfitting (Arashi et al., 2019).

Multicollinearity

Multicollinearity occurs in a data set when two or more independent variables are highly correlated with each other. Structural multicollinearity occurs when a new feature is derived from one or more other independent variables, such as when introducing polynomial expressions into the model. Data-based multicollinearity occurs in a data set when there is a natural relationship between independent variables that are each relevant to predicting the dependent variable. It may also occur when data are inappropriately sampled from a subset of the population. Most often, data-based multicollinearity occurs in associational studies in which the variables cannot be manipulated (Chan et al., 2022).

Linear regression is robust against most degrees of structural multicollinearity when the feature is a power of other variables, but evidence is lacking for when it is the product of other variables (Arashi et al., 2019). Linear regression has likewise been shown to be robust against low and moderate data-based multicollinearity. In the presence of high multicollinearity, the recommended approach is to remove one or more of the correlated independent variables (Stevens et al., 2017). However, when the primary goal of a study is inference and not prediction, this approach may not be feasible.

When multicollinearity is present, the standard error increases, and coefficients become sensitive to small changes in the model (Guilkey, 1975). Multicollinearity also reduces the precision of the estimated coefficients, which weakens the statistical power of the regression model (Heikkilam, 1988). Under these conditions, it may not be possible to trust *p* values to identify independent variables that are statistically significant in the model (Kraha et al., 2012; Stevens et al., 2017).

Multicollinearity can be detected by inspecting a correlation matrix of the independent variables in the linear model. A variance inflation factor (VIF) quantifies how much the variance in the model is inflated by a correlation between predictors. A VIF of 1 indicates no correlation between a given predictor and the other independent variables in the model. A VIF of 5 suggests further inspection is required, while a VIF of 10 or higher indicates significant multicollinearity that requires correction (O'Brien, 2007). VIF measures can be compared after adjustments have been made to the linear model to determine if multicollinearity has been appropriately addressed.

Regularization

Regularization is an analytic approach that constrains estimates in the linear model toward zero to avoid overfitting, reducing the risk of validity shrinkage when applying a predictive model to a new set of data (Gareth et al., 2013). This approach likewise addresses multicollinearity in the linear model by adding a tuning parameter to it. Unlike multiple linear regression, regularization requires variables to be standardized before analysis.

Ridge Regression

Ridge regression, or L2 regularization, is a regularization technique that examines the relationship between dependent and independent variables when multicollinearity is present. The goal of ridge regression is to reduce the standard error caused by multicollinearity by inserting some bias into the estimates in the model. Reducing the standard error in regression estimates significantly increases their reliability (Arashi et al., 2019). This technique also helps manage the bias-variance trade-off inherent in regression modeling, in which a lower bias in parameter estimation is associated with a corresponding higher variance, and vice versa.

Multiple linear regression estimates the coefficients by minimizing the RSS through linear algebra. Given a matrix X constructed of values for an independent variable x, the coefficient can be estimated by multiplying the inverse of the cross-product matrix $X'X$ by the transpose of the X matrix and the dependent variable y:

$$\beta_i = (X'X)^{-1}X'y \tag{7}$$

When the matrix is centered and scaled, the cross-product matrix is nearly singular when the columns in it are highly correlated, leading to inflated variances and standard error. Ridge regression addresses this problem by adding a ridge parameter of the identity matrix *I* – lambda (*λ)* – to the product matrix:

$$\beta_i = (X'X + \lambda I)^{-1}X'y \tag{8}$$

This ridge parameter is a shrinkage estimator in ridge regression, producing new estimates that have been shrunk to give a value closer to the real parameters. A multiple linear regression estimate can be shrunk using a ridge estimator to improve the estimate, especially when multicollinearity is present. Geometrically, ridge regression adds a penalty, *L2*, to the model, which is equal to the sum of the square of the absolute magnitude of given coefficients:

$$L2 = RSS + \lambda + \Sigma|\beta_i|^2 \tag{9}$$

When lambda approaches zero, the penalty has no impact, and the model fitted is comparable to multiple linear regression. When lambda approaches infinity, *L2* becomes very large, and the coefficient estimates will approach – but never equal – zero.

LASSO Regression

Least Absolute Shrinking and Selection Operator (LASSO) regression, or L1 regularization, is a regression technique similar to ridge regression, but the constrained coefficients are allowed to equal zero, resulting in them being removed from the model.

This feature of LASSO regression makes it optimal for feature selection and producing simpler models. Like ridge regression, LASSO regression addresses multicollinearity in the underlying matrix algebra when selecting coefficients. Geometrically, LASSO adds a penalty, *L1*, to the model, which is equal to the sum of the absolute value of the magnitude of the coefficients:

$$L1 = \; RSS + \lambda + \Sigma|\beta_i| \tag{10}$$

When lambda approaches zero, the penalty has no impact, and no parameters are eliminated. When lambda approaches infinity, more coefficients are set to zero and eliminated. Higher values for lambda produce more bias, while lower values produce more variance.

Selecting Lambda

Selecting the appropriate value for lambda is "somewhere between science and art form" (Bobko, 2001). The literature currently recommends four approaches to selecting this value, all of which seek to maintain predictive accuracy while minimizing the mean squared error.

Ridge trace. A ridge trace plots the values of the coefficients in the regression model for various levels of lambda. Because higher values of lambda yield smaller coefficient values, the goal of a ridge trace plot is to identify the smallest value of lambda where the various coefficients stabilize. Selecting the smallest value for lambda using a ridge trace plot ensures the least amount of bias is introduced into the model. However, this approach relies on a potentially subjective visual inspection of the plot to select a value for lambda.

VIF estimation. VIF estimation seeks a value for lambda that produces VIFs for each of the coefficients that is smaller than a selected cutoff value. This value can be identified through iteration or trial and error.

Cross-validation. *K*-fold cross validation is a technique in which the data are partitioned into approximately *k* equal-sized groups, typically five or ten. For any value of lambda and each value of *j* between 1 and *k*, the ridge regression coefficients are calculated based on all the data in the partitions except the *j*th one. These coefficients are then used to predict the values in the *j*th partition and calculate the associated residuals.

Evaluating the Model

Coefficient of Determination

The coefficient of determination (R^2) is a goodness of fit measure that reflects the percentage of variance in the response variable explained by the linear model. It shows how close the data are fitted to the regression line and falls between 0 and 1, with larger values representing better model fit. The value for R^2 is calculated by dividing the amount of variance explained by the total amount of variance in the model:

$$R^2 = \frac{\Sigma(y' - \hat{y})^2}{\Sigma(y - \bar{y})^2} \tag{11}$$

Independently, the coefficient of determination is only a descriptive value. To determine whether it is statistically significant requires conducting an *F*-test based on the

F statistic. The null hypothesis for this test states that the slope in a regression model is equal to zero, while the alternative hypothesis states that it is not equal to zero. The *F* statistic is calculated as the ratio of two measures: the mean square regression (MSR) and the mean squared error (MSE), discussed more fully later in this chapter. The MSR is calculated as the RSS divided by the number of predictors, while the MSE is calculated as the average squared difference between the observed and predicted values. A large *F* statistic indicates that the regression model is effective in its explanation of the variation in the dependent variable, while an *F* statistic of zero indicates it is not.

The accepted rules of thumb for interpreting R^2 suggest that a value less than 0.3 indicates no significant effect, a value between 0.3 and 0.5 indicates a weak effect, a value between 0.5 and 0.7 indicates a moderate effect, and a value greater than 0.7 indicates a strong effect (Starnes et al., 2010). While these cutoffs are useful for interpreting the goodness of fit, they cannot be used to determine whether the coefficient estimates and predictions are biased, nor do they indicate whether the regression model provides an *adequate* fit to the data.

The coefficient of determination increases as independent variables are added to the model, so one with more predictors may superficially appear stronger based on the value of R^2 alone. Overfitting can occur when the model contains an excessive number of independent variables, interaction effects, or polynomial terms, producing a deceptively high R^2 and a decreased capacity for precise predictions (Stevens et al., 2017). The adjusted R^2 can instead be used to compare models with large or different numbers of predictors. This goodness of fit measure adjusts for the number of terms in the model, and its value increases only when the new term improves the model fit more than expected by chance alone. The adjusted R^2 likewise decreases when the term does not sufficiently improve the model fit.

The coefficient of determination is largely accepted as applicable to ridge regression and LASSO regression for evaluating the goodness of fit (Arashi et al, 2019; Gibbons, 1981; Marquardt, 1975; Lawless, 1981; Tibshirani, 1996). However, Sánchez et al. (2022) argue that the coefficient of determination used in multiple linear regression cannot be generalized for application in ridge regression, and they offer an alternative definition derived from the residual sum of squares of the initial and restricted models. This approach has not yet been explored in the literature.

However, ridge regression and LASSO regression have both been shown to reduce validity shrinkage, in which the coefficient of determination shrinks when applying a predictive model to a new set of data (Bobko, 2001). This feature of the techniques reflects the stronger predictive accuracy they offer over multiple linear regression.

Standard Error of the Estimate

The standard error of the estimate (*S*), also known as the standard error of prediction, the standard error of the regression, and the residual standard error, measures the average distance that the observed values fall from the regression line. Like the coefficient of determination, the standard error of the estimate is a goodness of fit measure in regression analysis, indicating how inaccurate the linear model is on average by using the units of the response variable. Small values are better because they indicate that the observations are closer to the regression line.

Unlike the coefficient of determination, the standard error of the estimate can be used to assess the precision of predictions. Under the empirical rule, approximately 95% of observations should fall within plus or minus two standard error units from the regression line. Because the standard error of the estimate signifies the distances between the data points and the fitted values, it is also useful for comparing the fit between different linear models as they are being constructed and evaluated.

Mean Squared Error

Mean squared error (MSE) measures the amount of bias in a statistical model by assessing the average squared difference between the observed and predicted values:

$$MSE = \frac{\Sigma(y - \hat{y})^2}{n} \tag{12}$$

In regression analysis, the MSE represents the average squared residual. When a model has no bias, the MSE equals zero. As model bias increases, the MSE increases. As the data points fall closer to the regression line, the MSE decreases. A model with less bias yields more accurate predictions and therefore better fit to the data.

Squaring the differences eliminates negative values and ensures that the MSE is always greater than or equal to zero, but it can make the interpretation of the measure less intuitive. Squaring the differences also increases the impact of larger biases in the model. The MSE calculation disproportionately penalizes larger biases than smaller ones. The root mean squared error (RMSE), calculated by taking the square root of the MSE, allows interpretation of the data using natural units in the data.

Ridge regression and LASSO estimators are constructed to have a smaller MSE than multiple linear regression. Reviewing the ratio of the estimated MSE for a particular regularization estimator to the estimated MSE for multiple linear regression can indicate the degree of improvement between the models. This difference has also been proposed as one method for selecting the tuning parameter in ridge regression and LASSO regression (Gibbons, 1981).

Interpreting Coefficients

Multiple linear regression is a form of inferential statistics in which the *p* value for each independent variable tests the null hypothesis that the variable does not correlate with the dependent variable. If there is no correlation, no statistically significant association exists between changes in the independent and dependent variables. Statistical significance is determined by calculating a *t* statistic by dividing the coefficient by its standard error. The resulting ratio indicates how many standard-error units the coefficient is away from zero, with larger units indicating statistical significance through a *t*-test.

Standard practice in multiple linear regression analysis is to retain independent variables in the model that are statistically significant at the desired confidence level, usually .05 in the social sciences (Arashi et al., 2019). When polynomial or interaction terms are present in the model and statistically significant, it is recommended to retain the main effects from which they are derived as well, regardless of whether those variables are themselves statistically significant (Stevens et al., 2017). Other variables that are not statistically significant may be removed from the model through an iterative development process until all predictors remaining are statistically significant. When an independent variable is statistically significant, the corresponding coefficient can be interpreted as the degree of change in the dependent variable for each one-unit change in the predictor, holding all other variables in the model constant. When the dependent variable has been transformed, such as using a log function, the interpretation of the degree of change must be adjusted accordingly.

Unlike multiple linear regression, there is currently no recommended mechanism for conducting significance testing with ridge regression or LASSO, as they do not produce *p* values associated with the coefficients (Price, 1977). Smaller mean squared error values imply an improvement for some coefficients when applying these techniques, but not necessarily for all. Instead, ridge regression and LASSO regression are better adapted for increasing predictive accuracy and producing solutions that suggest directions for further investigation that multiple linear regression might not suggest (Price, 1977).

The results of this study contributes to the literature with a substantive exploration of regularization in a pay equity study using empirical data obtained from a sample organization. While studies in the literature have to date explored multiple linear regression in this context, no studies have explored the application of regularization in pay equity studies. Regularization potentially offers a more robust method of analysis and increased predictive accuracy than multiple linear regression when faced with multicollinearity or many independent variables, which have yet to be discussed or presented in the literature when applied to studying pay equity.

Like multiple linear regression, ridge regression and LASSO regression allow for the analysis of multiple independent variables when predicting pay as an outcome. These factors may be both legitimate or illegitimate in the context of statutory and case law with respect to pay discrimination. Although they may offer a means to address the limitations of multiple linear regression, ridge regression and LASSO regression may not be feasible given the trade-offs they incur by introducing bias into the coefficients in the model. The lack of ability to interpret statistical significance may likewise be unacceptable in a pay equity study, limiting the utility of regularization as an alternative to multiple linear regression. The present empirical study seeks to determine which modeling technique – multiple linear regression, ridge regression, or LASSO regression – is the most feasible and the most useful for organizations and the courts in conducting a pay equity study.

Prevalence of Multiple Linear Regression

A review of the literature found several common themes with respect to why multiple linear regression has been used so prevalently in pay equity studies:

Wide Acceptance

The most frequent reason in the literature given in support of the use of multiple linear regression in pay equity studies was that it has been widely accepted as the gold standard for establishing or refuting the presence of pay discrimination since the late

1970s, particularly by the courts. Interestingly, this support for the use of multiple linear regression represents a form of circular reasoning – the technique is now widely accepted primarily *because* it has been widely accepted. While multiple linear regression has understandably received more support in evaluating disparate impact rather than disparate treatment – which is more difficult to provide statistical evidence for at the individual or small group level – no commentary explicitly noted this distinction.

Simplicity

Multiple linear regression is widely used due to its simplicity. Developing a multiple linear regression model can be easily done without specialized software or advanced statistical knowledge, and results are readily interpreted and communicated to stakeholders. The nature of multiple linear regression is also easily understood by lay readers, who may struggle to interpret more complex methods of analysis.

Parsing Sources of Variance

Multiple linear regression is widely used due to its ability to simultaneously identify different sources of variance in explaining pay as an outcome. Multiple linear regression models yield coefficients that can be interpreted as the degree of influence for specific independent variables, representing their unique and shared contributions to explaining differences in pay.

Membership in Protected Classes

Multiple linear regression is widely used because it allows for modeling membership in protected classes, which in turn may be used to support or refute claims of discrimination. The coefficients for the independent variables representing this membership can be subjected to hypothesis testing to indicate whether discrimination has occurred depending on if the result is statistically significant or not.

Support for Causality

While multiple linear regression by itself does not meet the requirements to establish a causal relationship between independent and dependent variables, it may provide evidence to support one. To establish causality, the influence of other potentially explanatory variables must be non-significant; because multiple linear regression models allow for hypothesis testing of the significance of specific sources of variance via independent variables, they allow researchers to discard potentially explanatory variables that do not influence pay as an outcome.

Alignment with Common Employment Frameworks

Multiple linear regression lends itself well to human capital theory because it allows for the inclusion of variables related to characteristics that ostensibly increase employees' value in the labor market, such as years of experience or holding a graduate degree. It also aligns well with equity-theory models, which suggest that productivity is closely related to employees' sense of fairness in the workplace. This approach requires testing the work environment for issues of inequity and inequality to better understand employee outcomes, which multiple linear regression easily supports.

Criticism of Multiple Linear Regression

A review of the literature also found several common criticisms that have been raised about the use of multiple linear regression in pay equity studies:

Interaction Effects

Multiple linear regression has been criticized in the literature because of the lack of inclusion of interaction effects, particularly between demographic variables such as legal sex and race/ethnicity. This criticism is not directed at multiple linear regression itself as a technical approach but rather at the application of it in the context of pay equity studies as a design oversight. Excluding interaction effects ignores the unique sources of variance that may be explained by intersectional identities.

Model Specification

Multiple linear regression has been criticized in the literature due to the possibility of incorrect model specification and the inclusion of poorly defined variables. For example, job value is not federally defined and can refer to the value of a position to a company, the salary band to which it belongs, or other parameters depending on the organization. Because statistical significance and coefficients generated are based on the input variables, incorrect model specification can yield results that appear reasonable but are nonsensical from a practical perspective.

Tainted Variables

Multiple linear regression has been criticized in the literature because it allows for tainted variables to be included in the model without regard to their origin or quality. Variables may be considered tainted when they are of poor quality or do not accurately represent real-world information. The inclusion of potentially explanatory variables in a model when they are under the direct management of the employer – who may have been accused of pay discrimination – complicates the use of these variables. Nevertheless, multiple linear regression models will still run correctly with these variables, yielding potentially incorrect conclusions based on the coefficients produced.

Lack of Causal Explanations

Multiple linear regression has been criticized in the literature because it does not provide direct evidence for causality. Nevertheless, the results of a multiple linear regression model are often interpreted to mean there is a causal relationship between the predictor variables and pay as an outcome. This concern suggests that when other potentially explanatory variables have been excluded based on lack of statistical significance, the implication stands that the remaining significant variables – particularly demographic variables related to membership in a protected class – are the result of intentional acts of discrimination.

Lack of Statistical Power

Multiple linear regression has been criticized in the literature because models may lack the statistical power necessary to account for variables with small group sizes. This concern is particularly salient with demographic variables such as race/ethnicity, where certain groups such as American Indians are likely to be well below the accepted thresholds for multiple linear regression. This limitation may result in the aggregation of groups to meet the required thresholds or their exclusion from the model entirely.

Proposed Alternatives to Multiple Linear Regression

To date, only a few alternatives to multiple linear regression have been proposed in the literature:

Matching Models

Matching has been suggested as an alternative to multiple linear regression in pay equity studies. Under this approach, propensity score analysis could be applied to a set of employees to identify and create pairs of similarly matched individuals, with membership in a protected class as a "treatment" variable for analysis. This approach has the advantage of handling small group sizes, which are common with certain demographic categories.

Stratification Models

Stratification combined with multiple linear regression has been proposed as an alternative approach in pay equity studies. Commonly used in epidemiology, this approach requires employee data to first be disaggregated into sub-groups, such as by occupational category, and then fitted with a regression model to identify racial inequities. The results are then compared across groups to determine if any racial inequities observed were present across all groups or a plurality of them.

Ordinal Logit Regression

Original logit regression has been suggested as an alternative to multiple linear regression when pay data are not stored as interval-ratio data, as is commonly done in dollars, but instead as equidistant ordinal categories. Although it is rare to code organizational data this way, pay may be coded as ordinal when a goal is compared against banded external compensation or census data, such as from the General Social Surveys (GSS). This approach handles the ordinal data by employing logistic regression, where the log odds of a response variable are linearly related to the independent variables.

Peters-Belson Regression

Peters-Belson regression has been proposed to address the issue of assuming discrimination affects all members in a protected class the same way. This technique suggests first fitting a regression model to members who are not in a protected class, such as White employees, and then using that model to predict pay for other employees who are in a protected class, such as People of Color. A supplemental advantage of this approach is that it resolves the issue of aggregation of members of groups to meet required regression thresholds for cell sizes.

Explored Alternatives to Multiple Linear Regression

A review of the literature found no examples of matching or stratification used in pay equity studies. These techniques have been used more extensively in medical studies and therefore may not apply to traditional employment settings. However, several instances were found of Peters-Belson regression being applied in pay equity studies and discussed from both a theoretical and practical perspective.

Gastwirth (2010) established a foundation for applying Peters-Belson regression in pay equity studies using local linear approaches. This technique was designed to resolve the issue of determining which majority employees to compare against which ones from the protected class, as including too many irrelevant employees may introduce bias into the estimated disparity (Hikawa et al., 2010). To demonstrate this approach, Gastwirth (2010) revisited *EEOC v. Shelby County Government* (1983), an early employment discrimination case that employed multiple linear regression. Using multiple linear regression and augmented local linear Peters-Belson regression, Gastwirth demonstrated that the estimated pay differential from each method was large, accompanied by small *p* values of test statistics. He therefore confirmed the court's decision that female employees in the suit were discriminated against in their pay.

Three studies were present in the literature reporting the outcomes of a pay equity study using Peters-Belson regression, interestingly all from the health care field. Jagsi et al. (2013) employed this technique to examine salary differences by gender in a cohort of early-career physician researchers. Jagsi et al. (2016) examined the compensation of cardiologists by gender. Eichelberger (2018) used Peters-Belson regression to examine pay equity in academic obstetrics and gynecology. Each of these reports examined pay outcomes by gender alone; none included race/ethnicity or interaction terms with it.

Conclusion

The literature review suggests that multiple linear regression, while both feasible and useful in pay equity studies, carries significant limitations. Models constructed to explore the relationship between compensation and potentially explanatory variables require a deep understanding of both statistical modeling and the context of pay equity to yield interpretable and applicable results. The rationale for multiple linear regression and the concerns raised about them do not vary across time or the field of study and therefore should be considered relatively uniform within the literature. Interestingly, these commentaries do not raise other technical issues commonly raised about multiple linear regression, such as failure to meet assumptions for the model, overfitting, and outlier values. These findings informed the models developed in the empirical study described in the next chapter, particularly with respect to variable selection, multicollinearity, and group aggregation.

CHAPTER 3: METHODS

In this chapter, an empirical study is described to analyze a pay equity data set using multiple linear regression, ridge regression, and LASSO regression. I begin by describing the data set and how it was prepared. Next, I outline the procedures for the study and the analytic methods used to address the research questions.

Sample Data

The present study used historical human resources microdata from a four-year research university in the Western United States to examine how ridge regression and LASSO regression perform compared to multiple linear regression in predicting employee salary while examining racial inequity in pay. The study was guided by the findings of the literature review presented in Chapter 2 to address the concerns raised in the literature about the application of multiple linear regression in pay equity studies. This study examined regularization as a new alternative to multiple linear regression in investigating pay equity.

The analysis variables for the study were selected based on affirmative defenses outlined by the EEOC and the stated compensation philosophy of the university, which were largely in alignment. The university's compensation plan places the greatest weight on job value in determining pay rates; this value is based on a classification system for each position at the university developed from the most recent job structure study. Other factors used to determine pay include annual review scores and personal qualifications, such as years of service and whether the employee holds an advanced degree. Demographic variables, such as legal sex, age, and race/ethnicity were included but hypothesized to be non-significant as the university has undergone two pay equity studies over the last ten years, presumably resulting in equitable pay for members of protected classes.

Participants

Participants in the study were 1,025 full-time permanent non-instructional non-union staff employed at the university as of the November 1, 2019, census date. Senior executive staff, athletics coaches, classified staff, and other employees whose pay is negotiated by contract or outside the standard university compensation program were excluded from the study. Part-time, seasonal, and temporary staff were likewise excluded to avoid negatively skewing salary data.

The university employs a pay-for-performance compensation model in which pay is determined largely based on an annual review score after initial pay is set. This approach is unform across units in the university, although individual supervisors have discretion in assigning performance ratings. The study therefore only considered new and continuing employees who had an annual review score on record in the most recent year relative to the census date.

The university is in a state that at the time of the census date did not have an "equal pay for equal work" law in place addressing pay inequity for People of Color, but it has since adopted one.

Procedure

An anonymized data set for the present study was provided by the human resources office at the university. These data were taken from employment, demographic, and compensation records frozen as of the selected census date. This date was chosen because it aligns with federal compliance reporting, and associated records had undergone the highest level of scrutiny and cleaning, ensuring a more credible analysis.

Each employee in the university's data warehouse is associated with one or more job records on a given census date. Only the primary position was considered for the study, provided it met the criteria outlined above. The university initiated a robust data collection program for demographic data after the selected census date. Current demographic records were therefore used in lieu of frozen historical data to support a more complete set of data for the predictors.

The university captures birthdate and legal sex at the point of hire from official government documents. Race/ethnicity is self-reported, and employees may update how they identify independently at any time. Race/ethnicity is derived from responses to two demographic questions, which are set by law to ensure consistency across official government reporting and analysis. Employees are first asked whether they identify as Hispanic or Latino, which is referred to as ethnicity. Next, employees are asked to identify their racial affiliation(s) from five categories: American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, and White. Employees who identify as Hispanic or Latino are reported as such, regardless of their racial identification. Non-Hispanic employees who identify as more than one race are reported as Multiracial, separate from the individual racial identities. Together, this information is referred to as an employee's race/ethnicity. Employees who do not self-identify their race and ethnicity are reported as White in accordance with guidance issued by the EEOC. This practice is designed to incentivize organizations to fully collect demographic data on their employees, and it presumes their workforce is predominantly White.

This approach presents several challenges from a QuantCRiT perspective. First, it assumes that race and ethnicity are natural identities and stable over time, but these census categories have evolved over the last two hundred years as social constructs, reflecting new perspectives on demography and population shifts (Omi & Winant, 2014; Zuberi, 2001). The Hispanic or Latino category is particularly problematic, as being centered on ethnicity stands in contrast to the standard definition of the word in sociology, which relates broadly to cultural affinity. The label itself reflects a compromise on how to identify the desired group of people: Hispanic refers to individuals from areas with Spanish-speaking origins, while Latino refers to individuals with origins uniquely in Latin America, who may or may not be Hispanic. Culturally, these two identities are distinct, yet they are reported together as a single category for ease of analysis. Further, individuals who identify as Hispanic or Latino *and* one or more races provided are not reported under their selected racial identities. Instead, they are subsumed under the Hispanic or Latino category, denying the intersectional nature of their racial identity. The same issue arises for non-Hispanic individuals who identify as two or more races: they are reported collectively as Multiracial, which likewise presumes a shared experience or identity that is unlikely given the multiple variations on race that are possible from the five categories available. Finally, certain individuals may find it difficult to identify as Hispanic or Latino or as one of the race categories provided. For example, individuals of North African or Middle Eastern descent would be classified as White under the current census definitions, but these individuals may not perceive themselves as such.

Zuberi (2001) notes that "the definition and classification of race are essential for the enforcement of civil rights". While the government's coding system of race/ethnicity is not perfect, it is the one used by organizations in most pay equity studies in alignment with federal reporting practices. For consistency and real-world applicability, this study therefore used it as well, with the understanding that it represents an imperfect compromise on an important variable of interest in the modeling process.

Preprocessing

Legal Sex

Legal sex was coded as a binary variable (0/1) with men as the reference category. In accordance with federal guidelines on reporting legal sex, employees at the university must be identified as men or women; no third category for legal sex is currently available.

Race/Ethnicity

Race/ethnicity posed a unique challenge for coding given the variable includes seven levels, most of which had small cell sizes and were unbalanced on comparison. To address this issue, statisticians would conventionally aggregate members of smaller categories into larger ones, such as People of Color. However, QuantCRiT suggests that aggregation of demographic variables such as race/ethnicity can be problematic as it obscures the unique experiences of individual groups, as is currently done with Hispanic or Latino and Multiracial people. It also warns against centering whiteness by comparing aggregate groups for People of Color against White individuals. The literature review likewise found that coding race/ethnicity assumes all individuals within the group share the same experience of discrimination across all levels of the organization (Fogel, 1986).

To resolve these issues, contrast coding was employed to code race/ethnicity within the regression model, with two planned contrasts: 1) Hispanic and White employees and 2) Black and White employees. While this solution did not resolve the issue of centering whiteness, it did align the modeling process with accepted standards for identifying pay inequities recommended by the courts and the EEOC. The contrasts were coded such that the values for the two race/ethnicity categories (1, -1) were summed to zero. With multiple linear regression, these contrasts would allow the model to indicate if the difference in means between the two groups were different from zero, suggesting that race/ethnicity for the categories were statistically significant. This approach stands in contrast to conventional pay equity studies, in which dummy coding is used to code race/ethnicity without concern for group sizes or balance between them.

As a trade-off, this approach does not allow for the representation of other race/ethnicity groups in the model; the two groups of interest were selected *a priori* based on which ones could yield group sizes large enough for regression analysis. Current approaches do not present an alternative that would allow for the inclusion of smaller race/ethnicity categories while meeting the assumptions for regression.

Graduate Degree

Highest degree earned was coded as a binary variable (0/1) to indicate whether the employee held a master's or doctoral degree or not as of the census date. Degrees earned were self-reported but also managed by the university in its central information system. The literature review suggested that variables that are directly controlled by the employer may be suspect or tainted; careful attention was therefore given to this variable to ensure its completeness using multiple available human resources data sources.

Descriptive Statistics

The dependent variable in the study was annual pay (salary). Independent variables included legal sex, a contrast between Hispanic and White employees, a contrast between Black and White employees, age, total years of service, years of service in the position, performance rating, whether the employee held a master's or doctoral degree, whether the employee was a supervisor, and job value. These variables were all input simultaneously into the model.

Participants in the study were 1,025 full-time permanent non-instructional non-union staff at a four-year research university whose pay was determined within the organization's standard compensation program and pay-for-performance structure. New and continuing employees without a performance rating score were excluded from the study to ensure completeness of the data in the regression model. Sixteen employees were identified as outliers based on *z*-scores for salary and removed from the data set; each of these employees was White. These outliers were removed to avoid skewing salary data for the group, which would have exaggerated potentially significant differences when comparing pay. The remaining 1,009 employees were 66.2% women and 22.7% People of Color (American Indian or Alaska Native: 0.99%, Asian: 5.75%, Black or African American: 6.05%, Hispanic or Latino: 9.42%, Native Hawaiian or Other Pacific Islander: 0.20%, Multiracial: 0.30%, White: 77.3%). Table 1 shows the largest percentage of women of color were Hispanic (8.5%), Asian (6.4%), and Black (5.5%). The average age of employees was 40.7 years, 39.0% held a graduate degree, and 26.5% were supervisors. The average total years of service for employees was 6.5 years, the average years of service in the position was 3.4 years, and the average job value was 9.0 out of 15.0. The average performance rating for employees was 3.89 out of 5.0.

Table 1

Employees by Race/Ethnicity and Legal Sex

Race/Ethnicity	*Men*	*Women*	*Total*
American Indian or Alaska Native	4	6	10
Asian	15	43	58
Black or African American	24	37	61
Hispanic or Latino	29	66	95
Native Hawaiian or Other Pacific Islander	1	1	2
Multiracial	2	1	3
White	266	514	780
Total	341	668	1,009

The mean annual salary for all employees was $60,004.24. Table 2 shows that Hispanic employees were paid on average 22.4% less than White employees, while Black employees were paid 4.1% less. Looking at pay by race/ethnicity and gender, Hispanic women were paid 27.2% less than White women, while Black women were paid 7.4% less. By contrast, Hispanic men were paid 11.7% less than White men, while Black men were paid 5.0% less. The differences in pay for Hispanic and Black employees overall therefore appear to be largely due to differences associated with women of color. The distribution of employees by job value, race/ethnicity, and gender was relatively similar, suggesting that the difference in pay was not due to Hispanic and Black employees holding positions with lower job values. Standard deviations for several race/ethnicity groups were unexpectedly large and inconsistent across race/ethnicity groups, including when split by legal sex. Although outliers are often a cause of large standard deviations, those had been removed from the data set and were therefore not likely to be the cause. Salary data were also normally distributed and homogeneous, indicating that the large standard deviations were not due to non-normal data or heterogeneity of variance. While sample sizes were unbalanced across race/ethnicity groups, the size of the group did not seem to be related with how large or small the standard deviations associated with it were. Finally, the definition for salary was carefully selected to avoid confusion about which components of compensation to include; only base pay was reported. The large standard deviations observed in the data would therefore likely appear to be due to either natural variability in the data set or other external non-statistical factors that were not reported. This observation highlights the challenge of determining the creditability of a pre-existing data set in the absence of greater context about how it was generated.

Table 2

Salary by Race/Ethnicity and Legal Sex

Race/Ethnicity	*N*	*M*	*SD*
Men			
American Indian or Alaska Native	4	$67,318.25	$8,277.96
Asian	15	$68,722.46	$19,015.70
Black or African American	24	$59,842.80	$1,769.20
Hispanic or Latino	29	$55,602.54	$19,257.87
Native Hawaiian or Other Pacific Islander	1	-	-
Multiracial	2	-	-
White	266	$62,960.68	$19,348.65
Subtotal	341	$62,287.49	$19,205.66
Women			
American Indian or Alaska Native	6	$57,156.41	$12,456.74
Asian	43	$63,686.18	$19,224.79
Black or African American	37	$56,039.55	$18,095.79
Hispanic or Latino	66	$44,078.37	$15,407.32
Native Hawaiian or Other Pacific Islander	1	-	-
Multiracial	1	-	-
White	514	$60,517.01	$19,171.64
Subtotal	668	$58,838.64	$19,401.24
All Employees			
American Indian or Alaska Native	10	$61,221.14	$11,766.17
Asian	58	$64,988.67	$19,135.14
Black or African American	61	$57,535.91	$17,995.00
Hispanic or Latino	95	$47,596.28	$17,389.47
Native Hawaiian or Other Pacific Islander	2	-	-
Multiracial	3	$60,330.11	$14,634.17
White	780	$61,350.37	$19,254.58
Grand Total	1,009	$60,004.24	$19,385.02

Correlations between continuous variables were inspected next to identify relationships between salary and each other. As Table 3 shows, moderate statistically significant correlations were observed between age and total years of service (R = 0.545) and age and years of service in the position (R = 0.506). A strong statistically significant correlation between total years of service and years of service in the position (R = 0.745) was also observed, suggesting the potential for multicollinearity. However, a later review of the VIF for each variable using a threshold value of five indicated no multicollinearity was present in the data. The strongest correlation with salary was observed with job value (R = 0.906), while the correlations with the other variables were weak. A scatterplot revealed the relationship with job value to be curvilinear, while the remaining relationships were generally linear.

Table 3

Means, Standard Deviations, and Correlations of Continuous Variables [a]

	M	*SD*	*Age*	*Total Years of Service*	*Position Years of Service*	*Performance Rating*	*Job Value*
Age	40.70	11.83	-				
Total Years of Service	6.54	6.94	0.545***	-			
Position Years of Service	3.37	2.60	0.506***	0.745***	-		
Performance Rating [b]	3.89	0.63	0.016	0.089**	0.138***	-	
Job Value [c]	8.98	2.81	0.198***	0.158***	0.168***	0.089**	-
Salary	$60,004.24	$19,385.02	0.285***	0.260***	0.254***	0.117***	0.906***

* Statistically significant at .05. ** Statistically significant at .01. *** Statistically significant at .001.

[a] N = 1,009.
[b] Ranges from 1.0 to 5.0.
[c] Ranges from 1.0 to 15.0.

Analysis

The purpose of the study was to compare multiple linear regression, ridge regression, and LASSO regression models when applied to a pay equity study. Although alternative models have been proposed and reported in the literature, regularization was selected for the present study because it has not been explored in the context of pay equity to date. Measures of model evaluation included the coefficient of determination, the standard error of the estimate, and the mean squared error. Collectively, these three measures suggested which model performed the best on an empirical data set.

The study began with exploratory data analysis to determine the suitability of the selected data set for analysis and to ensure it satisfied all assumptions for multiple linear regression. Linear relationships between the selected predictor variables and salary were examined using scatterplots and correlation matrices.

Next, frequencies were examined to guide the re-coding of categorical variables. Contrast coding was employed rather than dummy coding as the study does not require assigning different weights to the various levels of the categorical variables, and group sizes are expected to be unbalanced. Homogeneity of variance was also examined to determine whether the regression model would be robust to different group sizes, supporting the use of contrast coding. Group sizes were reviewed to confirm the data set contained sufficient observations for multiple linear regression given the selected number of predictors.

Once pre-modeling assumptions for multiple linear regression were satisfied, a multiple linear regression model was developed with nine independent variables recommended in the literature and associated with affirmative defenses against charges of discrimination under Title VII and the EPA: legal sex, the two race/ethnicity contrasts, age, total years of service, years of service in the position, whether the employee held a master's or doctoral degree, and performance evaluation score, whether the employee was a supervisor, and job value. Interaction terms between legal sex and race/ethnicity were also examined.

The data were then standardized in preparation for the development of the ridge regression and LASSO models. Lambda was selected for these two models using an automated, iterative program that identified the best performing value based on the mean squared error it generated. This value was compared against a ridge trace to confirm its effectiveness in the model. This value was then used to develop ridge regression and LASSO regression models with the same predictors as the multiple linear regression model.

Multicollinearity was examined by calculating the VIF for each predictor and generating a correlation matrix between all the independent variables. A moderate VIF of five was used as a threshold for identifying problematic correlational relationships.

The coefficient of determination was used to indicate the strength of each model and compare their effect sizes. Of particular interest was the comparison between multiple linear regression and LASSO. Because LASSO allows certain coefficients in the model to reach zero, the model might have had fewer variables than multiple linear regression and therefore yield a lower R^2. Residual plots were examined for each model to ensure errors were uncorrelated and model diagnostics could be appropriately interpreted.

The standard error of the estimate and mean squared error were used to evaluate model accuracy. The analysis would determine whether regularization applied to a pay equity study conformed to expectations documented in the literature for improved model accuracy. The value for lambda was selected in part by identifying the value in the model that yielded the lowest mean squared error. Regularization models were expected to yield higher mean squared error values than multiple linear regression due to introducing bias into the model.

Software

Analysis was conducted in R (R Core Team, 2021). Regression analysis was conducted using the *glmnet* package (Friedman et al., 2010). Figures were produced using the *ggplot2* package (Wickham, 2009).

CHAPTER 4: RESULTS

In this chapter, the results from the empirical study conducted for this research are reported and synthesized.

Research Question 1

Which of the three models yields the highest coefficient of determination (R^2), the lowest standard error of the estimate (S), and the lowest mean squared error (MSE)?

In preparation for modeling, continuous variables were standardized, a requirement for regularization techniques such as ridge regression and LASSO. The literature offered little direct guidance on standardizing dummy variables, such as legal sex and whether the employee held a graduate degree. For this study, dummy variables were not standardized; neither were the two contrast variables. The rationale for this decision was that dummy and contrast variables representing categorical data do not have a natural mean or standard deviation, which standardization requires. However, this decision concurrently introduced the possibility of inadvertently giving greater weight to the continuous variables in the model.

A multiple linear regression model was developed next for comparison with the ridge regression and LASSO models. To address the curvilinear relationship between salary and job value, a quadratic term for job value was introduced into the regression model. An initial multiple linear regression model indicated moderate heteroskedasticity. To address this violation of the assumptions for linear regression, a log transformation was performed on salary before standardizing the variables for analysis. The final multiple linear regression model was statistically significant, $F(13, 995) = 893.6$, $p < .001$, $R^2 = 0.9211$, accounting for 92% of the variance observed in the data. The model yielded a standard error of the estimate of 0.2827 and an MSE of 0.0788. A review of the QQ plot, shown in Figure 1, indicated the residuals were normally distributed, and heteroskedasticity was no longer present, as shown in Figure 2.

Figure 1

QQ Plot – Multiple Linear Regression Model

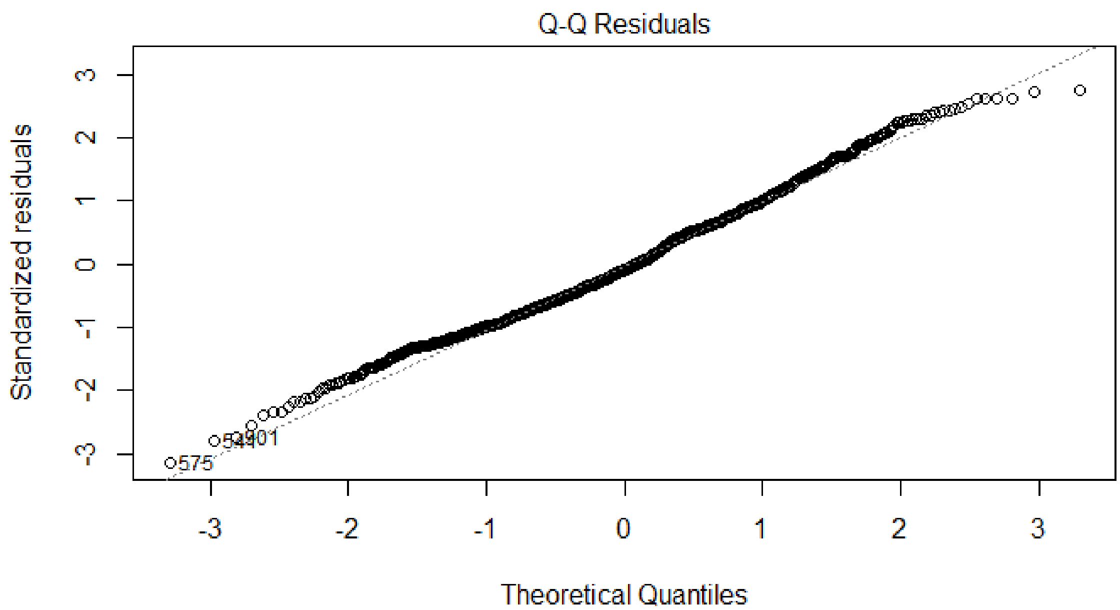


Figure 2

Residuals vs. Fitted Plot – Multiple Linear Regression Model

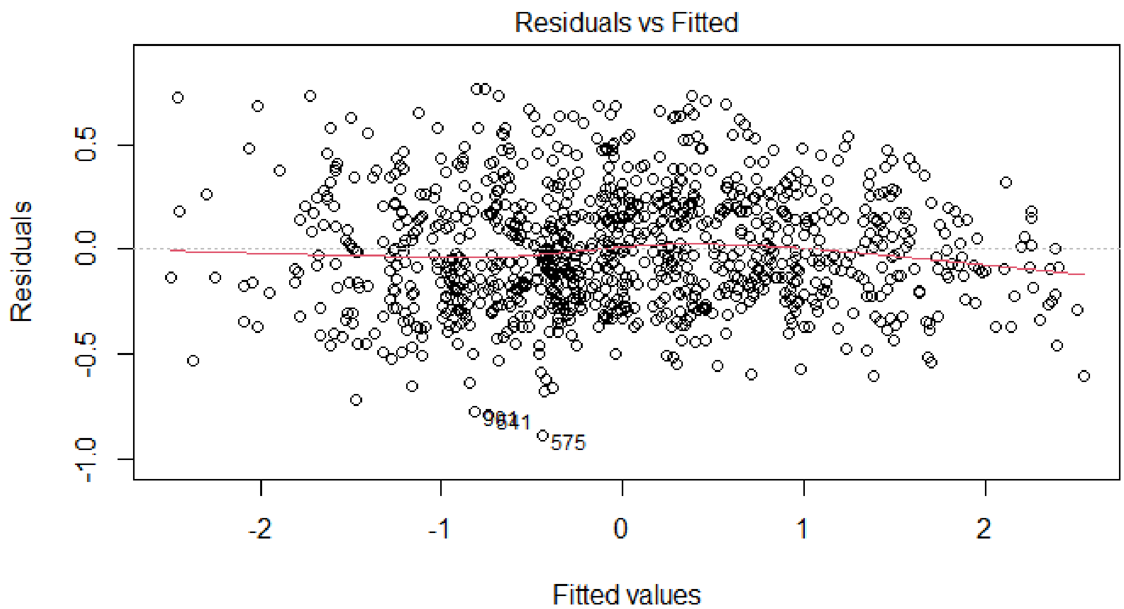


Figure 3 shows that the predicted values for the multiple linear regression model fell close to the line with no fanning. The model indicated that all variables except years of service in position and whether the employee held a graduate degree were statistically significant at $p < .05$. Table 4 shows that the most important variables were job value, the contrast between Hispanic and White employees, and the contrast between Black and White employees, each at $p < .001$. The interactions between sex and the race/ethnicity contrasts were also statistically significant, although the coefficients were smaller.

Figure 3

Observed vs. Predicted Salary – Multiple Linear Regression Model

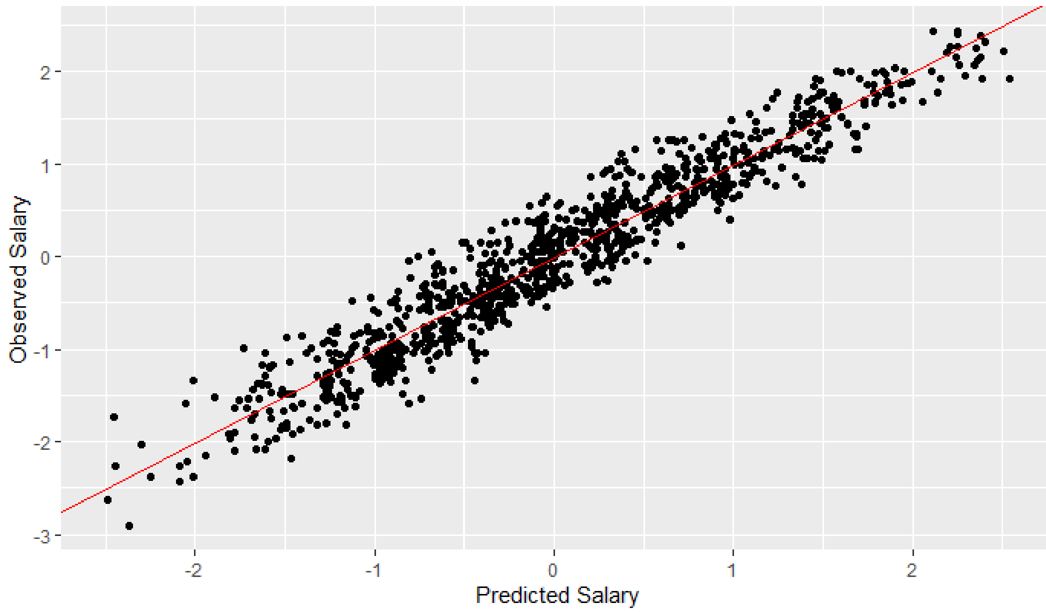


Table 4

Summary of Multiple Linear Regression Model

Variable	Estimate	*SE*	*t*	*p* value
Intercept	-0.1932	0.0209	-9.326	< .001***
Legal Sex	0.1071	0.0307	3.490	< .001***
Age	0.0477	0.1134	4.208	< .001***
Race Contrast – Hispanic vs. White	0.5846	0.0284	20.581	< .001***
Race Contrast – Black vs. Hispanic	-0.3405	0.0327	-10.409	< .001***
Legal Sex * Race Contrast – Hispanic vs. White	-0.2162	0.0493	-4.385	< .001***
Legal Sex * Race Contrast – Black vs. White	0.1367	0.0537	2.547	.011*
Total Years of Service	0.0760	0.0141	5.393	< .001***
Position Years of Service	-0.0067	0.0138	0.480	.6312
Performance Rating	0.0231	0.0092	2.507	.0123*
Graduate Degree (Y/N)	0.0082	0.0193	0.423	.672
Job Value	0.8829	0.0101	87.458	< .001***
Job Value^2	0.0907	0.0091	10.014	< .001***
Supervisor (Y/N)	0.0665	0.0228	2.920	.004**

* Statistically significant at .05. ** Statistically significant at .01. *** Statistically significant at .001.

Ridge regression was performed next on the data. Using cross-validation, the optimal lambda that minimized the MSE was 0.0918. The ridge regression yielded an R^2 of 0.9124, a standard error of the estimate of 0.3084, and an MSE of 0.0938. As there is currently no accepted method of conducting hypothesis testing with ridge regression, no tests of statistical significance could be performed on the model. A review of the QQ plot, shown in Figure 4, indicated the residuals were normally distributed, and heteroskedasticity was not observed, as shown in Figure 5. As with the multiple linear regression model, the predicted values for salary fell close to the line with no fanning, as shown in Figure 6. The most important variables in the model based on the size of the coefficients were job value, the contrast between Hispanic and White employees, whether the employee was a supervisor, and the contrast between Black and White employees.

Figure 4

QQ Plot – Ridge Regression Model

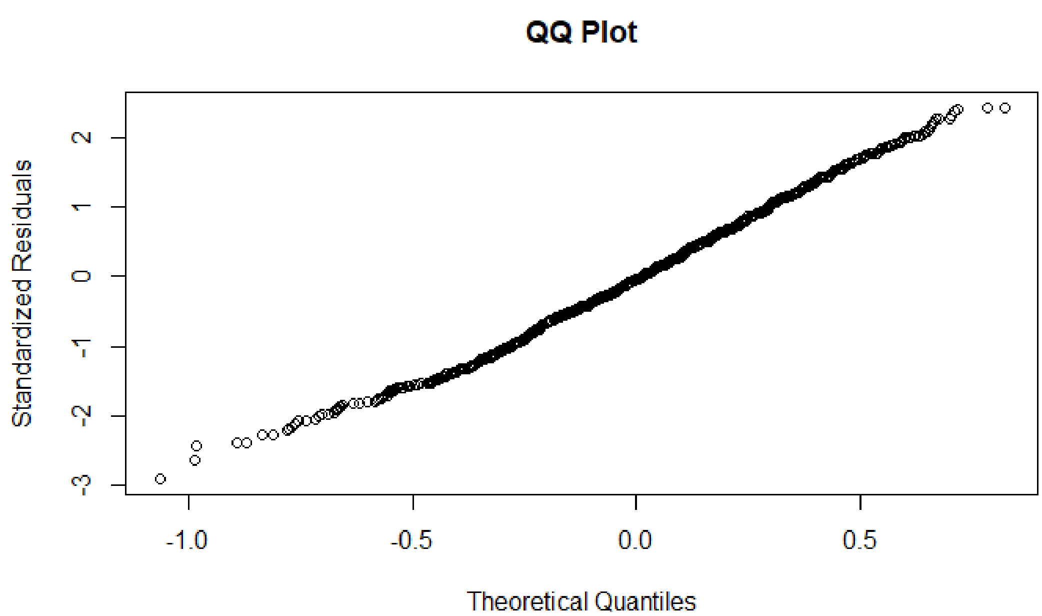


Figure 5

Residuals vs. Fitted Plot – Ridge Regression Model

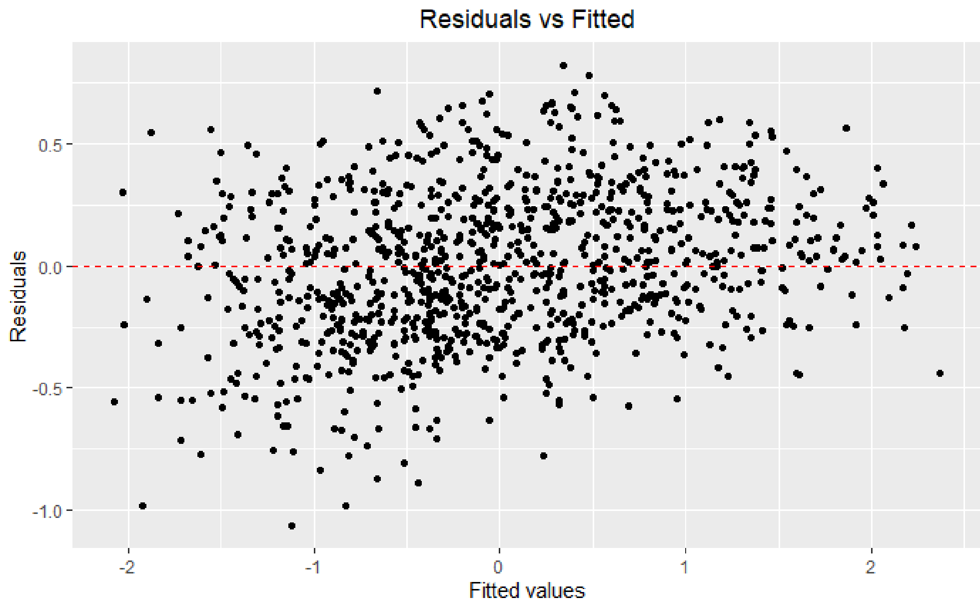


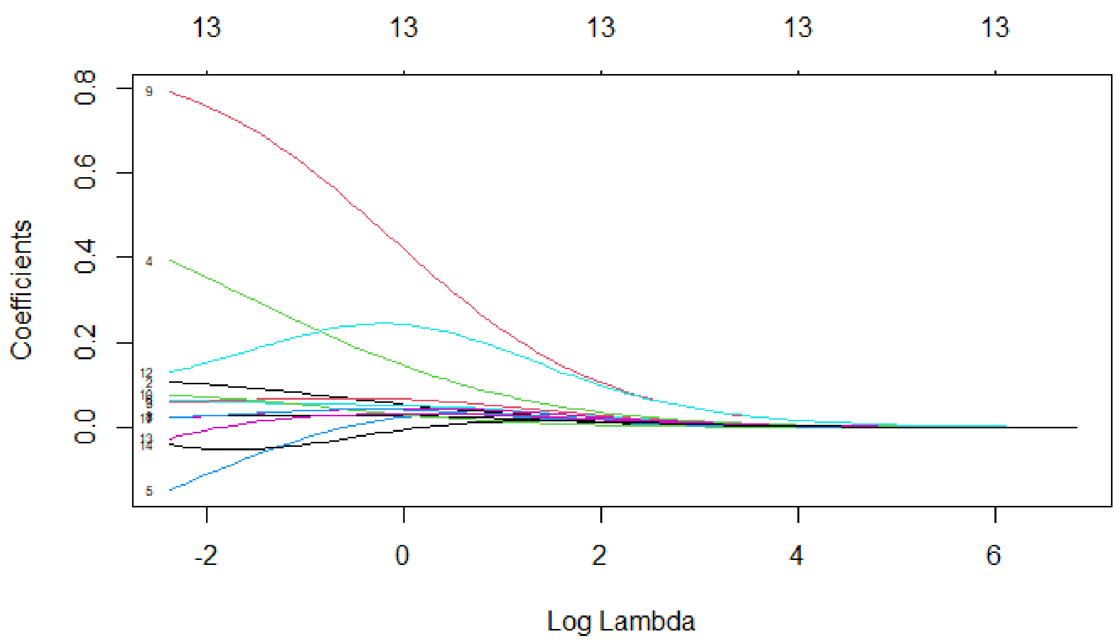
Figure 6

Observed vs. Predicted Salary – Ridge Regression Model

Figure 7 shows the effect of penalization via ridge regression on each coefficient in the model. As additional bias is introduced, indicated by the log of lambda, the coefficients begin converging towards – but never reach – zero. The variables that converge the slowest are the most important in the model. Mirroring the coefficients from the model, these variables were again job value (9), the contrast between Hispanic and White employees (4), whether the employee was a supervisor (12), and the contrast between Black and White employees (5).

Figure 7

Effect of L2 Regularization on Model Coefficients



LASSO regression was performed last on the data. Using cross-validation, the optimal lambda that minimized the MSE was 0.0008. LASSO yielded an R^2 of 0.9209, a standard error of the estimate of 0.2693, and an MSE of 0.0790. Like ridge regression, there is currently no accepted method of conducting hypothesis testing with LASSO, so no tests of statistical significance could be performed on the model. A review of the QQ plot, shown in Figure 7, indicated the residuals were normally distributed, and heteroskedasticity was not observed, shown in Figure 8. As with the multiple linear regression and ridge regression models, the predicted values for salary fell close to the line with no fanning, as shown in Figure 9. The most important variables in the model based on the size of the coefficients were job value, the contrast between Hispanic and White employees, the contrast between Black and White employees, and the interaction between legal sex and the contrast between Hispanic and White employees.

Figure 8

QQ Plot – LASSO Model

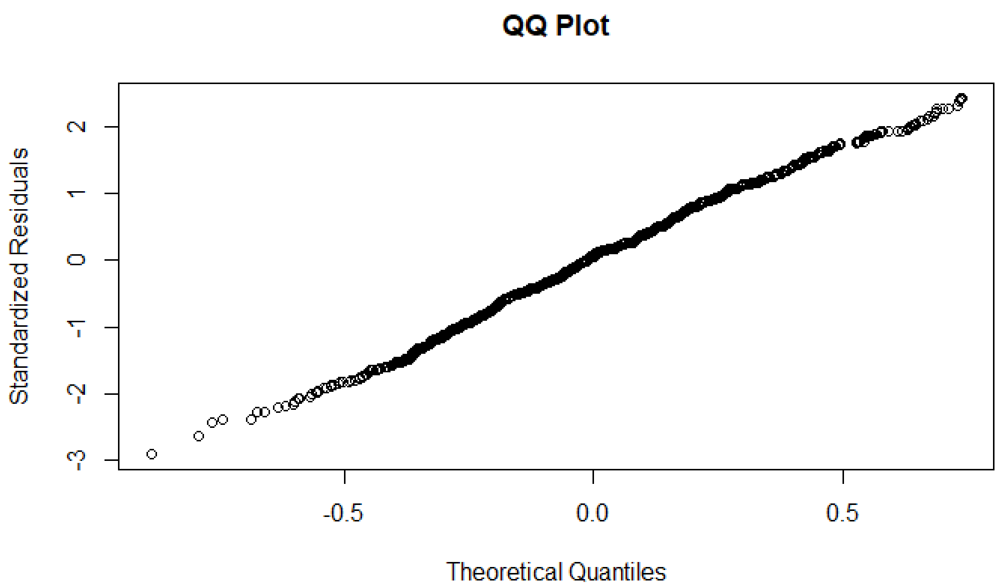


Figure 9

Residuals vs. Fitted Values – LASSO Model

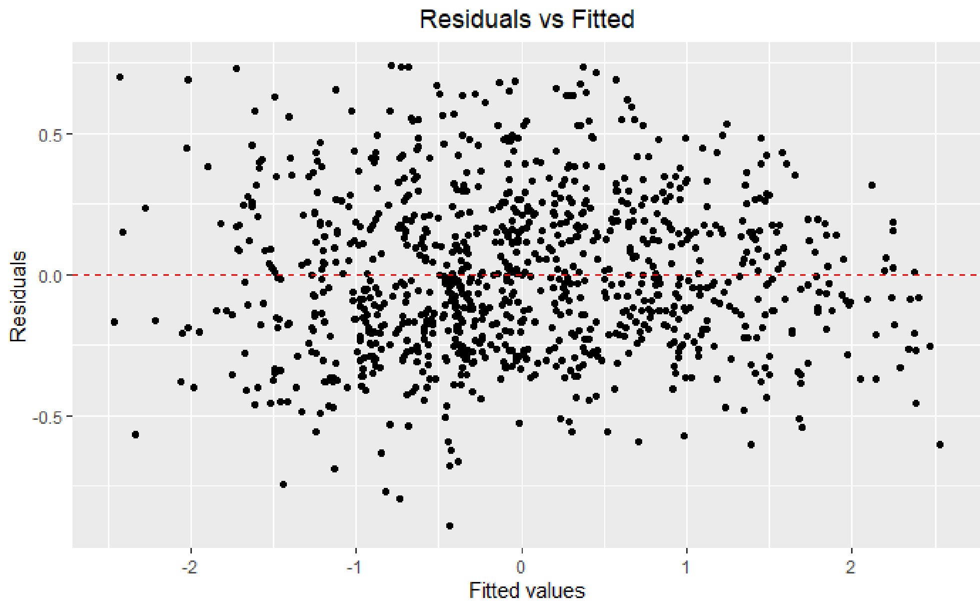


Figure 10

Observed vs. Predicted Salary – LASSO Model

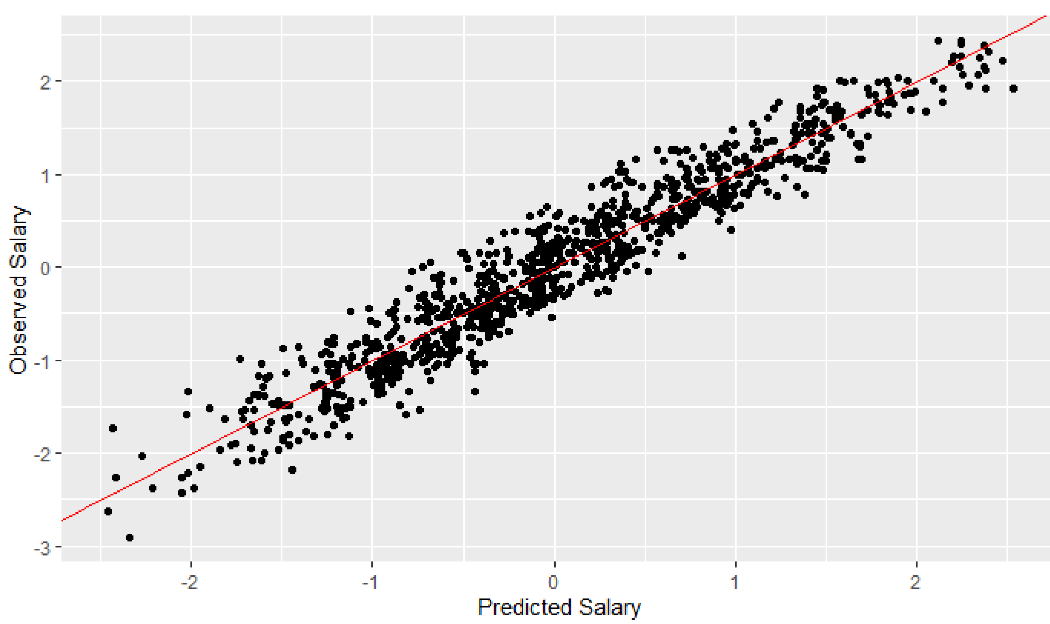


Figure 11 shows the effect of penalization via LASSO on each coefficient in the model. As additional bias is introduced, the coefficients begin converging towards zero, which they are allowed to reach with LASSO. As with ridge regression, job value (9), the contrast between Hispanic and White employees (4), and the contrast between Black and White employees (5) converged more slowly than with other variables. However, with LASSO, the fourth slowest converging variable was the interaction between legal sex and the contrast between Hispanic and White employees (13).

Figure 11

Effect of L1 Regularization on Model Coefficients

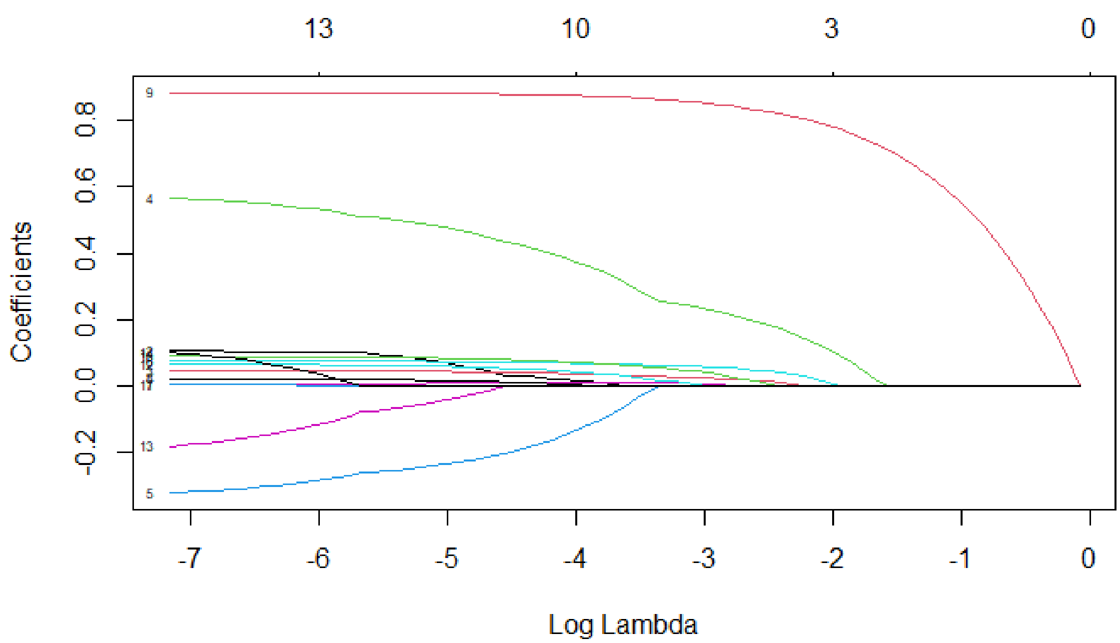


Table 5

Intercept and Standardized Coefficients by Regression Model

	Multiple Linear Regression	Ridge Regression	LASSO Regression
Intercept	-0.1932***	-0.1935	-0.1925
Legal Sex	0.1071***	0.0605	0.1038
Age	0.0477***	0.0571	0.0472
Race Contrast – Hispanic vs. White	0.5846***	0.2907	0.5520
Race Contrast – Black vs. White	-0.3405***	-0.0655	-0.3058
Legal Sex * Race Contrast – Hispanic vs. White [a b]	-0.2162***	0.0114	-0.1519
Legal Sex * Race Contrast – Black vs. White	0.1367*	-0.0309	0.0739
Total Years of Service	0.0760***	0.0672	0.0762
Position Years of Service	-0.0067	0.0268	0.0072
Performance Rating	0.0231*	0.0259	0.0228
Graduate Degree (Y/N)	0.0082	0.1460	0.0046
Job Value [c d]	0.8829***	0.8000	0.8830
Job Value^2	0.0907***	0.0780	0.0902
Supervisor (Y/N)	0.0665**	0.0878	0.0630

* Statistically significant at .05. ** Statistically significant at .01. *** Statistically significant at .001.

[a] Second slowest converging variable in ridge regression.
[b] Second most important variable in LASSO regression.
[c] Slowest converging variable in ridge regression.
[d] Most important variable in LASSO regression.

Table 5 provides a comparison of the coefficients yielded by the multiple linear regression, ridge regression, and LASSO models. LASSO returned coefficients close to those of multiple linear regression, which was expected given the small value for lambda in the model. In each of the three models, job value was the most important variable.

Table 6 shows that the multiple linear regression model returned the highest coefficient of determination (R^2 = 0.9211), although only slightly higher than the LASSO model (R^2 = 0.9209). The LASSO model returned the lowest standard error of the estimate (S = 0.2693), while the multiple linear regression model returned the lowest MSE (0.0788).

Table 6

Comparison of Regression Model Performance

Measure	Multiple Linear Regression	Ridge Regression [a]	LASSO Regression [b]
R^2	0.9211	0.9124	0.9209
S	0.2827	0.3084	0.2693
MSE	0.0788	0.0938	0.0790

[a] Lambda = 0.0918.
[b] Lambda = 0.0008.

Research Question 2:

To what extent does racial disparity in compensation exist within the university's workforce, and what are the measurable factors contributing to this inequity?

Multiple linear regression found that the contrast variable modeling the difference in mean salary between Hispanic and White employees was statistically significant, β = 0.5846, $t(995) = 20.581$, $p < .001$. Likewise, the contrast variable modeling the difference in mean salary between Black and White employees was statistically significant, β = -0.3405, $t(995) = -10.409$, $p < .001$. Looking at interaction terms, the variable modeling the difference in mean salary between Hispanic women and White women was statistically significant, β = -0.2162, $t(995) = -4.385$, $p < .001$. Finally, the variable modeling the difference in mean salary between Black women and White women was statistically significant, β = 0.1367, $t(995) = 2.547$, $p = .011$. Because the coefficients for these variables were statistically significant, it can be concluded that race/ethnicity was a significant factor in predicting pay in the model. The interactions between legal sex and the two race/ethnicity contrast terms were also statistically significant, indicating further that Hispanic and Black women also experienced a unique form of pay inequity related to their intersectional identity.

While regularization does not allow testing for statistical significance, coefficients can be interpreted based on their relative importance in the model. In the ridge regression model, the two contrast terms and the interaction terms were among the most important variables after job value. A trace plot of the coefficients likewise confirmed that these variables were among the slowest to converge as lambda approached zero, reflecting their importance in the model. Although LASSO regression yielded different coefficients from ridge regression, the relative size was comparable. That is, after job value, the two contrast terms and the interaction terms were the most important in the model, which was likewise reflected in a trace plot. The two regularization models therefore also indicate that pay inequity by race/ethnicity is present at the university for Hispanic and Black employees overall, as well as for Hispanic and Black women as particular subgroups.

Independent *t*-tests with Bonferroni correction found that Hispanic employees were paid significantly less (M = \$47,596, SD = \$17,482) than White employees (M = \$61,350, SD = \$19,267), $t(873) = -6.633$, p < 001. Hispanic women were also paid significantly less (M = \$44,078, SD = \$15,525) than White women (M = \$60,517, SD = \$19,190), $t(578) = -6.682$, $p < .001$. However, race was not a significant effect for Black employees despite them being paid less (M = \$57,536, SD = \$18,144) than White employees (M = \$61,350, SD = \$19,267), $t(839) = -1.495$, $p = .135$. Similarly, the interaction between race and gender was not a significant effect for Black women, despite them being paid less (*M* = \$56,040, *SD* = \$18,345) than White women (*M* = \$60,517, *SD* = \$19,190), $t(549) = -1.375$, $p = .170$.

The multiple linear regression, ridge regression, and LASSO regression models all indicated that race was an important factor in predicting salary for Hispanic and Black employees, as well as for Hispanic and Black women uniquely. However, independent *t*-tests found that only Hispanic employees and Hispanic women employees were paid significantly less than their White peers. It can therefore be concluded that while racial inequity is present at the university for Hispanic and Black employees, it only manifests in specific pay inequities for Hispanic employees and Hispanic women employees. If pay discrimination is present at the university, it stands in contrast with legal protections for People of Color under Title VII and under the Equal Pay Act for women. For employees who are both women and People of Color, this finding further suggests a unique violation of both laws for them. Had the focus of this study been on racial identity only, the intersectional nature of this violation would have gone unreported.

CHAPTER 5: DISCUSSION

In this chapter, the findings from the literature review and empirical study conducted for this research are synthesized, followed by a discussion of the implications of the proposed research questions. Limitations of the research, recommendations for further study, and reflections by the researcher are also provided.

Summary of the Study

A literature review was conducted to examine conflicting views on the use of multiple linear regression in pay equity studies. The review found agreement on several reasons why multiple linear regression has been used so prevalently in pay equity studies. The most often cited reason for the use of multiple linear regression was that it has been widely accepted by the courts and practitioners in the field since the late 1970s. This technique has been widely accepted in part because it allows for the parsing of multiple sources of variance, including membership in protected classes. Multiple linear regression therefore lends itself well to studies where the objective is inference, either to confirm or dispute unlawful discrimination or to ensure an organization's compensation philosophy is reflected in actual pay practices. The process and results of multiple linear regression are likewise easy to understand and communicate to stakeholders who may not have a statistical background.

Multiple linear regression has also been widely accepted because it provides support – but not evidence – for a causal relationship between pay as an outcome and potentially discriminatory acts. This modeling technique allows practitioners to identify potential sources of variance that are not statistically significant, excluding them as alternative explanations of differences in pay. Likewise, multiple linear regression in this application lends itself well to commonly used theoretical frameworks in employment studies, such as human capital theory and equity-based theory, which seek to identify the relationship between employee qualifications and pay as a decision-based outcome.

A literature review also revealed several shared criticisms of multiple linear regression in pay equity studies. The most common reason given was model specification, particularly the selection of the appropriate potentially explanatory variables for modeling. Authors in the commentaries noted the frequent lack of interaction terms, perhaps because of the lack of statistical power to analyze small group sizes in multiple linear regression. The commentaries also spoke about the potential inclusion of tainted variables, which may be inaccurate or incomplete. Worse, these variables may be under the control of an organization accused of discrimination, representing a gross conflict of interest.

Despite these concerns and limitations of multiple linear regression, few alternatives were found to have been discussed or explored in the literature. Matching models and stratification models were recommended in the mid-1980s as potential alternatives but have so far not been applied in reported pay equity studies. Peters-Belson regression has also been proposed as an alternative, but primarily as a conceptually different way of employing multiple linear regression rather than a technically distinct one. Several studies in the health care field were found that had successfully employed Peters-Belson regression in a pay equity study, although only for gender.

An empirical study was conducted next, with methods guided by the results of the literature review. The study used historical human resources microdata from a four-year research university in the Western United States to examine how ridge regression and LASSO regression perform compared to multiple linear regression in predicting employee salary. Participants included 1,009 full-time permanent non-instructional non-union staff employed at the university as of the November 1, 2019, census date. Senior executive staff, athletics coaches, classified staff, and other employees whose pay is negotiated by contract or outside the standard university compensation program were excluded from the study. Part-time, seasonal, and temporary staff were likewise excluded from the study.

Independent variables in the study included legal sex, age, race/ethnicity, total years of service, years of service in the position, performance rating, whether the employee held a graduate degree, whether the employee was a supervisor, and job value. Race/ethnicity posed a unique challenge for coding in the regression models. Due to small cell sizes and unbalanced groups, only two race/ethnicity categories were considered: Hispanic and Black employees. Data for these employees were input using contrast coding, in which they were compared against their White peers. The goal of the contrast coding approach was to determine whether the difference in mean pay for each of the two groups was statistically significantly different from zero. Interaction terms for legal sex and the two contrasts were also introduced into the model to account for the unique experiences of women of color. While contrast coding addressed several technical concerns in model development, it still centered on whiteness and removed other race/ethnicity categories from consideration.

A comparison of the three models found that multiple linear regression yielded the highest coefficient of determination and the lowest mean squared error. This model found that all variables except years of service in the position and whether the employee held a graduate degree were statistically significant at $p < .05$. Because the two contrast variables were statistically significant at $p < .001$, and Hispanic and Black employees were paid on average significantly less than their White peers, it was inferred that racial inequities are present at the study university. Whether these inequities are due to unlawful discrimination or disparate impact would require supplemental study beyond the scope of this research.

Ridge regression performed the worst of the three models, although only marginally. LASSO produced the lowest standard error of the estimate, indicating it yielded the best predictive accuracy of the three models. Because multicollinearity was not present in the data, the primary benefit of these two regularization models was to prevent overfitting, which LASSO was found to do. The coefficients it yielded were close to those produced by multiple linear regression, which was expected given the low value selected for the optimal tuning parameter lambda.

Each of the three models selected for study in this dissertation indicated that racial inequity in pay was present at the university. However, the study also found that the variables related to employee qualifications and the job value of the position were also important, indicating that the university's compensation philosophy was at least partially reflected in predicted pay.

Major Findings by Research Question

Research Question 1

Which of the three models yields the highest coefficient of determination (R^2), the lowest standard error of the estimate (S), and the lowest mean squared error (MSE)?

The coefficient of determination is a goodness-of-fit measure that describes the amount of variance explained by a model. Comparing multiple linear regression, ridge regression, and LASSO, the empirical study found that multiple linear regression yielded the lowest coefficient of determination, accounting for 92.1% of the variance in the data. However, this value was not significantly different from that of ridge regression (91.1%) and LASSO (92.1%). The small difference is not surprising given the relatively low values selected for the tuning parameter lambda (ridge regression = 0.0918, LASSO = 0.0008).

The standard error of the estimate is a goodness-of-fit measure that describes how well the model fits the data by assessing its predictive accuracy. Examining performance for the three models, the empirical study found that LASSO yielded the lowest standard error of the estimate, 0.2693, compared to 0.2827 for multiple linear regression and

0.3084 for ridge regression. The higher value for ridge regression is expected given the higher value for the tuning parameter lambda, thereby introducing more bias into the model. These findings suggest that LASSO may leverage the bias-variance trade-off offered by regularization more effectively to deliver greater predictive accuracy using regression to understand pay as an outcome. While the standard error of the estimate is normally interpreted using the natural units of the dependent variable, the process of log transforming pay and standardizing it makes analysis difficult, weakening its utility.

MSE is a general measure of the residuals in a regression model, describing how far on average predicted values are from corresponding observations in the data. The empirical study found that multiple linear regression yielded the lowest MSE of the three models (0.0788), although only marginally higher than LASSO (0.0790). Ridge regression yielded the highest MSE (0.0938), again due to the introduction of more bias into the model via the tuning parameter lambda. Conversely, the MSE values for multiple linear regression and LASSO were so similar due to the low value selected for lambda in the LASSO model.

Research Question 2

To what extent does racial disparity in compensation exist within the university's workforce, and what are the measurable factors contributing to this inequity?

The multiple linear regression, ridge regression, and LASSO models all indicated that racial inequity in pay was present at the university. The multiple linear regression delivered statistically significant results for both racial contrast variables examined in the study, indicating that Hispanic and Black employees were being paid less than their White peers. More specifically, the results of the multiple linear regression model suggested that the differences in pay were not due to random fluctuations in the data.

Ridge regression and LASSO – the two regularization models selected for this study – also indicated that racial inequity in pay was present at the university. Rather than relying on statistical inference, the interpretation of these models was based on the magnitude of the coefficients and their relative importance in analyzing variables as lambda converged toward zero using trace plots. The relative strength of coefficients examined in each model was consistent across the models, although the actual values varied, with the contrast between Hispanic and White employees being one of the largest. LASSO delivered these results with higher predictive accuracy than multiple linear regression or ridge regression, which is an additional benefit of the technique. It also carries the potential for feature selection in future studies by identifying variables to be dropped from the model. Ridge regression appeared to offer little advantage over multiple linear regression beyond addressing potential multicollinearity.

The interactions between legal sex and the two contrast variables were also statistically significant in the multiple linear regression model, indicating that Hispanic and Black women were paid statistically significantly less than White women. This finding was mirrored in the ridge regression and LASSO models, which also found these interaction terms to be among the most important in the model. Including the interaction terms – as recommended in the literature but commonly omitted in practice – therefore yielded an important insight into the unique experience of women of color at the university with respect to pay inequity.

The university examined in this study operates under a pay-for-performance philosophy, in which pay is determined in part based on annual performance, reflected in a performance review rating. Interestingly, while the performance rating was found to be statistically significant in the multiple linear regression model, it was also one of the weakest variables and significant at only $p < .05$. Because performance ratings were found to be less important than the racial contrasts in all three models, the study indicates that differences in pay cannot be fully explained by differences in performance.

Pay at the university is also determined by a combination of factors related to employee qualifications and the nature of the position, particularly job value. This last variable was found to be the most significant in all three models, suggesting that while pay inequity was present at the university, its compensation philosophy was also reflected, as expected; job value, total years of service, and whether the employee was a supervisor were all statistically significant in the multiple linear regression model. However, years of service in the position and whether the employee held a graduate degree were not significant, indicating the university's compensation philosophy is only partially reflected in the pay of its employees. This finding was mirrored in the ridge regression and LASSO models.

Based on the three models examined in this study, the university should more closely examine how compensation for employees is set, with the understanding that

Hispanic and Black employees – and potentially other people of color – are currently underpaid compared to their White peers in a way that cannot be explained by chance or factors related to the position or employee. A recommended next step would be to identify individual employees whose predicted pay could be classified as outliers with respect to mean or median pay for their job value. The background of these employees could be examined in more detail to determine whether there are factors not captured in the model that would explain the observed differences in pay. However, a comprehensive review of the pay equity structure would likely still be required to address the inequity found in this study.

Feasibility of the Models

Developing an effective regression model requires some understanding of statistical principles and proficiency in using software and programming packages, regardless of the technique employed. However, developing a ridge regression or LASSO model requires more technical skill than multiple linear regression due to the high degree of programming and manual calculations and plotting required. Moreover, these modeling approaches require specific knowledge of how to select a value for the tuning parameter lambda; differing values for the parameter can produce substantially different results in the model.

The literature review found that model specification – the selection of the appropriate variables for analysis – was a primary concern in conducting a pay equity study. Ironically, it is the simplicity of multiple linear regression that makes it so dangerous with respect to this concern: a multiple linear regression model can be easily programmed and interpreted without regard to whether the values selected for the study are reasonable or appropriate, which the courts have noted and cautioned against. This limitation highlights the important role both statistical acumen and content expertise play in conducting a pay equity study.

However, the empirical study suggests that multiple linear regression remains the most feasible modeling technique for a pay equity study due to its simplicity and ease of development. While it can be built in programming packages in R, Python, or STATA, it can also be run using graphical user interfaces (GUI) in Excel and SPSS. Given that organizations may not have the resources to hire a statistician or consulting company to conduct a pay equity study externally, multiple linear regression allows them to reduce costs by completing the work in-house using available staff resources.

Nevertheless, while multiple linear regression remains the most feasible technique for conducting a pay equity study – at least in comparison to ridge regression and LASSO – assumptions for linear regression must still be reviewed and confirmed to be met. For example, the empirical study found heteroskedasticity in an initial multiple linear regression, which, if left unchecked, would have invalidated the model. Log transformation was selected in this study to address the problem, at the expense of ease of interpretation of the results. Less mindful practitioners might have ignored the issue entirely and continued with an interpretation of the model without addressing how or why doing so would be problematic.

From a modeling perspective, ridge regression and LASSO require standardizing the variables and selecting a value for the tuning parameter lambda, which can be described as both an art and a science. These techniques also require programming a regression model that is more complex than multiple linear regression. To evaluate the models, values such as the coefficient of determination, the standard error of the estimate, and the MSE must be calculated programmatically; packages such as *glmnet* do not provide these natively. Further, they do not provide direct support for diagnostic plots to evaluate the normality of the residuals or to detect heteroskedasticity, which are needed to ensure assumptions for linear regression have been met.

From a practical perspective, multiple linear regression was found to be the most feasible for pay equity studies because it is the simplest to implement. Regression models may be developed by lawyers, human resources consultants, or analysts as part of a pay equity study, each with varying degrees of statistical and programming knowledge. Multiple linear regression is the most accessible modeling approach, the easiest to develop, and the easiest to interpret, posing the fewest technical challenges and barriers to using it in a pay equity study.

Utility of the Models

A review of the literature found that multiple linear regression was the preferred technique for conducting a pay equity study due to its simplicity. As noted in previous chapters, the ease with which a multiple linear regression analysis can be conducted is both a benefit and a drawback: while the models can be developed and understood by lay people without difficulty, they can also be misapplied and misinterpreted just as easily.

The literature review also found that multiple linear regression is the preferred technique for conducting a pay equity study due to its ability to simultaneously parse various sources of variance and to determine whether membership in a protected class is statistically significantly related to pay as an outcome. Two specific race/ethnicity groups – Hispanic and Black employees – were examined in the empirical study to analyze these benefits from a technical perspective. Due to small cell sizes and unbalanced groups across race/ethnicity categories, these two groups were entered into the model using contrast coding, where Hispanic and Black employees were coded with a value that summed to zero when compared against a value for White employees.

The purpose of the contrast coding was to determine whether the difference in mean pay between Hispanic and Black employees and their White peers was statistically significantly different from zero. The empirical study found that the coefficients for both contrast variables were statistically significant, indicating membership in each of these groups was related to pay as an outcome. The literature review cautioned that statistical significance does not establish a causal relationship, but it can provide evidence for one. Coupled with the descriptive analysis that found that Hispanic employees were significantly underpaid relative to White employees, these results indicate the presence of racial pay inequity at the study organization, for Hispanic employees and Hispanic women, whose interaction term was also statistically significant.

This finding was determined using hypothesis testing, which only multiple linear regression could provide; ridge regression and LASSO regression plots can at best only confirm the importance of specific independent variables in the models. Ridge regression is still best applied in the presence of multicollinearity, which was not found in the data set used for the empirical study. While LASSO can support the refinement of model specification by identifying variables that do not contribute to explaining variance, no variables were dropped in the model, despite years of service in the position and whether the employee held a graduate degree not being statistically significant. LASSO therefore can support model specification, but it cannot replace the value of hypothesis testing offered by multiple linear regression.

However, if the goal of the pay equity study is prediction rather than inference, LASSO does offer a slight advantage based on the empirical study, which found it yielded a lower standard error of the estimate. Prediction is often a focus of a pay equity study when the goal is to identify specific individuals who may be over- or underpaid relative to employees in a majority class. As part of a proactive pay equity study, this technique may help employers address pay inequities before they lead to a legal suit. In a reactive pay equity study, it may also help individual employees support a claim of discrimination by showing their pay was substantially different from what was predicted by the model.

In the empirical study, LASSO produced the most accurate model, with only minor losses in the overall fit of the model and the MSE. This result may be due in part to the ability of LASSO to address the problem of overfitting, where the model too closely accounts for noise or random variance in the data. These findings suggest that while multiple linear regression may be the most useful modeling technique for a pay equity study *overall*, pairing it with LASSO may support deeper analysis of the data. The current approach to pay equity studies is generally to leverage one model type alone; instead, two or more techniques could be employed in partnership to take advantage of their strengths.

The utility of each of the three models depends on the methodological focus of a pay equity study: inference or prediction. While each of the models has its own advantages and disadvantages, both technical and conceptual, they can only be properly understood and evaluated in the context of these two areas of focus.

In the case of inference, the goal of the study is to identify statistically significant factors related to predicting salary as an outcome. In a *proactive* pay equity study, an organization may seek to determine whether the factors recognized in its stated compensation policies are in fact related to assigned pay, such as years of service or level of education. A proactive study can also be used to determine whether membership in a protected class is statistically significantly related to salary, which legally it should not be. In a *reactive* pay equity study, a plaintiff may seek to establish disparate impact, while an offending organization may seek to defend against claims of it. In all cases where inference is the goal of the study, hypothesis testing is required to test for statistical significance. There is currently no accepted methodology for conducting hypothesis testing for ridge regression and LASSO; multiple linear regression is therefore the only one of the three models suitable for pay equity studies with a focus on inference.

However, in the case of prediction, the question of utility is more nuanced. With prediction, the goal of the pay equity study is to accurately predict pay for individual employees, regardless of the coefficients assigned to each variable or whether they are statistically significant or not. In such an approach, the goal is not to develop a generalizable model for the future but to develop an accurate one for the current set of data. This approach is most often used as a follow-up step to inference testing to identify employees in protected classes who may be underpaid relative to their peers. These cases can then be investigated more thoroughly to determine if there is additional relevant information related to the employees' pay outside unlawful discrimination.

A primary technical concern noted in the literature in developing a model for a pay equity study is the threat of multicollinearity, which can inflate the apparent influence of strongly correlated variables. Ridge regression offers a potential remedy for this issue but at the cost of inference testing. The empirical study found a strong statistically significant correlation between total years of service and years in the position as well as one between age and total years of service, each of them logical and reasonable from an employment perspective. In a data set with correlations that rise to the level of multicollinearity, ridge regression would yield the highest utility.

A primary conceptual concern noted in the literature is model specification – selecting the correct variables for the study. While this problem is commonly associated with the goal of inference, it also poses a challenge for prediction: the predictive accuracy of the model depends largely on the variables selected for the study. LASSO offers a potential remedy for this issue by allowing coefficients to reach zero, effectively removing them from the regression to produce the most parsimonious model that minimizes the MSE. Additionally, LASSO offers a similar remedy to multicollinearity as ridge regression, which retains all variables originally specified. In a data set where model specification is in question, LASSO would yield the highest utility.

Pay Inequity in Higher Education

Chapter 1 presented several historical reasons for pay inequity in the United States rooted in system and institutional racism. Heckman & Verkerke (1990) found that federal legal protections such as the Equal Pay Act and Title VII of the Civil Rights Act of 1964 reduced pay inequity in the country between 1965 and 1975, but their effects have waned since then. It is therefore important for practitioners in the pay equity field to consider the lasting historical effects of racism in the United States and how they might manifest in policies and practices in organizations.

The American Association for University Professors (AAUP) has conducted periodic studies of pay equity among faculty – albeit with a focus on gender equity – but there exist no studies in the literature examining pay equity for non-instructional staff at colleges and universities. This gap in the literature may be because staff in higher education are functionally comparable to employees in other industries, although there have also been few studies published detailing pay equity studies in the field outside of healthcare and athletics. It should be noted that while higher education employs staff to complete many of the same functional responsibilities as other industries, such as budget development, financial processing, and information technology, the purpose of higher education is to serve the public good rather than return profit to shareholders. The nature of employment in higher education therefore has a different set of priorities and expected outcomes than in other industries.

Still, higher education is not immune from the historical effects of racism in the United States and may experience racial pay inequity as other industries do. Because Hispanic and Black employees were often relegated to menial, low-paying positions until the mid-twentieth century, generations of employees have been denied access to education and employment advancement, setting them back further than their White peers. Further, only ten states currently require employers to disclose pay ranges to job applicants, further complicating the problem by denying Hispanic and Black employees an understanding of their pay relative to a median or expected range. Discrimination in the hiring and promotion process also often still results in People of Color being paid less and promoted less frequently than White employees (Heckman & Verkerke, 1990).

Historically Black Colleges and Universities (HBCUs) and other minority-serving institutions experience a unique form of racism rooted in government policy. During and after Reconstruction, small public schools emerged to provide basic education and trade skills to formerly enslaved Black Americans. Over time, these schools evolved to offer higher education opportunities as well, since public and private colleges and universities remained closed to People of Color, particularly Black students (Wade, 2021). These colleges and universities did not receive substantial funding from federal and state governments until the establishment of the land-grant programs of the late nineteenth century, resulting in smaller resources to recruit, retain, and pay staff compared to predominantly White institutions (Barr & McClellan, 2011). While predominantly White institutions today can leverage returns from endowments seeded two centuries ago or longer, minority-serving institutions find themselves significantly underfunded and struggling for survival, resulting in lower pay for People of Color compared to White employees across the field (Barr & McClellan, 2011; Wade, 2021).

The present study examined staff data at a four-year university in the Western United States and found pay inequity was present for Hispanic and Black employees. More specifically, Hispanic employees and Hispanic women employees were found to earn significantly less than White employees. While the literature does not provide other studies for comparison, it does offer potential explanations for why this inequity may exist today at the university. Other institutions should carefully consider these factors when conducting their pay equity studies to ensure they are not overlooked or reproduced. Whether the pay equity study is conducted by a single practitioner or a team of them, inclusive leadership, critical race consciousness, and a respect for the unique challenges and obstacles faced by people of color can help ensure the study delivers meaningful results to inform decision-making and address racial pay inequity.

Significance of the Study

The findings from the study described in this dissertation make several important contributions to the pay equity study literature.

First, the literature review synthesized support for and criticisms of the use of multiple linear regression in the literature spanning over thirty years and multiple disciplines, which to date had not been done. This review therefore serves as a meaningful resource for practitioners seeking a deeper understanding of the methodological, technical, and conceptual concerns associated with the use of multiple linear regression in pay equity studies. The goal of this study was to advance racial equity by empowering practitioners to make better, more thoughtful decisions in the research process associated with pay equity studies.

Second, the empirical study explored the use of regularization techniques in comparison to multiple linear regression in a pay equity study of real-world employment data. To date, only a few alternatives to multiple linear regression have been proposed in the literature, and only one method – Peters-Belson regression – has been implemented in pay equity studies and published. While the results of the study supported the strengths of multiple linear regression, developing ridge regression and LASSO models highlighted several salient issues associated with the modeling process, such as how to code contrast and interaction terms in regularization models.

Third, in the empirical study, multiple statistical methods were applied and compared regarding their utilization in the field of pay equity studies. Reflections were explored and recommendations were proposed based on the feasibility and utility of these regression models to enhance researchers' understanding of these methods. Regardless of the focus of the pay equity study – inference or prediction – multiple linear regression was found to be the most feasible of the three techniques explored as it is the simplest to conduct, it needs the least amount of technical expertise, and it does not require standardization of variables. Multiple linear regression therefore poses the fewest technical challenges and barriers to practitioners who choose to implement it in their studies when compared to ridge regression and LASSO regression. The utility of each of the three models was found to be more nuanced. In the presence of multicollinearity, ridge regression offers a potential remedy, at the cost of losing the ability to conduct significance testing. If the goal of the pay equity study is prediction rather than inference, LASSO regression offers a slight advantage over multiple linear regression and ridge regression in that it can produce the most accurate model, and it supports model specification by allowing variables to drop out of the model.

For practitioners, like institutional research professionals, human resources staff, and outside consultants, this study provides a roadmap for how a pay equity study can be conducted using regularization as a technique to replace or supplement multiple linear regression alone. For institutional leaders, who are often charged with leading a pay equity study, this research provides a discussion of how to consider the principles of QuantCRiT and race consciousness in the process.

Limitations

The literature offered few studies in which the technical details of regression techniques used in a pay equity study were explained. Instead, these studies focused on the outcome of whether discrimination was found or not, particularly in the legal literature. Likewise, publicly available non-academic studies reporting the results of university pay equity studies failed to provide insight into the mechanics of how multiple linear regression was employed. These reports and studies did not indicate how certain contextual realities were addressed, such as when a particular employee is taking on the responsibilities of multiple staff. While it is possible realities like this are built into job value, it is neither certain nor clear. How these variables are modeled can further mask pay inequities.

The empirical study adopted a set of variables *a priori* in accordance with guidelines from the EEOC, the university's compensation philosophy, and findings from the literature review. Had different variables been selected, the results of the empirical study may have been different, particularly in the presence of notable multicollinearity.

The literature currently offers conflicting opinions on whether to standardize categorical variables in ridge regression and LASSO regression that have been coded using dummy coding: some authors argue that standardization is unnecessary given the already-present binary values, while others argue that not standardizing gives undue weight to the continuous variables in the regression model (Hardy, 1993; Joffe & Greenland, 1995). The present study did not standardize binary variables – legal sex and whether the employee held a graduate degree – as their values would be problematic for interpretation.

No guidance was found in the literature on whether to standardize contrast variables in ridge regression and LASSO regression. Because the purpose of the contrast coding for these variables was to resolve the issue of small and unbalanced group sizes and to allow for inferences about the differences in means between two race/ethnicity groups in the multiple linear regression model, these variables were not standardized in the analysis. The literature also offered no guidance on whether to standardize interaction terms that were the product of two dummy coded variables or one or more contrast variables with ridge regression or LASSO. The empirical study did not standardize the interaction terms to maintain consistency with the underlying terms used to calculate them.

Recommendations for Future Research

The process of conducting a pay equity study presumes equity in pay outcomes despite the presence of inflation, merit increases, cost of living adjustments, and individualized raises over time. Currently, no modeling technique can handle all these sources of variance, including multiple linear regression. One approach to minimize this complexity would be to analyze employee qualifications and position attributes *at the point of entry* into the position. This type of analysis could be more useful for a pay equity study because it distinguishes between potential discrimination in the pay set when the position was started from discrimination that occurred afterwards.

However, this approach would have several substantive drawbacks. If performance ratings were to be retained in the model, new employees would have to be removed, as they would not have a rating for the position yet. As a result, this constraint could inadvertently remove new employees who entered as part of new hiring initiatives to increase the diversity of employees, which would in turn increase group sizes in certain race/ethnicity groups. A pay equity study would then have to be conducted one or more years after the hiring change to appropriately analyze the data. The difference in points of entry would also require consideration of whether to use nominal values for pay or adjust for inflation over time.

Future studies could also compare the utility of multiple linear regression, ridge regression, and LASSO in the presence of multicollinearity, which regularization effectively addresses. Given the descriptive analysis of the present study, it may be challenging to find microdata that meet this requirement. One approach to such research could therefore be to use a simulated employment data set with multicollinearity and racial discrimination present to fully examine the performance of ridge regression and LASSO models. Constructing a simulated data set with these parameters would be difficult but not impossible given advances in packages in R and Python.

A qualitative study to further examine a college or university's experience with a pay equity study might also be a fruitful addition to the pay equity study literature. Future studies could use a case study approach paired with document and archival analysis to examine their processes, with a focus on how well equity-mindedness and race consciousness were reflected in them. A mixed-methods approach might also help researchers better understand the experience of conducting a pay equity study from the perspective of practitioners and stakeholders. An iterative process could be implemented to examine pay equity studies holistically and provide context for statistical findings returned from a quantitative portion of the study.

Researcher's Reflections

During my career, I have participated in two staff pay equity studies in academic settings, each troubling and problematic. A lack of inclusive leadership marred both studies from the beginning, and poor research design, analysis, and interpretation of findings exacerbated the problem. At the time, I lacked the positional authority to influence the direction of the projects, so I turned to scholarship to give voice to my concerns. However, conducting this research posed many challenges from a personal, technical, and conceptual perspective. In the spirit of self-reflexivity, I offer these thoughts and considerations to future researchers who may struggle with the same issues I encountered in this research.

As I began the research design process, I frequently encountered technical limitations that forced me to make decisions that reproduced the very inequity I sought to address. The most immediate limitation was the inability of regression to accommodate small and unbalanced group sizes. The implication for members of already minoritized communities was immediately obvious: race/ethnicity could not be fully reflected in the model if I were to adhere to the guidelines for the appropriate use of regression. Interaction terms to represent intersectionality would likewise be problematic. Removing certain groups from the model was difficult, as it forced me to make invisible certain communities whose identities may have had a meaningful relationship to their experiences as employees.

As a compromise, I employed contrast coding for the two largest groups: Hispanic and Black employees. However, to implement this approach, I also needed to establish a baseline group for comparison; logically, this group was White employees as the majority class. Doing so centered whiteness in my study though, contributing to a further sense of “othering” People of Color by defining them in contrast to White employees. I continue to struggle with how to decenter whiteness when the study of progress and equity requires us to make such comparisons. Historically, pay equity studies have focused on comparing men and women, which also requires defining one group in comparison to another. We are fortunate today to understand that identity is far more complex than simple binaries, but that framework of identity will likely remain with us as long as inequity does.

A closely related issue I encountered was the use of federal census categories for this study. Having worked for over a decade in institutional research, I was acutely aware at the start of this study of the limitations of these categories, particularly for individuals who do not see themselves accurately reflected in them. As noted in Chapter 3, the Hispanic or Latino category is not only problematic because it obscures many unique cultural identities, but it also supersedes all other racial affiliations an individual may have. I have struggled throughout my life with the issue of whether being Hispanic is a White ethnicity or if it is a unique racial identity due to historical oppression and colonization associated with it. Similar issues can be seen in the American Indian, Asian, and Black categories as well, where many meaningful subgroups exist. Perhaps most egregious is the Multiracial category, which conflates a wide variety of bi- and multi-racial identities into one shared "experience."

I explored whether I could code the different categories as dummy variables with overlapping categories to at least partially represent racially intersectional identities, but once again I faced the issue of small and unbalanced group sizes. To align with generally accepted pay equity practices and provide a framework for future studies, I chose to use the federal census categories for this study. Doing so did not feel right, but it was the only avenue available within the constraints of regression.

Garcia and Mayorga (2018) noted that "a dataset is constructed with a particular theoretical framework that impacts the entire research design and process." As I began exploratory data analysis, another issue I encountered was the quality and credibility of the data I was given for this study and whether I as a researcher was accepting the framework in which they were created. The file I received contained anonymized records taken from a larger census reporting data set. While I was told these data had been thoroughly vetted before release, as they are used for federal compliance reporting, I did not have direct knowledge of how individual data points were collected or maintained over time. I also had to trust that the data themselves were accurate, despite unusual observations in the records. For example, the standard deviations for salary were large for certain race/ethnicity groups, despite my removing outliers from the data set. Without more information about the nature of these data, I could only accept them as they were and move forward with the records as the university would have done had it conducted the study itself. However, it is important to note that data management can significantly impact the findings from a pay equity study, so careful attention should be paid to it where possible.

The design and structure of the variables can either limit analysis or create opportunities for it. For example, the variable in my models that reflected whether the employee was a supervisor was given as a binary variable in part to anonymize the data. Had the number of employees supervised been provided instead, the results of the analysis might have been different given the addition of another continuous variable. This variable could also have been transformed into a categorical one with the number of employees supervised dummy coded to capture whether it was low, medium, or high. This change could have likewise changed the results of the study. This study also examined pay as an outcome for a group of individuals employed at a particular point in time. Had the data been collected and analyzed longitudinally, the models would have had to account for multiple merit raises, cost of living adjustments, market adjustments over time. These changes are difficult to model in combination with each other, but they could also provide a more complete picture of the lifecycle of contributing factors in predicting pay.

Similarly, certain variables were not available for this study, such as total years of experience; the data set provided only included total years of service at the university and in the current position. An employee with fifteen years of experience but who started at the university only a year before the study could therefore appear less “valuable” from an organizational perspective when looking at years of service. Likewise, credentials and certifications beyond holding a graduate degree were not available for employees at the university. These data might have been an additional source of variance in the models to explore and should be considered when available.

A final issue I encountered was how to address the heteroskedasticity observed in the model. Normal distribution of error terms is required to appropriately interpret the results of a regression analysis, so the violation of this assumption posed a technical challenge. I opted to perform a log transformation on salary as the outcome variable, which resolved the heteroskedasticity but resulted in the loss of simple interpretation of salary in its natural units. Alternative options available include reciprocal, square root, cube root, and square transformations. Researchers should likewise be prepared to explore an appropriate remedy to this problem in future empirical studies.

Although I believe in the value of my research, it does not offer solutions to the issues I have described. Instead, I acknowledge the limits of my knowledge, and I continually seek insight from other individuals who are also passionate about advancing the work of equity-focused study. It is my hope that this research will provide a starting point for richer conversations about the implications of choices made in pay equity studies.

REFERENCES

Altonji, J., Smith Jr, A., & Vidangos, I. (2013). Modeling earnings dynamics. *Econometrica*, *81*(4), 1395-1454.

Ananda, S., & Gilmartin, K. (1991). Inclusion of potentially tainted variables in regression analyses for employment discrimination cases. *Indus. Rel. LJ*, *13*, 121.

Anderson, J. (1993). Race, meritocracy, and the American academy during the immediate post—World War II era. *History of Education Quarterly*, *33*(2), 151-175.

Arashi, M., Saleh, A., & Kibria, B. (2019). *Theory of ridge regression estimation with applications*. John Wiley & Sons.

Barbezat, D. (2002). History of Pay Equity Studies. *New Directions for Institutional Research*, *115*, 9-39.

Barnett, A. (1982). An Underestimated Threat to Multiple Linear Regression Analyses Used in Job Discrimination Cases. *Indus. Rel. LJ*, *5*, 156.

Barr, M. & McClellan, G. (2011). *Budgets and Financial Management in Higher Education.* San Francisco, CA: Jossey-Bass.

Barrett, G., Alexander, R., Anesgart, M., & Doverspike, D. (1986). Frequently encountered problems in the application of regression analysis to the investigation of sex discrimination in salaries. *Public Personnel Management*, *15*(2), 143-157.

Barrett, G., & Doverspike, D. (1989). Fundamental errors and erroneous beliefs in using regression analysis as a test for sex discrimination. *Law & Psychol. Rev.*, *13*, 1.

Beeby, J., & Nieman, D. (2002). The Rise of Jim Crow, 1880–1920. *A Companion to the American South*, 336-347.

Best, R., Edelman, L., Krieger, L., & Eliason, S. (2011). Multiple disadvantages: An empirical test of intersectionality theory in EEO litigation. *Law & Society Review*, *45*(4), 991-1025.

Beyer, R. (2019). *The Positive Effects of Equitable Pay on Outcomes that Employers Value* (Doctoral dissertation, Roosevelt University).

Bielby, W., & Coukos, P. (2007). Statistical Dueling with Unconventional Weapons: What Courts Should Know about Experts in Employment Discrimination Class Actions. *Emory LJ*, *56*, 1563.

Billard, L. (2017). Study of salary differentials by gender and discipline. *Statistics and Public Policy*, *4*(1), 1-14.

Bobko, P. (2001). *Correlation and regression: Applications for industrial organizational psychology and management*. Sage Publications.

Bura, E., Gastwirth, J., & Hikawa, H. (2010). Local linear logistic Peters–Belson regression and its application in employment discrimination cases. *Statistics and its Interface*, *3*(2), 125-144.

Carter, R. (2015). *Perceptions of strategic compensation* (Doctoral dissertation, Lipscomb University).

Chan, J., Leow, S., Bea, K., Cheng, W., Phoong, S., Hong, Z., & Chen, Y. (2022). Mitigating the Multicollinearity Problem and Its Machine Learning Approach: A Review. *Mathematics*, *10*(8), 1283.

Chatterjee, S., & Hadi, A. (1986). Influential Observations, High Leverage Points, and Outliers in Linear Regression. *Statistical Science*, *1*(3), 379–393.

Connell, E., & Mantoan, K. (2017). Mind the Gap. *ABA Journal of Labor & Employment Law*, *33*(1), 1-30.

Connolly, C. (1991). The use of multiple linear regression analysis in employment discrimination cases. *Population Research and Policy Review*, *10*(2), 117-135.

Davis, M. (2010). Contrast coding in multiple linear regression analysis: Strengths, weaknesses, and utility of popular coding structures. *Journal of data science*, *8*(1), 61-73.

Dempster, A. (1988). Employment discrimination and statistical science. *Statistical Science*, 149-161.

Eichelberger, K. Y. (2018). Equal pay for equal work in academic obstetrics and gynecology. *Obstetrics & Gynecology*, *131*(2), 224-226.

Finkelstein, M. (1980). The judicial reception of multiple linear regression studies in race and sex discrimination cases. *Columbia Law Review*, *80*(4), 737-754.

Fogel, W. (1986). Class pay discrimination and multiple linear regression proofs. *Neb. L. Rev.*, *65*, 289.

Follett, R., & Welch, F. (1983). Testing for discrimination in employment practices. *Law & Contemp. Probs.*, *46*, 171.

Fox, J. (1997). *Applied regression analysis, linear models, and related methods*. Sage Publications, Inc.

Friedman J., Hastie T., Tibshirani R. (2010). “Regularization Paths for Generalized Linear Models via Coordinate Descent.” *Journal of Statistical Software*, *33*(1), 1–22. https://www.jstatsoft.org/v33/i01.

Fusfield, D., & Bates, T. (2005). The Black sharecropping system and its decline. *African Americans in the US economy*, 32-37.

Garcia, N., & Mayorga, O. (2018). The threat of unexamined secondary data: A critical race transformative convergent mixed methods. *Race Ethnicity and Education*, *21*(2), 231-252.

Gareth, J., Daniela, W., Trevor, H., & Robert, T. (2013). *An introduction to statistical learning: with applications in R*. Springer.

Gastwirth, J. (1988). [Employment Discrimination and Statistical Science]: Comment. *Statistical Science*, *3*(2), 175–183. http://www.jstor.org/stable/2245562

Gibbons, D. (1981). A simulation study of some ridge estimators. *Journal of the American Statistical Association*, *76*(373), 131-139.

Guerin-Gonzales, C. (1994). Conversing across boundaries of race, ethnicity, class, gender, and region: Latino and Latina labor history. *Labor History*, *35*(4), 547-563.

Guo, Shenyang, and Mark W. Fraser. *Propensity score analysis: Statistical methods and applications*. Vol. 11. SAGE publications, 2014.

Guilkey, D. K., & Murphy, J. L. (1975). Directed ridge regression techniques in cases of multicollinearity. *Journal of the American Statistical Association*, *70*(352), 769-775.

Gillborn, D., Warmington, P., & Demack, S. (2018). QuantCrit: Education, policy, 'Big Data' and principles for a critical race theory of statistics. *Race ethnicity and education*, *21*(2), 158-179.

Godfrey, L., Orme, C., & Santos Silva, J. (2006). Simulation-based tests for heteroskedasticity in linear regression models: Some further results. *The Econometrics Journal*, *9*(1), 76-97.

Hadi, A., & Chatterjee, S. (2009). *Sensitivity analysis in linear regression*. John Wiley & Sons.

Hardy, M. (1993). *Regression with dummy variables* (Vol. 93). Sage.

Heikkila, E. (1988). Multicollinearity in regression models with multiple distance measures. *Journal of Regional Science*, *28*(3), 345-362.

Heckman, J., & Verkerke, J. (1990). Racial Disparity and Employment Discrimination Law: An Economic Perspective. *Yale L. & Pol'y Rev.*, *8*, 276.

Hikawa, H., Bura, E. and Gastwirth, J. (2010). Robust Peters-Belson Type Estimators of Measures of Disparity and their Applications in Employment Discrimination Cases. Technical Report, Department of Statistics, George Washington University.

Jagsi, R., Biga, C., Poppas, A., Rodgers, G., Walsh, M. N., White, P. J., & Douglas, P. (2016). Work activities and compensation of male and female cardiologists. *Journal of the American College of Cardiology*, *67*(5), 529-541.

Jagsi, R., Griffith, M., Stewart, A., Sambuco, M., DeCastro, M., & Ubel, P. (2013). Gender differences in salary in a recent cohort of early-career physician-researchers. *Academic medicine: journal of the Association of American Medical Colleges*, *88*(11).

Joffe, M., & Greenland, S. (1995). Standardized estimates from categorical regression models. *Statistics in Medicine*, *14*(19), 2131-2141.

Kanazawa, S. (2005). The myth of racial discrimination in pay in the United States. *Managerial and Decision Economics*, *26*(5), 285-294.

Kraha, A., Turner, H., Nimon, K., Zientek, L., & Henson, R. (2012). Tools to support interpreting multiple linear regression in the face of multicollinearity. *Frontiers in Psychology*, *3*, 44.

Kennedy, D. M. (2009). What the New Deal Did. *Political Science Quarterly*, *124*(2), 251-268.

Lawless, J. (1981). Mean Squared Error Properties of Generalized Ridge Estimators. *Journal of the American Statistical Association*, *76*(374), 462–466.

Marquardt, D., & Snee, R. (1975). Ridge Regression in Practice. *The American Statistician*, *29*(1), 3–20.

McDermott, D. (2000). Whither the glass ceiling? Conducting a statistical pay audit. *Compensation & Benefits Review*, *32*(5), 17-22.

Muller, C. (2018). Freedom and convict leasing in the postbellum South. *American Journal of Sociology*, *124*(2), 367-405.

Neter, J., Kutner, M., Nachtsheim, C., & Wasserman, W. (1996). Applied linear statistical models.

O'Brien, R. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & quantity*, *41*(5), 673-690.

Omi, M., & Winant, H. (2014). *Racial formation in the United States*. Routledge.

Peterson, D. (1999). Pay discrimination models. *Journal of Forensic Economics*, *12*(2), 111-124.

Powell Jr, L. (1973). McDonnell Douglas Corp. v. Green. *Opinion of the United States Supreme Court*, *6*.

Price, B. (1977). Ridge regression: Application to nonexperimental data. *Psychological Bulletin*, *84*(4), 759.

Price, R., & Mills, E. (1985). Race and residence in earnings determination. *Journal of Urban Economics*, *17*(1), 1-18.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org.

Roth, P., Bobko, P., & Switzer III, F. (2006). Modeling the behavior of the 4/5ths rule for determining adverse impact: Reasons for caution. *Journal of Applied Psychology*, *91*(3), 507.

Rouen, E. (2017). *Rethinking Measurement of Pay Disparity and its Relation to Firm Performance*. (Doctoral dissertation, Columbia University).

Rumrill, P., Fitzgerald, S., & Merchant, W. (2010). Using scoping literature reviews as a means of understanding and interpreting existing literature. *Work (Reading Mass)*, *35*(3), 399‑404.

Sarkar, D., & Haverstick, A. (2010). The Preventive Pay Equity Audit: Guidance for Modeling the Regression Analysis. *ABAJ Lab. & Emp. L.*, *26*, 61.

Shlomowitz, R. (1979). The origins of Southern sharecropping. *Agricultural History*, *53*(3), 557-575.

Simson, D. (2018). Fool Me Once, Shame on You; Fool Me Twice, Shame on You Again: How Disparate Treatment Doctrine Perpetuates Racial Hierarchy. *Hous. L. Rev.*, *56*, 1033.

Sklaroff, L. R. (2009). *Black culture and the New Deal: the quest for civil rights in the Roosevelt era*. Univ of North Carolina Press.

Starkweather, J. (2010). Categorical variables in regression: Implementation and interpretation. *Benchmarks Online*.

Starnes, D., Yates, D., & Moore, D. (2010). *The practice of statistics*. Macmillan.

Tabron, L. & Thomas, A. (2023). Deeper than wordplay: A systematic review of critical quantitative approaches in education research (2007–2021). *Review of Educational Research*, 00346543221130017.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*(1), 267–288.

Tolnay, S. (2003). The African American "great migration" and beyond. *Annual Review of Sociology*, *29*(1), 209-232.

University of California, Irvine. (2021). *Fifty-state pay equity law summary*. The Equity Project. https://www.law.uci.edu/centers/pay-equity-project/images/50-state-law-chart.pdf.

Vasu, E. S. (1979). *Non-normality in regression analysis: A Monte Carlo investigation under the condition of multi-collinearity*. Institute for Research in Social Science, University of North Carolina at Chapel Hill.

Wade, E. (2021). The History of HBCUs: Lessons on Innovation from the Past. In *Reimagining Historically Black Colleges and Universities: Survival Beyond 2021* (pp. 5-13). Emerald Publishing Limited.

Weisberg, H., & Tomberlin, T. (1983). Statistical evidence in employment discrimination cases. *Sociological Methods & Research*, *11*(4), 381-406.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag: New York. https://ggplot2.tidyverse.org.

Williams, J. (1981). BKWPAC: Single Observation Regression Diagnostics with Plots and Reweighted Least Squares. *The American Statistician*, *35*(3), 163–163.

Zuberi, T. (2001). *Thicker than blood: How racial statistics lie*. U of Minnesota Press.