Rationalizing Equal Pay:
Monotonicity and the Limits of Aggregation*

Bo Cowgill† Bobak Pakzad-Hurson‡

March 27, 2026

Working Paper

Abstract

When are the relative wages a firm pays to its workers "fair"? Theoretically, we characterize "Equal Pay for Equal Work" laws in the US as restricting wages to be monotonic in productivity at the individual level. Empirically, we study the pay and productivity of researchers at top public universities, where salaries and research output are largely observable. Is there a weighting of output metrics (e.g. publications in one journal versus another) that rationalizes the wages within a firm as monotonic? We show that the answer is no in 61% of cases, and that inter-university moves by researchers contribute to violations of monotonicity. While individual-based Equal Pay for Equal Work is restrictive, we show that US-style laws nearly satisfy the desirable, symmetric fairness goal of "Equal *Work* for Equal *Pay*." By contrast, Equal Pay for Equal Work approaches in other countries which focus on group-level averages in pay controlling for output are less likely to symmetrically treat "pay" and "work."

JEL Classification: J31, J71, M52, K31, D63.

Keywords: Equal Pay for Equal Work, Wage Fairness, Pay Monotonicity, Outcome Tests, Personnel Economics.

*We thank Arjada Bhardi, Fabrizio Dell'Acqua, Laura Doval, Huseyin Yildirim, and seminar audiences for constructive discussions and feedback. We also thank Maxwell Cozean, Noa Rosinplotz, and Ian Cheng for outstanding research assistance. Cowgill thanks the Kauffman Foundation and the Columbia Center for Political Economy.

†Columbia University. bo.cowgill@gsb.columbia.edu.

‡Brown University. bph@brown.edu.

1 Introduction

A common societal goal is to ensure relative pay among coworkers is fair. Various laws and regulations speak to this—mandated pay transparency, restrictions on the ratio of CEO-to-worker pay, and protections for remote work vs in-person employees. The most consistent of these initiatives—found in 98 countries around the world[1]—are "equal pay for equal work" (EPEW) laws that explicitly restrict the relative pay of workers doing similar tasks within a firm. In the United States, the Equal Pay Act of 1963 imposes this restriction, and all 50 states have adopted similar legislation.

EPEW requirements, particularly in the US, frequently exhibit three common features. First, they collapse a worker's "work" into a single-dimensional, omnibus measure which we refer to in this paper as *productivity*.[2] This unidimensionality allows for a complete ranking over workers' productivities. Second, they bind on an individual basis, meaning that a firm is in violation if even one worker is paid less than another and the pay difference is not justified by productivity differences.[3] These individual-basis restrictions frequently are not meaningfully

[1] https://www.worldbank.org/en/news/press-release/2024/03/04/new-data-show-massive-wider-than-expected-global-gender-gap

[2] For example, New York's, Connecticut's, and Canada's EPEW defines a worker's work as a "composite" of factors such as skill, effort and responsibility (see, respectively, NY Labor Law § 194; CT Gen Stat § 31-75; Pay Equity Act, S.C. 2018, c. 27, s. 416(47-50)). In some countries such as Germany (Entgelttransparenzgesetz [EntgTranspG] [Transparency in Wage Structures Act], June 30, 2017, BGBl. I. at 2152, §§ 3–4, last amended by Gesetz, July 5, 2021, BGBl. I at 3338, art. 25), the equal pay restriction is stated as equal pay for *work of equal value*, explicitly requiring a single value to measure potentially multi-dimensional factors of output.

[3] For example New York, Illinois, Germany, and Ontario specifically state that it is in violation for any employee to be paid less than even a single coworker when productivity differences cannot justify unequal pay (see, respectively, NY Labor Law § 194; 820 Ill. Comp. Stat. 112/10; Entgelttransparenzgesetz [EntgTranspG] [Transparency in Wage Structures Act], June 30, 2017, BGBl. I. at 2152, §§ 3–4, last amended by Gesetz, July 5, 2021, BGBl. I at 3338, art. 25; Ontario's Employment Standards Act, 2000, S.O. 2000, c. 41, s. 42(1)). More broadly, legal proceedings regularly find in favor of plaintiffs who are able to find only one coworker, or a small number of coworkers, who are paid higher wages without the requisite productivity argument—see EEOC v. AH 2007 Management, LP d/b/a Courtyard Monroe Airport, and Aimbridge Hospitality, LLC d/b/a Aimbridge Hospitality, LLC, No. 19-cv-914 (W.D. La. Feb. 3, 2021); EEOC v. Jerry's Chevrolet Inc., et al., No. 21-cv-02464 (D. Md. Nov. 30, 2022); EEOC v. Dell, Inc., No. 3:20-cv-03131 (N.D. Tex. Sept. 2, 2021); EEOC v. First Metropolitan Financial Services, Inc., 1:18-cv-177 (N.D. Miss. March 18, 2021); EEOC v. Enoch Pratt Free Library, No. 8:17-cv-02860 (D. Md. Dec. 23, 2020); EEOC v. Covenant Medical Center, Inc., 2:20-cv-10662 (E.D. Mich. Sept. 2, 2020); EEOC v. Cummins, Inc., d/b/a Cummins Business Services, No. 3:17-cv-01306 (M.D. Tenn. Mar. 29, 2019); EEOC v. Prince George's County, Maryland, No. 1:15-cv-02942 (D. Md. June 1, 2017); Samira Ahmed v BBC [2020] UKET 2206858/2018. The largest source of variation between EPEW on this dimension is whether the restriction applies to all pairs of workers, or only to workers in different protected classes (e.g. a woman cannot be paid less than a man if she is equally productive,

constrained by the notion of "protected classes."[4] Third, they are broader than merely ensuring equal pay for "equal" work, and instead imply that a worker cannot be paid less than a peer who is equally or less productive, i.e., "equal pay for equal *or better* work."[5]

Perhaps the largest qualitative difference that exists between EPEW laws is whether the restriction binds on an individual basis, as discussed, or on an aggregate, group-level basis. Specifically, group-level EPEW restrictions differ from individual-level EPEW found in the US in that they ask whether, controlling for productivity, members of one sex are paid less than the other on average. Section 50 of Canada's Pay Equity Act (S.C. 2018, c. 27, s. 416), for example, specifically defines the "equal line method" as a comparison between male and female "regression lines" for workers performing similar tasks. Moreover, if the female "regression line" falls below the male "regression line," the law prescribes raising women's wages to equalize the two; similar adjustments are often advocated by policy and consulting specialists as a way of increasing pay fairness (for details, see Anderson et al., 2019).

In this paper, we positively and normatively study pay "fairness" as restricted by EPEW. Taking a positive approach, we theoretically characterize the restrictions on the set of *worker types*—each specifying a wage and a productivity—that can simultaneously be employed by a firm under the individual- and group level-restrictions. Using a novel dataset containing details on the pay and productivity of US research faculty, we ask whether the pay structures of US universities for research faculty can be rationalized as satisfying legal restrictions imposed by EPEW. We find that in most cases, the answer is "no."

---

but she could be paid less than an equally productive woman). We return to these differences later in our analysis.

[4]While all EPEW restrictions in the US, to our knowledge, specify that the restriction of equal pay binds only for workers across defined protected classes, there are often many specified protected classes. Because plaintiffs alleging they were unfairly paid can present a comparison worker who differs along any protected class, the number of "groups" is combinatorially defined by the intersection of any subset of protected classes. For example, New York specifies 11 protected classes (NY Labor Law § 194), and even if these are each binary categories, there are 2,048 groups across which EPEW binds. Given typical firm sizes, it is plausible that many workers share no coworkers within their group, which collapses EPEW restrictions fully to the individual level.

[5]See Morgan v. United States Soccer Fedn., 25 F.4th 1102 (9th 2022); Bence v. Detroit Health Corp., 712 F.2d 1024, 1027 (6th Cir. 1983); EEOC v. Chas. S. Winner, Inc. d/b/a Winner Ford of Cherry Hill d/b/a Winner Ford, No. 1:16-06137 (D.N.J. Nov. 22, 2017); EEOC v. Mechanical Design Systems, Inc., No. 8:22-cv-02463 (D. Md. May 9, 2023); EEOC v. AH 2007 Management, LP d/b/a Courtyard Monroe Airport, and Aimbridge Hospitality, LLC d/b/a Aimbridge Hospitality, LLC, No. 19-cv-914 (W.D. La. Feb. 3, 2021); EEOC v. Enoch Pratt Free Library, No. 8:17-cv-02860 (D. Md. Dec. 23, 2020).

Taking a normative approach, we study whether individual- or group-level EPEW restrictions are desirable. As we explain shortly, we consider a relative pay restriction to be "desirable" if it treats wage and productivity symmetrically. We show that individual-level EPEW is not symmetric, but it is nearly so, and we characterize the minimum additional restriction to add to EPEW to ensure symmetry: a condition we refer to as "Equal Work for Equal Pay" (EWEP). By contrast, we show that group-level EPEW restrictions are far from symmetric, and indeed, that "flipping" the role of pay and productivity can lead to contradictory assessments of the fairness of a firm's pay structure. These results suggest that individual-level EPEW restrictions may be seen as more fair than group-level EPEW.

Our analysis begins by considering a firm and a finite set of workers it employs, each with a wage and a productivity. (Individual-level) EPEW imposes a version of Becker's "outcome test" for discriminatory treatment against a worker: given the worker's input (productivity), are there weakly lower-input workers who receive a weakly higher output (wage)? We theoretically show that a *firm structure*—a joint collection of pay and productivity among the workers—satisfies EPEW if and only if the wage schedule follows a non-decreasing function of productivity, a restriction which we refer to as *pay monotonicity*. Pay monotonicity is demanding, and becomes more demanding as the number of employed workers grows. In large firms, pay monotonicity affords almost no flexibility on the pay structure: if the firm employs a set of workers across the entire productivity distribution, then the firm has almost no flexibility when it comes to setting the wage of a new worker it hires. For similar reasons, the firm also lacks the ability to match competing offers or to poach employees from other firms at a wage premium, and it removes the ability of a worker to individually negotiate higher pay.

We then empirically test compliance with legal restrictions, i.e., pay monotonicity. An initial difficulty we must contend with are the measurement challenges associated with pay monotonicity. If each worker had only a single dimension of productivity (e.g. number of widgets constructed per hour), it would be trivial to check whether or not high performers earn more than low performers. However, many jobs contain multiple dimensions of productivity

(e.g. representing different tasks) that must be aggregated to determine aggregate productivity. An observer likely does not know the relative weighting of these dimensions, meaning that a given pay structure may or may not satisfy wage monotonicity given different potential weights.

We develop an empirical framework to overcome this measurement challenge. Our framework helps measure pay monotonicity in real-world settings with an arbitrary number of productivity metrics. It poses the following question: Does a linear weighting of the productivity metrics exist that could rationalize observed pay as monotonic? Or, is the set of weights that rationalize observed pay as monotonic empty? Our technique allows for interactions and arbitrary transformations of productivity metrics, however, as the set of inputs grows, it becomes easier to rationalize pay as monotonic. As such, the effectiveness of our methods relies on the ability to limit the complexity of pay schemes, and to ensure they remain interpretable and practical. We also develop a related new technique to address the role of unobserved productivity, and to quantify the "distance" from monotonicity (in units of pay).

As a proof of concept for our rationalizability approach, we apply it to professors in academic departments at top public universities (2008-2019), where pay information is publicly available. We use publications in different journals and citation counts as productivity metrics, allowing for the idea that publications in certain journals might be weighted more highly than others. Throughout, we implement our method in the way most generous to employers and look only at within-job-title comparisons, i.e. we enact a very conservative approach to declaring a firm is in violation of monotonicity. Despite this, we find that 61% of department$\times$rank$\times$year cells fail the monotonicity test: The set of journal weights rationalizing the pay distribution as monotonic is empty. To rationalize it with unobserved factors (such as teaching or service), the strongest unobserved contribution would have to count for about 30% as much as the most heavily weighted observable metric in the cell (on average).

To study whether larger units are less likely to satisfy pay monotonicity, we use a shift–share design based on school-level cumulative external hiring interacted with baseline title shares. We interpret this exercise as suggestive evidence on the scale–fairness link rather than as a

clean external-shock design. In the IV estimates, each additional faculty member in a cell is associated with about a 3–4 percentage-point lower probability of a monotonic pay schedule and about $4.4–$4.9 thousand more salary gap that must be ignored to restore monotonicity. These patterns are consistent with the theoretical prediction that, as a firm expands, its pay schedule becomes more brittle: larger firms have fewer admissible "slots" for new hires without violating monotonicity, making it harder to accommodate poaching or outside offers without breaching the equal-pay restriction. Consistent with this mechanism, full-professor cells — which are on average twice as large as associate-professor cells — exhibit larger monotonicity violations, suggesting that bigger groups are more prone to these breaches.

Returning to the conceptual question of relative pay fairness, an important observation is that EPEW treats pay and productivity asymmetrically. A firm structure satisfies EPEW if all workers are paid the same wage, even if they have different productivities, but it does not satisfy EPEW if all workers have the same productivity but are paid different wages. The symmetric version of EPEW, which we call "equal work for equal pay" (EWEP), requires that if one worker is paid the same (or more) than another, then the former should not be less productive than the latter.

We propose that an ideal fairness restriction on relative wages should treat pay and productivity symmetrically for two reasons. First, symmetry of pay and productivity is popularly considered to be an important fairness notion. Complaints about EWEP violations often take the form: "A worker from a disadvantaged group must be twice as productive to receive the same pay as a worker of the majority group."[6] A recent survey (*The Pulse of Black America*, 2022) finds that violations of EPEW and EWEP are the top two labor-related concerns among employed black Americans (EWEP 57%, EPEW 56%).

Second, symmetry avoids a "chicken or the egg" phenomenon: do wages determine productivity, or does productivity determine wages? We note that economists have no clear answer to this question. Papers in the organizational economics literature (e.g.s Winter, 2004; Halac

[6]This idea has been stated in various ways. In her book, *Black Feminist Thought*, Patricia Hill Collins critiques "the basic economic exploitation where U.S. Black women [...] work twice as hard for the same pay[.]" (Collins, 1990, page 74). In her 2021 memoir *Becoming*, Michelle Obama wrote, "There's an age-old maxim in the black community: You've got to be twice as good to get half as far."

et al., 2021; Yildirim, 2025) consider the design of incentive schemes under the premise that workers pick their level of productivity in response to wage contracts, while papers in the labor economics literature (e.g.s Mortensen and Pissarides, 1994; Postel-Vinay and Robin, 2002; Cullen and Pakzad-Hurson, 2023; Cullen et al., 2025) take worker productivity as fixed and study the wage determination process.[7] We are agnostic about the direction of causation, but given that either is plausibly true, and that the answer may differ by labor market, symmetric treatment of pay and productivity is potentially important. In other words, because legal restrictions on pay fairness do not differ by labor market, imposing symmetry ensures that the legal restriction is "fair" across markets.

We use the notion of symmetry to evaluate restrictions on relative pay. As previously argued, the EPEW restriction alone does not imply symmetry, but we show that almost every firm structure that satisfies EPEW would also satisfy EPEW if pay and productivity were "swapped." Indeed, the minimal additional restriction to add to EPEW to ensure symmetry is EWEP, which are jointly equivalent to *strictly* increasing wages in productivity.

We also use symmetry to evaluate group-based, regression approaches to EPEW used in other countries. Specifically, group EPEW declares a firm structure to be unfair if, controlling for productivity, the regression coefficient on an indicator variable for being a member of a historically underrepresented group is negative.[8] We show formally that group EPEW is not consistent with the symmetry axiom; even if the desired coefficient is positive we show that reflecting the wage and productivity of each worker can lead to exactly the opposite conclusion, in other words, a violation of group EWEP. Specifically, workers in one group can be paid more on average than workers in the other controlling for productivity, while workers in the former

[7] Perhaps of interest, papers in both of these literatures, see e.g.s Winter (2004) and Cullen and Pakzad-Hurson (2023), find that violations of EPEW can arise in equilibrium.

[8] Group-based EPEW is used in a large literature aiming to measure differences in pay given equal work (or by holding work as similar as possible). Prominent examples include the literature on gender wage gaps (e.g., see Bertrand et al., 2010; Goldin, 2014; Blau and Kahn, 2017; Cook et al., 2021) and racial wage gaps (Neal and Johnson, 1996; Bertrand and Mullainathan, 2004; Bayer and Charles, 2018). However, EPEW studies have also focused on the wage gaps within other demographic variables such as parental status (Budig and England, 2001), sexual orientation (Carpenter, 2007), having a criminal record (Pager, 2003), height (Case and Paxson, 2008), physical attractiveness (Hamermesh and Biddle, 1993), and education (Dale and Krueger, 2002).

group are more productive on average than workers in the latter group controlling for wages. We show that this discrepancy between group EPEW and group EWEP occurs not just theoretically, but also empirically in several datasets. Moreover, modifications to the firm structure to "raise the regression coefficient" to satisfy group EPEW can create increased violations of group EWEP (and vice versa) *for the disadvantaged group*—if a firm structure satisfies group EWEP but not group EPEW, raising the wages of workers of the disadvantaged group can lead the "post intervention" firm structure to violate group EPEW. This suggests that wage increases for members of the minority group can decrease fairness, or, perhaps more likely, that group EPEW is not an ideal approach to pay fairness.

Taken together, our results suggest that pay structures satisfying individual EPEW are more likely to satisfy symmetry than those satisfying regression-based approaches to EPEW.

We review the related literature below. In Section 2, we formalize EPEW and present possibility/impossibility results around pay monotonicity. In Section 3, we discuss the symmetric notion of EWEP and conceptual and practical issues with group-level aggregation methods to studying pay fairness. In Section 4, we develop our empirical method. In Section 5, we describe our data and strategy to operationalize the method in this setting. Section 6 presents results from our empirical exercise. Section 7 concludes.

1.1 Related Literature

Individual-Level Fairness Notions. Fairness of outcomes for individuals with similar "relevant" characteristics are common in economics, and such "outcome tests" were popularized in Becker (1957, 1993). Becker's original formulation of these tests focused on marginal workers along a binary threshold assignment (hired/not). Outcomes tests have been applied to a variety of settings including lending (Peterson, 1981; Dobbie et al., 2021), police searches (Ayres, 2002), and other criminal justice settings (Hull, 2021; Canay et al., 2023), as well as job promotions (Benson et al., 2019; Huang et al., 2024). Because of the need to identify marginal subjects, papers in this setup often use an instrument and IV approach.

However, the concept can be generalized to a richer set of potential assignments (such as wages, sentence lengths, interest rates, etc.). Moving from a binary to a continuous assignment creates many observable "local margins," so in ideal circumstances the need for an external instrument disappears. However, that conclusion rests critically on the assignment rule being monotone and the same for every group.[9] Arnold et al. (2018) examine a continuous assignment (bail amount), yet the authors still exploit random judge assignment as an instrument because judges may differ in both level and slope of their bail-setting rules (violating the common-schedule assumption).

By contrast, our approach applies to within-firm, within-unit comparisons between workers. We assume a common within-firm rule that sets wages and productivities, and simply test whether there *exists* any nondecreasing mapping from observed productivity to wages that can generate the observed pay profile. Since every pair of workers are assumed to face the same rule, no exogenous variation or instrumental variable is needed.

Rationalization and Existence. To the best of our knowledge, there are no prior papers that jointly study EPEW and EWEP — either at the individual or group level — in the same data, or that give guidance about their theoretical relationship. Our rationalizability approach as a test for violations of monotonicity in multidimensional settings is conceptually akin to papers in information design (Bergemann and Morris, 2016, 2019) that ask whether private information exists to rationalize an agent's behavior. This approach has been used by other recent papers to study a variety of questions. Syrgkanis et al. (2017) use this approach for studying auctions; Magnolfi and Roncoroni (2023) use this approach in entry games; Bergemann et al. (2022) study counterfactuals with latent information, and Gualdani and Sinha (2019) study discrete choice. Most recently, Rambachan (2024) uses a similar approach to identify prediction mistakes from observational data, with an application to criminal justice.

[9] If decision makers tailor the schedule, if assignment is mismeasured, or if other unobservables contaminate the within-assignment comparison, the familiar infra-marginality problem returns, and an IV or structural approach is needed.

By contrast, group EPEW and group EWEP are less cohesive. We show that, at the same firm, the two regression-based tests can point in opposite directions: the same group of workers can have a group EPEW violation while the other group of workers can have a group EWEP violation. We document this discrepancy in three empirical settings using replication data from Bohren et al. (2023) and Cowgill et al. (2020), and in the academic research setting we study more broadly in this paper. To reconcile possible group EPEW/EWEP discrepancies, we contribute alternative fairness rules.

Determinants of Researcher Pay. There is also a literature that studies the determinants of pay among research faculty across several fields.[10] One strand of this literature examines whether faculty pay is "fair" across gender and demographic groups (Gordon et al., 1974; Hoffman, 1976; Hirsch and Leppel, 1982; Ginther and Hayes, 1999; Toutkoushian et al., 2007; Li and Koedel, 2017; Card et al., 2022; Harris and Maté-Sánchez-Val, 2022; Obloj and Zenger, 2022; Baker et al., 2023; Kim et al., 2024). These papers mostly take a regression-based approach to pay fairness, mirroring the approaches we discussed above, and which we study formally in Section 3.

This literature finds that journal publications, especially "top" journals, are particularly strong determinants of pay (Sauer, 1988; O'Keefe and Wang, 2013; Gibson et al., 2017; Lyu and Toda, 2019). We follow Card et al. (2022) in studying salaries in math, economics, and psychology, in which journal publications are the consensus "most important" form of productivity. Other forms of faculty effort, such as grants (Lyu and Toda, 2019) and teaching (Katz, 1973), contribute much less. Despite this, our methodological approach in Section 4.2 allows for "unobserved" forms of productivity besides research (such as teaching other contributions). Our procedure derives the minimum contribution of these other forms of productivity needed to rationalize the pay structure.

[10] This includes economics (Tuckman and Leahey, 1975; Sauer, 1988; Moore et al., 2001; Hilmer et al., 2015), finance (Swidler and Goldreyer, 1998; Garfinkel et al., 2024), marketing (Mittal et al., 2008), management (Gomez-Mejia and Balkin, 1992), and mathematics (Diamond Jr, 1986).

2 Model: Equal Pay for Equal Work

There is a single firm employing a finite set of workers I with $|I| \geq 2$. Each worker $i \in I$ is associated with a *productivity* $\theta_i \in \Theta \subset [0,1]$ and a *wage* $w_i \in [0,1]$. For $i \in I$, we refer to (θ_i, w_i) as the worker's *type*. Throughout, we make no restrictions on the support of Θ. That is, our analysis allows for the distribution of productivities to be rich, $\Theta = [0,1]$, such that the difference in productivities between two workers can take on any value in the interval $[0,1]$, but all of our results will hold if Θ is a proper subset of $[0,1]$.[11] The *firm structure* is represented by $S := (I, (\theta_i, w_i)_{i \in I})$.

A *fairness rule* $\mathcal{F}$ is a mapping from the space of firm structures to $\{0,1\}$, where a value of 1 indicates that the firm structure is "fair" and a value of 0 indicates that the firm structure is "unfair."

There are many fairness rules. We begin by formalizing individual-level notions of fairness codified in many legal statutes, which are based on the premise of Equal Pay for Equal Work (EPEW). EPEW restricts a firm from paying a worker strictly less than an equally-or-less productive coworker. If productivity is thought of as fixed and wages are set after the firm observes productivity, then EPEW encodes "equal treatment of equals" meaning that a violation of EPEW is a violation of Becker's outcome test appropriately contextualized to a continuous set of potential assignments (i.e. wages).

Definition 1. Equal Pay for Equal Work (EPEW) *is violated for worker $i \in I$ if there exists a worker $j \in I$ with $\theta_j \leq \theta_i$ and $w_j > w_i$.*

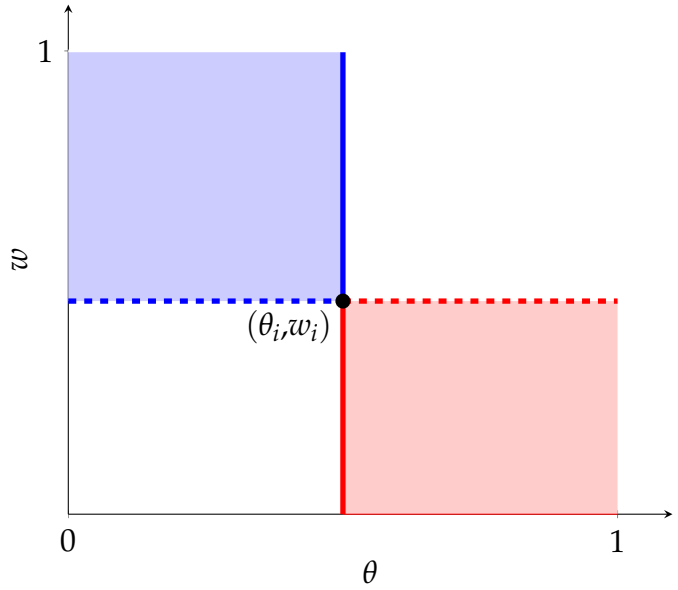
We say that a firm structure exhibits EPEW if there is no worker $i \in I$ for whom EPEW is violated. We emphasize that EPEW is an individual-level notion; a worker can be treated "unfairly" even if there is only one other worker who "gets a better deal." Note that the set of

[11] For modeling purposes, the support of Θ implicitly depends on specifics of the equal pay restriction. For example, equal pay for "similar" work laws coarsen the equal pay restriction, requiring it to bind for workers whose productivities are not necessarily equal (Gentile Passaro et al., 2024). In *Freyd v. University of Oregon* the 9^{th} Circuit Court of Appeals ruled that equal pay for similar work requires all professors of the same rank and within the same department to receive equal pay, thus, under an equal pay for similar work restriction, we would take Θ to be a singleton set. In our empirical setting, we briefly consider whether employers comport with equal pay for "similar" work following this definition, and in all cases, the answer is no.

firm structures satisfying EPEW is non-empty; if all workers earn the same wage then EPEW is satisfied regardless of their productivities. This definition bakes in a notion of anonymity: additional information about a worker's identity such as job title or demographic information do not play a role in EPEW. In reality, EPEW often applies only within a job type and across demographic groups. In Appendix A5, we enrich our model to include these real-world considerations, and discuss the robustness of our upcoming results.

Figure 1 demonstrates violations of EPEW. Specifically, we plot a single worker i employed by the firm, and represent her productivity on the horizontal axis and her wage on the vertical axis. EPEW restricts that there can be no other worker j employed by the firm who falls on the vertical blue line or shaded blue region (violation of EPEW for worker i), or on the vertical red line or shaded red region (violation of EPEW for worker j).

Figure 1: **EPEW Restriction**



Notes: This figure represents EPEW restrictions caused by a worker i; the firm structure is in violation of EPEW if there exists another worker j whose type falls in either the red or blue shaded regions, or on the red or blue solid lines.

Let $\mathcal{F}_L$ be a fairness rule such that a firm structure is deemed "fair" if and only if it satisfies EPEW—the "L" subscript references that $\mathcal{F}_L$ is the fairness rule prescribed by typical equal pay

laws. The following result characterizes the restrictions on firm structures induced by $\mathcal{F}_L$.

Proposition 1. *Let $S = (I,(\theta_i,w_i)_{i\in I})$ be a firm structure. $\mathcal{F}_L(S) = 1$ if and only if there exists a non-decreasing function $\phi : [\min_{i\in I} \theta_i, \max_{i\in I} \theta_i] \to [0,1]$ such that any worker i of productivity θ receives wage $w_i = \phi(\theta)$.*

This result finds that satisfying EPEW is equivalent to wages being non-decreasing in productivity. We therefore often refer to the EPEW restriction as inducing *pay monotonicity*. Pay monotonicity is easy to check via $|I-1|$ inequalities: ordering workers from least productive to most productive, is the wage of a worker lower than the next worker in the sequence? If the answer is "no" for any worker, then the firm structure S is not compliant with EPEW, that is, $\mathcal{F}_L(S) = 0$. If the answer is "yes" for all workers, then the firm structure S is compliant with EPEW, that is, $\mathcal{F}_L(S) = 1$.

2.1 Flexibility

How restrictive is pay monotonicity on firm structures? The following result discusses this topic. First, satisfying EPEW becomes more difficult as the number of workers $|I|$ increases—more workers means there are more opportunities for a violation. Therefore, satisfying pay monotonicity is more difficult as the number of workers increases (Item 1). Second, if there is a small distance between any potential new worker j of type (θ_j,w_j) to an already-employed worker i in either the productivity or wage dimension, the "flexibility" the firm has regarding hiring—that is, the set of types the firm can add while staying compliant with EPEW—shrinks to zero (Item 2).

Proposition 2 (Restrictive for Large Firms)**.** *Let $S := (I,(\theta_i,w_i)_{i\in I})$ be an arbitrary firm structure, and let $S^+ := (I\cup\{j\},(\theta_i,w_i)_{i\in I\cup\{j\}})$ be a firm structure that adds a worker $j \notin I$ to the firm.*

1. *$\mathcal{F}_L(S^+) = 1$ only if $\mathcal{F}_L(S) = 1$.*
2. *Fix any $\epsilon > 0$. There exists $\delta > 0$ such that if $\mathcal{F}_L(S) = 1$, and either:*

- *for any* $\theta \in [0,1]$ *there exists a worker* $i \in I$ *such that* $|\theta_i - \theta| < \delta$, *or*
- *for any* $w \in [0,1]$ *there exists a worker* $i \in I$ *such that* $|w_i - w| < \delta$,

then $\mathcal{F}_L(S^+) = 1$ *if and only if the new worker* j*'s type* $(\theta_j, w_j) \in V \subset [0,1]^2$ *where* $|V| < \epsilon$.

Item 2 has the following interpretation. The first premise is that the firm is initially compliant with the fairness rule; if the firm is not, then Item 1 states that the firm will also be non-compliant when it adds an additional worker. The second premise is that, for small δ, there is a fine "grid" of either wages or productivities already represented at the firm. Therefore, for small ϵ, the conclusion that the new worker j's type can only be drawn from a set V with (Lebesgue) measure strictly less than ϵ means that the firm has little flexibility when it comes to the type of a new worker it hires.[12] The overall implication is that a firm with many workers that remains compliant when adding an additional worker either must initially employ workers from a restricted set of types, or it is restricted to hiring a new worker from a restricted set of types.

3 Pitfalls of aggregation

Our analysis thus far formalizes EPEW as an *individual-level* restriction: within a firm, no worker can be paid strictly less than an equally or less productive coworker. In many policy discussions and in several jurisdictions outside the U.S., however, "equal pay" is operationalized using *aggregated* (group-level) comparisons—often by comparing average wages across groups controlling for some measure of output. This section explains why, in our setting, aggregation is a poor substitute for individual-level restrictions. In particular:

1. Aggregation can lead to conflicting takeaways depending on whether 'work' or 'pay' is treated as endogenous, and

[12] We note that our "flexibility" notion is based on the measure of the set V of worker types the can be hired. Alternative definitions can be defined by holding fixed one dimension of the worker's type. For example, if worker productivity is exogenous to the firm structure (i.e. there is no pay for performance), then a natural question is whether the wage of *any* worker can be raised by more than ϵ without violating monotonicity. The answer to this question is "no" under an alternative of the second premise of Item 2, wherein the "or" is replaced with "and."

2. Aggregation makes audits (including rationalizability-based audits) weak in settings with multidimensional productivity.

3. Aggregation can "average out" individual unfairness,

We explore each of these three issues in the sequel. To do so, we formalize a common, regression-based aggregation method to determine fairness that is used in Canada and in much of the academic literature (as discussed in the Introduction).

To formalize these ideas, we introduce additional notation. For any firm structure $S := (I,(\theta_i,w_i)_{i\in I})$, let $A \subset I$ and $B = I \setminus A$ form a partition of the set of workers. We will refer to each $i \in A$ as a member of group A and to each $i \in B$ as a member of group B. We take the convention that the B group is the "historically disadvantaged" group such that worse outcomes for this group trigger a firm structure to be labeled unfair. Below, we consider two separate forms of aggregation to compare disparities across groups. Implicitly, we limit our study to firm structures with rich enough sets of worker types represented in each group so that all upcoming concepts are well-defined; otherwise we simply define the firm structure in question as being fair.[13]

Iso-Productivity Bands. The first aggregation approach compares a weighted average of violations of EPEW across iso-productivity bands. This approach is more natural when there is a common support for productivities across the two groups in a dataset. Let $\hat{\Theta} \subset \Theta$ be the subset of productivity levels that are represented by workers in both groups: $\hat{\Theta} := \{\theta | \exists i \in A \text{ with } \theta_i = \theta \text{ and } j \in B \text{ with } \theta_j = \theta\}$.

Let

$$\beta^{isoprod} = \sum_{\theta \in \hat{\Theta}} \phi_\theta \tau(\theta), \tag{1}$$

where $\tau(\theta)$ represents the average group wage difference within the $\theta \in \hat{\Theta}$ iso-productivity band,

[13] For example, the group-level regression coefficients we discuss are unidentified if all workers belong to the same group.

and ϕ_θ is the weight associated with iso-productivity band $\theta \in \hat{\Theta}$. The weights ϕ_θ are given by:

$$\phi_\theta = \frac{\text{var}(B_i \,|\, \theta_i = \theta)\Pr(\theta_i = \theta)}{\sum_{\theta \in \hat{\Theta}} \text{var}(B_i \,|\, \theta_i = \theta)\Pr(\theta_i = \theta)},$$

where B_i is an indicator that equals 1 if worker i is a member of group B. Note that by construction, $\phi_\theta \in [0,1]$ for all $\theta \in \hat{\Theta}$, and $\sum_{\theta \in \hat{\Theta}} \phi_\theta = 1$.

Following Goldsmith-Pinkham et al. (2024); Angrist (1998), $\beta^{isoprod}$ as defined in Equation (1) can be represented via the following linear regression

$$w_i = \alpha + \beta^{isoprod} B_i + \underbrace{\sum_{\theta \in \hat{\Theta}} \gamma_\theta \mathbb{1}(\theta_i = \theta)}_{\substack{\text{Iso-productivity band} \\ \text{fixed effects}}} + \epsilon_i, \tag{2}$$

where $\mathbb{1}(\theta_i = \theta)$ is an indicator function that equals 1 if $\theta_i = \theta$, and 0 otherwise. Here, γ_θ represents the fixed effect associated with the θ iso-productivity band.

Linear Productivity Trend. The second aggregation approach considers the linear relationship of productivity on wages, and considers whether one group receives a larger average deviation from this linear trend. Therefore, this aggregation is more natural when there are few workers who have the exact same productivity.

In what follows, "barred" terms represent the average value of that variable across workers $i \in I$ (e.g. $\bar{\theta}$ represents the average value of θ_i across all workers $i \in I$). Let

$$\beta^{contprod} = \frac{\sum_{i \in I} \varepsilon_{w_i} \varepsilon_{B_i}}{\sum_{i \in I} \varepsilon_{B_i}^2}, \tag{3}$$

where, for $X_i \in \{B_i, w_i\}$, ε_{X_i} is the residual from regressing X_i on θ_i, that is:

$$\varepsilon_{X_i} = X_i - \bar{X} - \frac{\sum_{j \in I} (\theta_j - \bar{\theta})(X_j - \bar{X})}{\sum_{j \in I} (\theta_i - \bar{\theta})^2} (\theta_i - \bar{\theta}).$$

$\beta^{contprod}$ in Equation (3) can be represented as linear regression

$$w_i = \alpha + \beta^{contprod} B_i + \gamma\theta_i + \epsilon_i, \tag{4}$$

where θ_i is a linear control for productivity.

Therefore, $\beta^{isoprod}$ and $\beta^{contprod}$ measure notions of "group EPEW." Let $\mathcal{F}^{isoprod}$ and $\mathcal{F}^{contprod}$ be fairness rules such that a firm structure is deemed "fair" if and only if, respectively, $\beta^{isoprod} \geq 0$ and $\beta^{contprod} \geq 0$, i.e., that B-group workers receive no lower pay than A-group workers after controlling for productivity.[14]

3.1 Aggregation can generate "the contradiction": symmetry, endogeneity, and EWEP

The first issue is conceptual. Aggregated regression-style rules treat "pay" and "work" asymmetrically, and this asymmetry interacts poorly with the fact that in many environments it is unclear whether productivity determines wages or wages determine productivity (or both). Indeed, economists study both possibilities. Without clear evidence on which direction causality flows, it may be reasonable for an analyst to conclude unfairness if and only if a firm structure is deemed unfair when pay is viewed as the dependent variable *and* when work is viewed as the dependent variable.

To implement this restriction, we discipline the set of admissible fairness rules using a symmetry requirement.

Definition 2 (Reflection and symmetry). *For any firm structure $S = (I, (\theta_i, w_i)_{i \in I})$, define its* reflection $S^r := (I, (\theta'_i, w'_i)_{i \in I})$ *by $\theta'_i = w_i$ and $w'_i = \theta_i$ for all $i \in I$. A fairness rule $\mathcal{F}$ is* symmetric *if*

[14] These fairness rules are based on the point estimates of the regressions described above. Similar notions can be defined to account for statistical significance declaring a firm structure to be unfair if the p-value of the (one- or two-sided) null hypothesis that the relevant coefficient differs from zero. Doing so does not change our qualitative results; indeed, the tests we describe consider the limiting case as we approach infinite data. Also, our results are unchanged if these fairness rules define a firm structure to be fair if the respective coefficient falls within an arbitrary interval in $[-1,1]$, i.e. within the support of possible values the coefficients can take given our normalization that each worker's type falls in $[0,1] \times [0,1]$.

$\mathcal{F}(S) = \mathcal{F}(S^r)$ *for every* S.

Symmetry rules out "contradictory" conclusions regarding the fairness of a firm structure that arises if the analyst considers wages as the outcome and productivity as the control variable, or productivity as the outcome and wages as the control. Moreover, and as discussed in the introduction, symmetry is potentially seen as an important fairness concept innately, and it also may be appealing to consider for a policy maker who must impose a common restriction on firm structures both in industries where a worker's wage determines her productivity, and in industries where a worker's productivity determines her wage.

Of course, fairness rules that fail symmetry may do so to different degrees. The fairness rule that fails symmetry but satisfies the following condition is a "minor" failure in the sense that a firm structure "almost surely" does not yield contradictory fairness results under reflection.[15]

Definition 3 (Near symmetry)**.** *A fairness rule* $\mathcal{F}$ *is* nearly symmetric *if, for any* $\epsilon > 0$ *and any firm structure* S*, there exists an "*ϵ*-perturbed firm structure"* $S_\epsilon = (I, (\theta_i^\epsilon, w_i^\epsilon)_{i \in I})$ *such that:* $|\theta_i - \theta_i^\epsilon| < \epsilon$ *for all* $i \in I$*,* $|w_i - w_i^\epsilon| < \epsilon$ *for all* $i \in I$*, and* S_ϵ *and its reflection* S_ϵ^r *satisfy* $\mathcal{F}(S_\epsilon) = \mathcal{F}(S_\epsilon^r) = \mathcal{F}(S)$*.*

Below, we analyze individual and group-level EPEW in terms of symmetry.

3.1.1 Individual-level symmetry: Equal Work for Equal Pay (EWEP)

The symmetric notion of EPEW is that if a worker is paid more, she should not be less productive. We call this "Equal Work for Equal Pay" (EWEP).

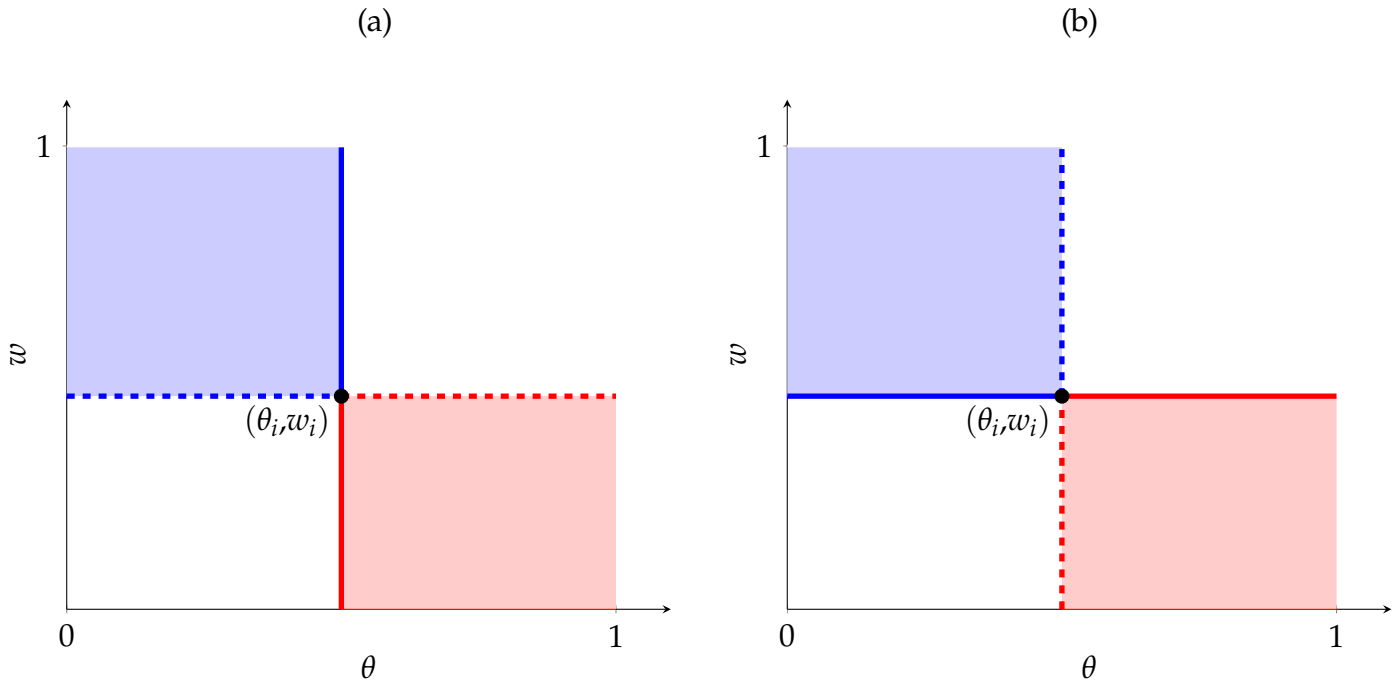
Definition 4. Equal Work for Equal Pay (EWEP) *is violated for worker* $i \in I$ *if there exists a worker* $j \in I$ *with* $w_j \geq w_i$ *and* $\theta_j < \theta_i$*.*

We say that a firm structure exhibits EWEP if there is no worker $i \in I$ for whom EWEP is violated. EWEP does not impose the same restriction as EPEW. For example, if all workers have the same productivities and all receive different wages, then EWEP is satisfied but EPEW

[15] Specifically, for a nearly symmetric fairness rule and holding fixed I, the set of wage and productivity tuples for workers in I that yield contradictory fairness results has zero Lebesgue measure.

is not. Figure 2 displays the difference graphically. Panel (a) replicates Figure 1 and displays violations of EPEW given the presence of a worker i. Panel (b) depicts violations of EWEP: there can be no other worker j employed by the firm who falls on the horizontal blue line or shaded blue region (violation of EWEP for worker i), or on the horizontal red line or shaded red region (violation of EWEP for worker j).

Figure 2: **EPEW and EWEP Restrictions**



Notes: Panel (a) represents EPEW restrictions caused by a worker i; the firm structure is in violation of EPEW if there exists another worker j whose type falls in either the red or blue shaded regions, or on the red or blue solid lines. Panel (b) represents EWEP restrictions caused by a worker i in a similar way.

Let $\bar{\mathcal{F}}_L$ be a fairness rule such that a firm structure is deemed fair if and only if it satisfies both EPEW and EWEP. Recall that $\mathcal{F}_L$ deems a firm structure fair if and only if it satisfies EPEW, which means that $\bar{\mathcal{F}}_L$ deems fewer firm structures fair than does $\mathcal{F}_L$. Formally, we define a partial order over fairness rules: fairness rule $\mathcal{F}$ is *stronger* than fairness rule $\mathcal{S}'$ if for any firm structure S, $\mathcal{F}(S)=1$ only if $\mathcal{S}'(S)=1$.

The following result characterizes $\bar{\mathcal{F}}_L$, and the compliance of both $\mathcal{F}_L$ and $\bar{\mathcal{F}}_L$ with our symmetry axioms.

Proposition 3.

1. *Let* $S := (I, (\theta_i, w_i)_{i \in I})$ *be a firm structure.* $\bar{\mathcal{F}}_L(S) = 1$ *if and only if there exists a strictly increasing function* $\psi : [\min_{i \in I} \theta_i, \max_{i \in I} \theta_i] \to [0,1]$ *such that any worker* i *of productivity* θ *receives wage* $w_i = \psi(\theta)$.

2. $\bar{\mathcal{F}}_L$ *is symmetric, and* $\mathcal{F}_L$ *is almost symmetric. Moreover, any symmetric fairness rule* $\mathcal{F}$ *that is stronger than* $\mathcal{F}_L$ *is also stronger than* $\bar{\mathcal{F}}_L$.

Item 1 of the proposition characterizes $\bar{\mathcal{F}}_L$—satisfying both EPEW and EWEP implies that wages are strictly increasing in productivity. Item 2 states that EPEW imposes a nearly symmetric fairness rule, and that EPEW and EWEP together impose a symmetric fairness rule. Moreover, EWEP is minimal restriction that must be added to EPEW to generate a symmetric fairness rule. In other words, if one wishes to strengthen existing equal pay laws to ensure symmetry is satisfied satisfied, then one needs to place an additional restriction that is no weaker than EWEP.

The following remark displays the "closeness" of symmetry and near symmetry in applied settings:

Remark 1. *The premise and conclusion of Proposition 2 are unchanged under fairness rule* $\bar{\mathcal{F}}_L$ *instead of* $\mathcal{F}_L$.[16]

3.1.2 Group-level regression rules are not symmetric

We define group EWEP by "flipping" the dependent and independent variables in our group EPEW machinery. Specifically, $\beta^{isowage}$ is defined analogously to $\beta^{isoprod}$ but where θ, θ_i, and $\hat{\Theta}$ are replaced, respectively, with w, w_i, and $\hat{W}$ in Equation (1) and Equation (2). That is, $\beta^{isowage}$ determines the weighted average difference in productivity across groups within iso-wage bands. Similarly, $\beta^{contwage}$ is defined analogously to $\beta^{contprod}$ but where θ_i and $\bar{\theta}$ are replaced,

[16]The closeness of symmetry and near symmetry is also borne out in our analysis of firm structures conducted in Section 6. Only 6% of university-department-rank-year cells contain two workers with the same salary and different productivity metrics, which implies that in at most 6% of cells could the firm structure satisfy EPEW but not the combination of EPEW and EWEP.

respectively, with w_i and $\bar{w}$ in Equation Equation (3) and Equation (4). That is, $\beta^{contwage}$ determines whether group B has, on average, a larger average deviation from the linear best-fit of productivity on wages.

Let $\mathcal{F}^{isowage}$ and $\mathcal{F}^{contwage}$ be fairness rules such that a firm structure is deemed "fair" if and only if, respectively, $\beta^{isowage} \leq 0$ and $\beta^{contwage} \leq 0$, i.e., that B-group workers have no higher productivity than A-group workers after controlling for pay.

We find that neither method of measuring group EPEW is symmetric or near symmetric:

Proposition 4. *$\mathcal{F}^{isoprod}$, $\mathcal{F}^{isowage}$, $\mathcal{F}^{contprod}$, and $\mathcal{F}^{contwage}$ are all distinct fairness rules, and all fail to satisfy both symmetry and near symmetry.*

Proposition 4 has two central takeaways: first, the four group-level regression approaches detailed in Equation (1) and Equation (3) which consider both productivity and wage as the dependent variable, separately, all admit different fairness rules. While it is well known that Equation (1) and Equation (3) are not equivalent *holding fixed the dependent variable*, we show that "flipping" the dependent variable can yield a different outcome. Second, all four of the group-level regression approaches fail to satisfy both symmetry and near symmetry.

To see that all four fairness rules fail to satisfy the symmetry axiom, consider the simple example depicted in Table 1. The table presents a firm structure for which, regardless of whether the approach in Equation (1) or Equation (3) is used, there is a fundamental conflict in determining fairness: if pay is the dependent variable (group EPEW) then the regression approaches deem the firm structure in question fair because B group workers are less productive on average within each iso-wage band (each row). On the other hand, if productivity is the dependent variable (group EWEP) then the regression approaches deem the firm structure in question unfair because B group workers are paid less on average within each iso-productivity band (each column). In other words, common regression approaches used to measure fairness are not well-founded in an axiomatic basis that symmetrically treats wages and productivities. The proof of Proposition 4 extends the argument to show that near symmetry is also not satisfied.

We also note that the firm structure in Table 1 violates (the individual notions of) EWEP and

EPEW discussed in the previous sections, as not all workers earning the same wage are equally productive, and not all workers who are equally productive earn the same wage.

Table 1: **Equal Work and Pay Counterexample**

	θ_ℓ	θ_h
w_ℓ	a,b,b	a,b
w_h	a,b	a,a,b

Notes: This table presents a firm structure S with 10 workers. There are 5 workers in group A and 5 in group B. There are two wage levels $w_\ell < w_h$, and two productivity levels $\theta_\ell < \theta_h$ represented in the firm structure. There is one worker of group A and two workers of group B who receive wage w_ℓ and are of productivity θ_ℓ, and so on.

Moreover, these regression-based approaches investigating group-level disparities do not even satisfy intuitive notions of monotonicity. Specifically, Remark 2 shows that a firm structure S can be deemed unfair when wage is the dependent variable and fair when productivity is the dependent variable, but a firm structure S' that differs from S only in that the wages of $B-$group workers are raised is deemed fair when wage is the dependent variable and unfair when productivity is the dependent variable.[17] Therefore, this finding suggests that Canada's prescription to raise women's wages if their "regression line" is below that of men's using a group-based EPEW approach can lead to a violation *against women* of the symmetric notion of fairness.[18]

Remark 2. *Let S be a firm structure and let S' be a "post-intervention" firm structure such that the wages of each B-group worker is higher in S' than S, and no other changes are made. It is possible that $\mathcal{F}^{isoprod}(S)=0$ and $\mathcal{F}^{isowage}(S)=1$, while $\mathcal{F}^{isoprod}(S')=1$ and $\mathcal{F}^{isowage}(S')=0$. By symmetry, similar statements can be made about a firm structure S and a "post-intervention" firm structure S' such that S' only changes the productivity of B-group workers compared to S.*

Proof. Remark 2 follows from inspection of the following two firm structures, which reveals

[17] As can be seen from Equation (3), fairness rules $\mathcal{F}^{contprod}$ and $\mathcal{F}^{contwage}$ depend on the "intensive margin" values that productivity and wages take, whereas Equation (1) reveals that fairness rules $\mathcal{F}^{isoprod}$ and $\mathcal{F}^{isowage}$ depend on the "extensive margin" values that productivity and wages take in that only workers within the same iso-productivity and iso-wage bands, respectively are considered. To simplify the exposition of the remark, we state it only in terms of $\mathcal{F}^{isoprod}$ and $\mathcal{F}^{isowage}$, although the finding can be extended to comparisons between $\mathcal{F}^{contprod}$ and $\mathcal{F}^{contwage}$ at the cost of additional specificity.

[18] For details on the specific language and citation for the policy, please recall the associated passage on page 2.

similar violations as in Table 1:

Table 2: **"Non-monotonicity" Example**

S

	θ_ℓ	θ_h
w_ℓ	a,b	a,b
w_h	a	$\varnothing$

S'

	θ_ℓ	θ_h
w_ℓ	a	a
w_h	a,b	b

Notes: These tables presents two firm structures with 5 workers. There are 3 workers in group A and 2 in group B. There are two wage levels $w_\ell < w_h$, and two productivity levels $\theta_\ell < \theta_h$ represented in the firm structure. The first firm structure S involves both $B-$group workers receiving pay w_ℓ, and the second firm structure S' differs from S only in that both $B-$group workers receive wage w_h.

Firm structure S violates group EPEW because $A-$group workers of productivity θ_ℓ earn more on average than the $B-$group worker of productivity θ_ℓ, but it does not violate group EWEP because workers earning wage w_ℓ are equally productive on average across groups. By contrast, firm structure S' (which increases the pay of both $B-$group workers) violates group EWEP because the $B-$group workers earning wage w_h are more productive on average than the $A-$group worker earning w_h, but it does not violate group EPEW which case be seen because both $B-$group workers earn the highest wage within the firm. □

Empirical Settings With Incongruous Regression-based Fairness Findings Is our finding in Proposition 4 that group EPEW and group EWEP differ just a theoretical curiosity, or do such cases appear in actual data? Appendix E presents three empirical settings containing firm structures S that confirm our theoretical finding: namely, that $\mathcal{F}^{isoprod}(S) \neq \mathcal{F}^{isowage}(S)$ and $\mathcal{F}^{contprod}(S) \neq \mathcal{F}^{contwage}(S)$. In all examples, workers are categorized by the authors as having a mutually exclusive group identity of either male or female. The first setting from a well-known paper (Bohren et al., 2023) is an experiment in an online labor market for coding jobs, where productivity is measured by performance on a math-based standardized test and subjects assign pay to workers. The second setting from Cowgill et al. (2020) is of the labor market for tech workers, where worker resumes are assigned pay statistician-developed inferences of

productivity. The final setting is of the labor market for university faculty that we describe in Section 5, where productivity is measured based on research metrics.

In all cases, we run regressions nearly identical to those relating to Equation (1) and Equation (3), separately considering productivity and wages as dependent variables.[19] In all of these examples, women receive lower pay than men controlling for productivity ($\beta^{isoprod}$,$\beta^{contprod} < 0$ if we take women to be group B) and are less productive than men when controlling for pay ($\beta^{isowage}$,$\beta^{contwage} < 0$ if we take women to be group B).

Moreover, Remark 2—that interventions to raise the pay of members of disadvantaged group can lead to violations of group EWEP—is also more than a theoretical possibility. Bohren et al. (2023) contains a experimental intervention. In Appendix E1, we analyze the effects of this treatment on group EPEW and group EWEP. Although our results are not always precise, they suggest the intervention harmed group EPEW and did not improve group EWEP; if anything, the group-EWEP point estimates move slightly in the opposite direction

Alternative aggregation methods There exist different group-level measures of fairness that satisfy our axioms. Given that the approaches above control for productivity when considering wages (or vice versa), it seems intuitive that relevant notions of fairness may take into account both wage and productivity discrepancies across groups. We outline two such approaches in Appendix A6.

3.2 Aggregation weakens auditability in multidimensional settings

The second issue is practical and especially salient in the empirical setting of this paper. Output is typically *multidimensional*, and any aggregated audit must choose a *scalar* notion of "work" (explicitly or implicitly). If the analyst (or the firm) is free to choose weights across many productivity dimensions, then aggregated fairness tests become extremely easy to satisfy ex post.

[19] The only difference from these two regression approaches is that for the settings described above that have multiple firms (i.e. the labor market for university faculty), we include firm fixed effects to separately estimate regression coefficients within, rather than across, firms.

Suppose observed productivity is a vector $\vec{\theta}_i \in [0,1]^K$, for some $K \geq 1$. Suppose that overall productivity is defined via a function $f : [0,1]^K \to [0,1]$, i.e. $f(\vec{\theta}_i) \in [0,1]$ is worker i's productivity.

Remark 3 ("Cherry-pick a dimension" makes group audits weak)**.** *If an equal pay audit compares groups only via averages controlling for productivity (as above), then an auditor can easily rationalize fair pay.*

Under the continuous productivity approach, if there exists any *productivity dimension k for which* $\gamma(\bar{\theta}_A^k - \bar{\theta}_B^k) \geq \bar{w}_A - \bar{w}_B$*, i.e., the weighted average difference in k-productivity across groups exceeds the average difference in wages across groups, then choosing* $f(\vec{\theta}_i) := \frac{\theta_i^k}{\max_{j \in I} \theta_j^k}$*—i.e. putting all weight on dimension k—rationalizes the observed group pay gap, even if* $B-$*group workers outperform* $A-$*group workers on other dimensions. The isoproductivity approach admits yet another issue: cherry picking the dimension of productivity can create an incomparability problem given that the support of the productivity distributions across groups may differ.*

This is not merely a philosophical objection: it is a direct warning about degrees of freedom in real-world equal-pay audits. When "work" is high-dimensional and weights are not fixed ex ante, aggregation creates substantial scope for post-hoc rationalization.

3.3 "Averaging out" individual unfairness

The third issue is mechanical. Aggregation that averages outcomes across groups can look "fair" even when most individuals are treated unfairly. A canonical pattern is that extreme observations offset a systematic disparity for many others. For example, consider a firm structure in which all workers have identical productivity (so individual-level EPEW would require equal pay across all workers). There is an arbitrary number of A-group workers who each earn a wage of 0.5 There are ten B-group workers; nine of them each earn a wage of 0.45, while one "superstar" $B-group$ worker earns 0.95. The average wage is 0.5 for each group, and given that productivities of all workers are equal, both $\beta^{isoprod} = 0$ and $\beta^{contprod} = 0$. That is, regression-based fairness rules $\mathcal{F}^{isoprod}$ and $\mathcal{F}^{contprod}$ view the given firm structure as fair even

though 90% of B-group members are strictly underpaid relative to every A-group member.

3.4 Takeaway

Taken together, these three points motivate the paper's focus on *individual-level* restrictions (pay monotonicity) and on transparency about how multidimensional performance is aggregated into pay. In the sections that follow, we develop and apply an empirical framework that asks whether there exists *any* reasonable aggregation of observable productivity metrics that can rationalize observed pay as monotone—an approach that is deliberately conservative in declaring violations, yet still frequently fails in the data.

4 Empirical Method

In this section, we develop a method for empirically measuring a firm structure's monotonicity. We provide both a binary test of monotonicity, as well as "distance to monotonicity" methods. These can be used as fairness tests, as well as an outcome metric to study whether policy changes move "closer" to monotonicity.

In settings where a firm has a single, known productivity metric, testing for monotonicity is straightforward. A more challenging case is when a firm has multiple productivity metrics because an observer testing for monotonicity does not know the relative weight put on each productivity dimension used by the firm. Moreover, a firm wishing to avoid a lawsuit may ex-post declare a set of weights that make it appear to be paying "fair" wages.

Our approach addresses both of these difficulties in as conservative a manner as possible. We ask whether there exist any linear combination of all productivity metrics that would rationalize the observed firm structure as satisfying monotonicity.

Preliminaries. Our methodology cleanly matches our theoretical setup. A firm employs (at a particular time) a set of workers I. Each worker $i \in I$ is associated with a *productivity vector*

of outcomes $\vec{\theta}_i$ of dimension $K \in \mathbb{N}$ and a *wage* w_i. Each element of $\vec{\theta}_i$ represents the worker's performance on a different productivity metric. Insofar as the firm cares about the square, cube, logarithm, or other transformation of a productivity metric, these transformations can be added to as a separate productivity metric. There is some function $f : \mathbb{R}^{\mathbb{K}} \to \mathbb{R}$ which maps the K productivity metrics into a single-dimensional productivity measure, which is unobserved by the econometrician.[20]

We assume that the function f is a linear mapping. That is, the firm places a weight $\vec{\beta}$ on the vector of productivity metrics. The worker's single-dimensional productivity can therefore be represented as $y_i := f(\vec{\theta}^i) = \vec{\theta}^i \cdot \vec{\beta}$, or equivalently $y_i = \theta_1^i \beta_1 + \theta_2^i \beta_2 + ... + \theta_K^i \beta_K$.

Definition 5 (Monotonicity Test)**.** *A wage distribution satisfies monotonicity if* $w_i > w_j \implies y_i > y_j$*, for all workers* $i,j \in \{1,...,|I|\}$.

Notice that we do not include pairwise inequalities corresponding to workers who have the exact same salary; only for pairs of workers with a strict salary difference do we require the higher-paid worker to be more productive. This is because we are testing for weak monotonicity, which reflects legal restrictions (Proposition 1).

Remark 4. *The monotonicity test for a firm structure can be represented as a system of at most* $\binom{|I|}{2}$ *pairwise inequalities.*

Although there are $\binom{|I|}{2}$ different pairs of workers, there are several reasons why a monotonicity test can sometimes be represented by strictly fewer than $\binom{|I|}{2}$ pairwise inequalities. First, pairwise constraints are not imposed on pairs of workers who earn the same wages. Second, different pairs of inequalities could be collinear, meaning some restrictions are redundant. Finally, and discussed more next, legal restrictions may not bind across all pairs of workers. For example, some states laws are are particularly concerned with pairwise comparisons between men and women (rather than between pairs of equally gendered workers).

[20] Our theoretical model took each worker's productivity to lie in the interval [0,1], however, this was merely a normalization. Naturally, productivity metrics such citation and publication counts fall in larger (closed and bounded) intervals within $\mathbb{R}$. All of our results would be unchanged given a rescaling of productivity metrics to fall in [0,1].

The following example displays the set of inequalities to be satisfied in a three-worker firm.

Example 1. Consider a three-work firm structure where worker 1 is paid strictly more than worker 2, and worker 2 is paid strictly more than worker 3. The monotonicity requirement can be represented by a system of $\binom{3}{2} = 3$ pairwise inequalities:

$$\vec{\theta}^1 \cdot \vec{\beta} > \vec{\theta}^2 \cdot \vec{\beta} \qquad \text{(1's productivity is greater than 2's)} \tag{5}$$

$$\vec{\theta}^1 \cdot \vec{\beta} > \vec{\theta}^3 \cdot \vec{\beta} \qquad \text{(1's productivity is greater than 3's)} \tag{6}$$

$$\vec{\theta}^2 \cdot \vec{\beta} > \vec{\theta}^3 \cdot \vec{\beta} \qquad \text{(2's productivity is greater than 3's)} \tag{7}$$

Definition 6. *The pairwise inequalities define a feasible set for the weight vector $\vec{\beta}$:*

$$\mathcal{B}^G \equiv \left\{ \vec{\beta} \in \mathbb{R}^K : w_i > w_j \implies \vec{\beta} \cdot \vec{\theta}^i > \vec{\beta} \cdot \vec{\theta}^j, \forall i \neq j \right\}.$$

A wage structure is rationalizable *if the feasible set $\mathcal{B}^G$ is non-empty.*

Stated differently, the wages are "rationalizable" if there exists a non-empty set of weights that satisfy the system of pairwise inequalities.

Satisfying Rationalizability. A few key properties make the system of inequalities easier or harder to rationalize. First, as the number of productivity dimensions grows, it is easier (mechanically) to rationalize the wage distribution. Similarly as the number of inequalities shrinks, the system becomes easier to satisfy (see Proposition 2).

More generally, rationalizability is generous to the firm. As long as there exists a set of weights that solve the system of inequalities, the test returns positive. Therefore, estimates from the implementation of this methodology will yield a lower bound on the true frequency of monotonicity violates.

Additional Constraints. Our test thus far is focused entirely on the minimum requirements of the monotonicity test in Definition 5 and Remark 4. In some cases, additional constraints can

be justified that may help narrow the set of weights that can rationalize the firm structure. For example, certain β_k coefficients could be constrained to be have a particular sign. A negative weight could be useful in cases where certain performance metrics (i.e., mistakes) should be negatively correlated with success. Imposing that a weight must be positive could be justified for metrics that are known to be valuable to the firm.

Other constraints may also be justified. To be solvable using our approach above, the constraints simply need to be expressed as an additional set of inequalities within the linear program. Although some constraints may be outside this paradigm, it is flexible. For example, it is possible to express constraints on the maximum or minimum of β_k weights as a set of linear inequalities (as we do next).

Implementation. Our definitions above deliberately use strict inequalities. When we bring this technique to data, one possibility is to use symbolic linear programming. This is computationally expensive, and so an alternative is to use numeric linear programming techniques, which seek solutions with weak inequalities on productivities (i.e., $w_i > w_j \implies y_j \geq y_j$). However a solution with weak inequalities on productivities leads to a violation of EPEW. In Appendix B1, we show how we rule out this possibility using weak inequalities alone.

4.1 Distance to Monotonicity: Dropping Minimal Salary Differences

The procedures above generates an empty set of productivity weights when the system of inequalities cannot be rationalized. However, empty sets are not all equal. Some monotonicity violations might be relatively minor, and some could be much larger.

In this section and the next, we pursue two approaches to measuring "distances" to rationalizability. These two measures can be motivated by distinct "fixes" to monotonicity violations.

1. The first (below) is motivated by the size of wage changes needed to eliminate non-monotonicity.
2. The second (in Section 4.2) is about the level of unobserved productivity needed to rationalize wages.

Both measures are equal to zero at companies whose wage structures are monotonic, but may differ about which wage structures are qualitatively "large" (discussed in Section 4.3). These two measures are not exhaustive, and there may be other ways to measure distance from monotonicity.

Minimum Salary Differences. Our first approach is to impose a minimum salary difference δ, and drop all pairwise inequalities where the salary difference lies weakly below δ. We then find the smallest δ that would rationalize the observed salaries as monotonic ("$\underline{\delta}$"). $\underline{\delta}$ represents a dollar figure which may or may not be "large" relative to other benchmarks. As motivating intuition, "small" values (say) $\underline{\delta} = \$1$ may still allow a firm structure to be considered close to fair, whereas a $\underline{\delta}$ equal to half a year's salary would likely not.

Example 2. Consider a three-worker firm structure as in Example 1: Worker 1 is paid more than worker 2, who is paid more than worker 3. We assume that all β_k terms are weakly positive. Suppose worker 1's productivity is higher on all dimensions than 2 and 3. However, worker 3 is more productive than worker 2 on all dimensions, but is paid \$X less than worker 2. The original system of inequalities therefore fails.

As a measure of the "distance" to monotonicity, we consider the revised system of inequalities using a $\delta > 0$ tolerance. If $\delta \geq \$X$, this would drop the third inequality in the system (the final one comparing workers 2 and 3). Assuming no other monotonicity violations, the simplified system will return a feasible set.

$$\beta_k \geq 0, \forall k \qquad \text{(All weights weakly positive)} \tag{8}$$

$$\vec{\theta}^1 \cdot \vec{\beta} > \vec{\theta}^2 \cdot \vec{\beta} \qquad \text{(1's productivity is greater than 2's)} \tag{9}$$

$$\vec{\theta}^1 \cdot \vec{\beta} > \vec{\theta}^3 \cdot \vec{\beta} \qquad \text{(1's productivity is greater than 3's)} \tag{10}$$

$$\cancel{\vec{\theta}^2 \cdot \vec{\beta} > \vec{\theta}^3 \cdot \vec{\beta}} \qquad \cancel{\text{(2's productivity is greater than 3's)}} \tag{11}$$

The smallest δ for which the revised system has a non-empty solution is $\underline{\delta}$, our first way to measure distance from monotonicity.

$\underline{\delta}$ can be interpreted in at least two ways. First, suppose that a worker must initiate a lawsuit in order to correct unfair pay. An absence of lawsuits against a firm given a value of $\underline{\delta} > 0$ implies that no worker's expected net winnings from engaging in the lawsuit is greater than $\underline{\delta}$. Second, suppose the econometrician observes wages with error but that market participants do not (e.g. if workers receive bonus payments that are not visible to the econometrician). Then an alternative explanation to a firm's structure as being unfair is that the "error" terms offset violations of monotonicity. The validity of this alternative depends on the size of $\underline{\delta}$; a high value indicates that, for some pair of workers, the sum of the error terms is $\underline{\delta}$, which may or may not be realistic given sources of error in wage observance.

4.2 Distance to Monotonicity: Unobserved Productivity

Our second way of measuring distance to monotonicity accounts for the possibility that some forms of productivity might be unobserved. This is a classic problem in organizational economics (Holmstrom and Milgrom, 1991). To handle this situation, we introduce a new term, η_i, representing worker i's unobserved productivity, i.e., $y_i = \vec{\theta}^i \cdot \vec{\beta} + \eta_i$. In most settings, η_i could represent a bundle of hard-to-measure factors that firms detect and incorporate into pay, but are not codified. Because each term η_i appears only in worker i's output, the model can now rationalize any wage ordering by appropriately choosing $(\vec{\beta}, \eta)$. In other words, the system of inequalities becomes trivially feasible once we allow unrestricted η_i terms.

To obtain a meaningful measure of distance to monotonicity, we constrain the size of the unobserved terms relative to the observed productivity metrics. Specifically, we require that the largest unobserved term be no larger than some multiple $\rho \geq 0$ of the largest weighted observable metric:

$$\rho \cdot \max_k \beta_k \geq \max_i \eta_i.$$

For a given ρ, we ask whether there exists $(\vec{\beta},\eta)$ that rationalizes the observed wages as strictly monotone in latent productivity. We then define $\underline{\rho}$ as the smallest value of ρ for which such a solution exists. This quantity $\underline{\rho}$ is our second measure of distance from monotonicity. Economically, $\underline{\rho}$ measures how large unobserved productivity must be—relative to the most heavily weighted observable metric—to rationalize the observed pay ranking.

In some settings, $\underline{\rho}$ could be implausibly large. For example, consider a sales context in which an "organizational citizenship" dimension (teamwork, initiative, mentoring) is hard to measure relative to sales revenue. Within this context, our measure can be read as saying, for example, "to rationalize these wages as monotone, we must believe that the best organizational citizenship term $(\max_i\{\eta_i\})$ counts twice as much ($\underline{\rho}=2$) as the most important observable metric $(\max_k\{\beta_k\})$."

Example 3. Consider the same three-worker firm structure as in Example 2, whose original system of equation returns an empty set (because of worker 2 is less productive than worker three on all dimensions, but is paid more). As a measure of the "distance" to monotonicity, we consider the revised system of inequalities adding unobserved terms η_i, and a new constraint on these terms.

$$\vec{\theta}^1\cdot\vec{\beta}+\eta_1>\vec{\theta}^2\cdot\vec{\beta}+\eta_2 \qquad \text{(1's productivity is greater than 2's)} \qquad (12)$$

$$\vec{\theta}^1\cdot\vec{\beta}+\eta_1>\vec{\theta}^3\cdot\vec{\beta}+\eta_3 \qquad \text{(1's productivity is greater than 3's)} \qquad (13)$$

$$\vec{\theta}^2\cdot\vec{\beta}+\eta_2>\vec{\theta}^3\cdot\vec{\beta}+\eta_3 \qquad \text{(2's productivity is greater than 3's)} \qquad (14)$$

$$\beta_k \quad \forall k\geq 0 \qquad \text{(Positive } \beta_k\text{s)} \qquad (15)$$

$$\eta_i \quad \forall i\geq 0 \qquad \text{(Positive } \eta_i\text{s)} \qquad (16)$$

$$\rho\cdot\max\{\beta_k\}\geq\max\{\eta_i\} \qquad \text{(Constraint on Maximum } \eta_i\text{)} \qquad (17)$$

The smallest ρ for which the revised system has a non-empty solution is $\underline{\rho}$, our second way to measure distance from monotonicity.

Whether a given value of $\underline{\rho}$ is implausibly large depends on the context. Even with the cap on η_i, allowing unobserved productivity to rationalize the wage schedule is generous to the firm: it assumes that the employer has perfectly allocated monetary credit to the workers who are unobservably productive. It is possible that an "overpaid" worker does nothing to deserve this extra payment, yet the procedure above still allows his pay to be rationalized by a hypothetical (unobserved) η_i. In this sense, our results based on unobserved productivity are likely to be conservative, in the sense of providing a lower bound on the extent of non-monotonicity.

Other constraints on unobservables. There are many ways to constrain the size of the η_i terms besides the max-based cap above.[21] The appropriate constraint is inherently context-dependent and reflects which relationships between the η_i and the β_k are economically implausible. As discussed above, a wide variety of such constraints can be accommodated within our solution concept as long as they can be expressed as linear constraints. In Appendix B2, we show that after normalizing by the largest observable weight, the max-based cap can be handled by solving a small family of linear programs, one for each candidate maximizer of β_k.

The η_i constraints can also be combined with the sign constraints in page 27. In some applications, sign constraints are essential to meaningfully restrict the system. For example, the cap $\rho \cdot \max\{\beta_k\} > \max\{\eta_i\}$ would be trivially satisfied if all η_i were allowed to be negative while all β_k were positive. For this reason, Example 3 includes additional constraints requiring all β_k and η_i to be weakly positive.

4.3 Agreement between Distance Measures

Although $\underline{\delta}$ and $\underline{\rho}$ are both "distances to monotonicity," they take different approaches, use different units, are not comparable, and are not guaranteed to agree qualitatively.[22] As an example, consider a two-worker firm structure where one worker is paid \$100,000 more, but has only slightly lower productivity, than the other. To "fix" this lack of monotonicity by

[21] For example, one could constrain the range of the η_i, rather than their maximum.

[22] In Section 3, we discuss further qualitative differences between fairness rules, and how they can "disagree."

dropping pay comparisons requires ignoring a relatively large gap ($\underline{\delta} = \$100K$). But to fix the issue by adding unobserved productivity, we only need to assume the higher-paid worker has modest unobserved contributions (low $\underline{\rho}$).

Similarly, consider a two-worker firm structure where one worker is paid a mere \$1 more, but has much lower productivity, than the other. To "fix" this lack of monotonicity by dropping pay comparisons requires ignoring a small gap ($\underline{\delta} = \$1$). But to fix the issue by adding unobserved productivity, we would need to assume an enormous difference in unobserved contributions (a high $\underline{\rho}$).

5 Empirical Setting and Data

To operationalize these methods and ideas, we now turn to a sample of researchers at top public universities in the United States. For these workers, salaries and output are largely observable. Salaries are often available on state level government websites, and research productivity can be measured with publicly-available publication and citation data. University websites contain rosters and job titles within departments, allowing us to consider the relative pay and productivity of workers who do the same "job."

As in our methodological setup above, productivity is multidimensional. In particular, each worker has a different mix of publications in various journals, and different journals could have different productivity weights. Like our methodological setup, there may be unobserved measures of performance such as departmental service and/or teaching. We can use the methods above to recover the minimum bounds needed to constraint the role of these unobservables.

Appendix C details how our data was assembled and merged. The availability of salary data affected the scope of our sample (both for dates and geographies). For research productivity outcomes, we used the data carefully assembled by Card et al. (2022).[23] We follow Card et al. (2022) in studying the disciplines of mathematics, psychology, and economics. For these

[23]The Card et al. (2022) replication package can be found at https://github.com/sdellavi/GenderGapsAcademies. We are grateful to Card, DellaVigna, Funk and Iriberri for making this data available.

disciplines, Card et al. (2022) identified the highest-impact journals in each field, and assembled an annual panel of publications by all authors in all these journals.

We augment the Card et al. (2022) data with each author's cumulative total citations across all publications (per year), using data from Scopus API from 1996-present.[24] This includes citations to papers not covered by the Card et al. (2022) data. Appendix C contains the full details of this dataset. Our final dataset consists of an author×year panel containing each authors publications and citations per year.

High-level academic service commitments (such as departmental chair terms and administrative roles like dean and vice-dean terms) are observable at the researcher-year level through titles. We exclude these workers from our tests, as their job expectations are likely different and unique. Insofar as we cannot fully remove additional significant service expectations, this is likely to affect only the "full professor" cells in our data. Our salary data contains total compensation paid to each worker, with universities in some states separately reporting base compensation and additional payments such as summer salary. We therefore consider total compensation for our analysis, but our qualitative results are broadly robust to investigating only base pay where available.

Table 3 summarizes the sample. We cover over 3,000 professors in 32 universities across 13 states over 12 years. The sample includes about 35,000 publications across 42 total journals. 55% of our observations are for full professors, 23% are for associate professors, and 21% are for assistant professors. In terms of coverage, 96% of researcher×year observations feature positive, non-zero productivity. By this, we mean that for at least one productivity dimension, the researcher had non-zero value (including cumulative journal articles and cumulative citations).

Our sample is not completely balanced because salary data is not available for some universities in early years of our sample. On average our data covers about eight years per university, and we track each professor for on average 5.6 years. On average, each professor generates 2.12 new publications per year, although this number is only 0.18 per year in the Card et al. (2022) top journals. However, these averages are dragged up by outliers, and the median observation

[24] https://dev.elsevier.com/

contains one publication per year and zero top publications.

Table 3: **Summary of Sample**

Variable	Count	Note
Professors	3,014	
U.S. States	13	
Universities	32	
Years	12	2008-2019
Departments	91	Avg. ≈3 departments/university
Average Years/School	8.03	
Average Years/Department	7.85	
Average Years/Professor	5.60	
Total "Top" Journals	42	from Card et al. (2022).
- Economics	14	
- Mathematics	15	
- Psychology	13	
Publications	35,991	
Citations	2,408,264	
Total Professor × Years	16,951	
Total Department × Years	715	

Notes: This table contains summary statistics from our sample.

In our raw data, we have annual data containing researcher's publication count for each of the 42 top journals in Card et al. (2022) by year. We also have a single variable for the number of total Card et al. (2022) publications, as well as another variable for the number of total *non*-Card et al. (2022) publications. We also have data about new citations (annually). The raw data above contain *new* publications and citations each year. As we describe next, we augment this raw data by creating separate variables representing cumulative publication counts (by journal and overall), as well as cumulative citations.

5.1 Operationalizing the Method

Using the data above, we operationalize the empirical methodology described in Section 4. Overall, our approach aims to be flexible towards the institutions: When a question of interpretation arises, we have aimed to make it easier to satisfy monotonicity constraints. For this

reason, our implementation choices will further contribute the interpretation of our results as a *lower bound* on the level of monotonicity violations.

Below we review a series of such interpretations and how we resolved them. First, we allow "productivity" in a given year to be a weighted combination of current-year publications and citations, or cumulative publications and citations. For cumulative metrics, we start with the Card et al. (2022) start date of 1960. This creates a new set of new variables for each journal representing the authors' cumulative publications in that journal. We have done the same for citations. This allows schools to rationalize pay inequality based on historical or cumulative performance.

We implement the linear program described in Example 1 for each Title$\times$School$\times$Discipline$\times$Year cell in our data. We remove all cells with one member, as monotonicity requires at least one pairwise wage comparison. This approach is generous to universities; we do not require departments to have a) consistent weights across multiple years, or b) consistent weights for assessing junior vs senior scholars. We also do not impose c) discipline-wide constraints, even though consensus about journal rankings sometimes exists within fields. In our procedure, departments are allowed to reward professors based on idiosyncratic, department-specific journal rankings that change every year and are different for seniors and juniors.

We also examine monotonicity only within job titles, i.e., by comparing pay and productivity only among professors of the same rank (assistant, associate, or full) among those without administrative roles. We do not examine differences between (for example) associate professors and assistant professors, even though there could be equal pay violations arising from promoting an unproductive assistant professor to a higher paying associate title. We focus only on within-rank violations, as performance expectations on non-research dimensions might change across ranks.

We have also restricted the weights placed on publications and citations to be weakly positive. This means that an additional publication or citation cannot harm productivity. To avoid the case where all β_k weights are zero (i.e., all workers mechanically must have productivity zero), we

require that at least one weight be strictly positive.[25] When we operationalize the "unobserved productivity" approach (Section 4.2, the η_is), we also constraint all η_is to be weakly positive. We also drop the first and last years of each researcher's time at a given university, as the salaries during these potential transition periods can be complicated by (say) half a year's pay.

Distances from Monotonicity. Our methodology specifies two approaches for distances from monotonicity. Both approaches involve running our procedure iteratively, tolerating increasingly large monotonicity violations, and reporting a threshold of tolerance under which each cell becomes rationalizable. The first (Section 4.1) involves dropping minimal salary differences. To operationalize this method, we examine salary differences between zero and the maximum pay difference between two workers in each cell. We conduct a binary search over this interval to find the smallest gap that can be dropped in order to rationalize the wages.[26]

In the second approach (Section 4.2), we tolerate increasingly high levels of performance differences on unobservable metrics (such as teaching quality) and observable metrics. To operationalize this, we impose a cap on the ratio between i) the largest unobserved term, $\max_i \eta_i$, and ii) the largest observable weight, $\max_k \beta_k$, where the latter may correspond to a journal-publication measure, a cumulative publication measure, or citations. Colloquially, this asks how important the strongest unobserved contribution would need to be relative to the most heavily weighted observable metric in the cell. As with our earlier problem, we conduct a binary search to find the smallest such cap.[27]

Gender Rationalizability. We also implement gender-restricted EPEW requirements which limit our system of inequalities only to pairs of different-gendered professors within the same

[25] We accomplish this by adding a constraint to the linear program requiring that $\sum_k \beta_k = 1$, and $\beta_k \geq 0$, $\forall k$. This means that at least one weight must be strictly positive.

[26] A binary search over an interval is an efficient method for finding the point where a property first becomes true within a sorted range. Starting with a lower and upper bound, the algorithm tests the midpoint of the interval. If the condition holds at the midpoint, the upper bound is moved down; if it fails, the lower bound is moved up. This halving process repeats until the interval narrows to the smallest value where the condition is satisfied.

[27] We programmed our script to search as high as $\rho = 10{,}000$, but for most cells the binary search terminated between zero and 1.

cell. We code each researcher's gender based on either a) the coding in Card et al. (2022), or b) an imputed gender using names (using the Genderize.io API, described in in Appendix C2.3). We then proceed with our rationalizability test using the "pared-down" set of inequalities (if all workers within a cell have the same gender, we code this as satisfying gender rationality.) We present formal details in Appendix B3.

5.2 Outcome Variables

We examine outcomes y_{sdjt} for schools s (one of 32 public universities), disciplines d (math, economics, or psychology), jobs j (Assistant, Associate or Full), at times t (years 2008-2019). The three outcomes we use for y_{sdjt} are:

1. "**Rationalizability**" — a binary representing if the set of β weights rationalizing the pay distribution as monotonic is *non-empty*.
2. "**Dollar Distance**" from Rationalizability — a number representing the result of our "Dropping Minimal Salary Differences" analysis (Section 4.1).
3. "**Unobservable Cap Distance**" (Relative to Max Observable Weight) — a number representing the smallest cap $\underline{\rho}$ such that $\max_i \eta_i \leq \underline{\rho} \max_k \beta_k$ can rationalize the observed pay ordering, i.e. the result of our "Unobserved productivity" analysis (Section 4.2).

Throughout, there is only one unobservables-based outcome in the paper, namely $\underline{\rho}$ defined relative to $\max_k \beta_k$.

5.3 Specifications

Our empirical analysis broadly takes one of two types. The first is descriptive, and the second uses shift–share IV evidence that we interpret suggestively. These use distinct specifications and empirical strategies, described below.

Descriptive. What types of departments and/or job categories exhibit high degrees of non-monotonicity? For this analysis, we run the following descriptive regressions.

$$y_{sdjt} = \gamma_s + \lambda_t + \alpha_d + \omega_j + \epsilon_{sdjt}, \tag{18}$$

where γ_s are school fixed effects, λ_t are year fixed effects, α_d are discipline fixed effects, and ω_j are job role fixed effects. We report the coefficients for discipline (omitted category = math) and job title (omitted category = assistant professor) to describe differences across academic fields and seniority.

Our baseline descriptive regressions use school, year, discipline, and rank fixed effects. In Appendix D2, we show that adding flexible controls for cell size and the number of journals in the cell attenuates, but does not overturn the main patterns. Such attenuation is consistent with Proposition 2, because of the mechanical negative impact of a large cell size on rationalizability.

Shift–Share Evidence. Because the size of a cell is endogenous, we develop a shift-share IV strategy. Our theoretical results (Proposition 2) noted the difficulty of maintaining monotonicity as a company becomes larger. However to measure this this relationship empirically, there are several possible threats to exogeneity. For example, teams with monotonicity violations might shrink as discouraged workers exit. Alternatively, larger teams might develop specialized HR functions, and transparency protocols due to scale economies. This institutional investment in fairness might lower inequality, and possibly overstate the vulnerability of larger cells to monotonicity violations.

Because the size of a cell is endogenous, we develop a shift-share IV strategy. Our second-stage IV equation is:

$$y_{sdjt} = \beta_1 \text{Cell Size}_{sdjt} + \gamma_s + \lambda_t + \alpha_d + \omega_j + \epsilon_{sdjt}, \tag{19}$$

where γ_s are school fixed effects, λ_t are year fixed effects, α_d are discipline fixed effects, and ω_j are job role fixed effects. The coefficient on Cell Size_{sdjt} is the variable of interest.

We call our instrument B_{sjt}, and it operates at the school×title×year level. As is common for Bartik-like exposure designs, our instrument combines a baseline exposure term ("shares") and a time-varying term ("shifts"). Let z_{sdj0} denote the baseline share of faculty in school s with discipline d and title j (computed in the first observed year of that $s \times d \times j$ cell), and let g_{st} denote the cumulative number of new external hires at school s through year t. Using the notation of Goldsmith-Pinkham et al. (2020), the instrument we construct is

$$B_{sjt} = \sum_{d=1}^{D} z_{sdj0} g_{st}. \tag{20}$$

Because g_{st} does not vary by d, this simplifies to $B_{sjt} = z_{sj0} g_{st}$ where $z_{sj0} := \sum_{d=1}^{D} z_{sdj0}$ is the baseline share of title j within school s. Within each school, the baseline shares satisfy $\sum_d \sum_j z_{sdj0} = 1$. Equation 20 is therefore a school×title×year instrument. Discipline enters our IV specification through the outcome index y_{sdjt} and the fixed effects, but the instrument itself varies only at the $s \times j \times t$ level. Standard errors for the regression are clustered both at the school×year and the discipline×title levels.

In our use of this instrument, the relevant variation comes from school-level cumulative external hiring g_{st} interacted with baseline shares z_{sdj0}, following the shift–share logic of Borusyak et al. (2022); Adao et al. (2019, "many shocks"). This approach allows the baseline shares z_{sdj0} to be correlated with time-invariant determinants of outcomes absorbed by the fixed effects; conditional on school, year, discipline, and title fixed effects, the remaining variation comes from school-level shifts g_{st} that load differentially across cells through these predetermined shares.

The school-level "shifts" g_{st} may reflect expansions or contractions in university resources arising from state budgeting decisions, alumni donation patterns, or other idiosyncratic school-specific factors. At the same time, because g_{st} is school-level and cumulative, such shifts could in principle affect pay-setting through channels other than cell size. We therefore interpret the IV estimates as suggestive evidence on the scale–fairness link, rather than as a clean estimate of a purely external hiring shock. The school-based shifts g_{st} in our setup have a mean of 92 and

a standard deviation of 49.[28] This implies that, at a given point in the sample, a typical school has accumulated about 92 external hires since the start of the sample.

Qualitatively, our instrument tells us how strongly title j at school s is exposed to school-level hiring shocks at time t, scaled by the baseline share of that title within the school. The mean value of the instrument is 33 (median 26), with a standard deviation of 25. If the instrument equals 33 for school s and title j at time t, this means that $z_{sj0} \times g_{st} = 33$, i.e. cumulative hires at school s through year t, scaled by the baseline title share z_{sj0}, equals 33. As such, a "high" instrument corresponds to a higher group size. Our first-stage regression is

$$\text{Cell Size}_{sdjt} = \delta_1 B_{sjt} + \gamma_s + \lambda_t + \alpha_d + \omega_j + u_{sdjt}. \tag{21}$$

Table 4 presents the first stage regressions (both with fixed effects and without). Our instrument has a strong first stage, featuring strong F statistics (as measured using the metrics proposed by Montiel-Olea and Pflueger 2013 and Stock and Yogo 2005; Kleibergen and Paap 2006). A one standard deviation shock raises the size of a cell by about 4-5 researchers.

Table 4: **Shift-Share First Stage**

	(1) Cell Size (Faculty)	(2) Cell Size (Faculty)
Instrument (Standardized)	4.7***	4.3***
	(1.1)	(.8)
Fixed Effects	-	Y
F-Statistic	17	28
Observations	1,926	1,926
R^2	.42	.6

Notes: This table reports 2SLS estimates relating cell size to pay-fairness outcomes, using the school×title×year shift–share predictor B_{sjt} as an instrument for the number of research faculty. Our instrument has been standardized in order to ease interpretation. In Column 2 fixed effects include school, year, discipline and job title are included.

[28] In the Stata implementation, g_{st} is the cumulative number of new external hires at school s through year t. It is therefore best interpreted as cumulative external hiring intensity rather than as net headcount growth by the school.

6 Empirical Results

When we operationalize our technique on our data, we find 61% of the cells in our data have pay structures that cannot be rationalized using the inequalities. Stated differently, in 61% of our cells, there are no set of weights on journal publications and citations (both current and cumulative) that would rationalize the observed pay as monotonic in productivity metrics. When we analyze distances from monotonicity, the average non-monotonic cell requires differences of about $48,000 to be ignored. Non-monotonic cells can be rationalized if the largest unobserved term is about 30% of the most heavily weighted observable metric on average (under the assumption that credit for unobserved performance was allocated to the right person).

Table 5 examines differences by seniority. The pay of full professors is least rationalizable by all measures demonstrated in column 1. The difference in rank is largely explained by the fact that full-professor cells are substantially larger (average 14 members vs 6 for assistant and associate). When we add flexible controls for cell size and the number of journals (Appendix Table D4), the gaps by rank shrink considerably. This is consistent with the scale–fairness mechanism formalized in Proposition 2.

One possible explanation for the non-rationalizability of full-professor wages is that senior faculty perform more service or other unobserved tasks. This issue is least likely to affect assistant professors (despite this, we find that ≈56% of assistant professor cells cannot be rationalized). Our distance-to-monotonicity measures allow us to directly assess this explanation for full professor non-monotonicity. If unobserved productivity were the primary driver, we would expect full-professor cells to require substantially larger allowances for latent performance. Instead, we find the opposite pattern: despite being more frequently non-rationalizable overall, full-professor cells require a *smaller cap* on unobserved productivity than assistant-professor cells to restore monotonicity (Column 2).

By contrast, the sharp difference across ranks appears on the pay margin. To rationalize full-professor wage schedules by relaxing pay comparisons, we would need to ignore roughly

$32,000 more in salary differences than for assistant professors (Column 3). This divergence suggests that the dominant source of non-rationalizability among senior faculty is not hidden productivity, but larger and more dispersed pay gaps.[29]

Table 5: **Differences by Seniority**

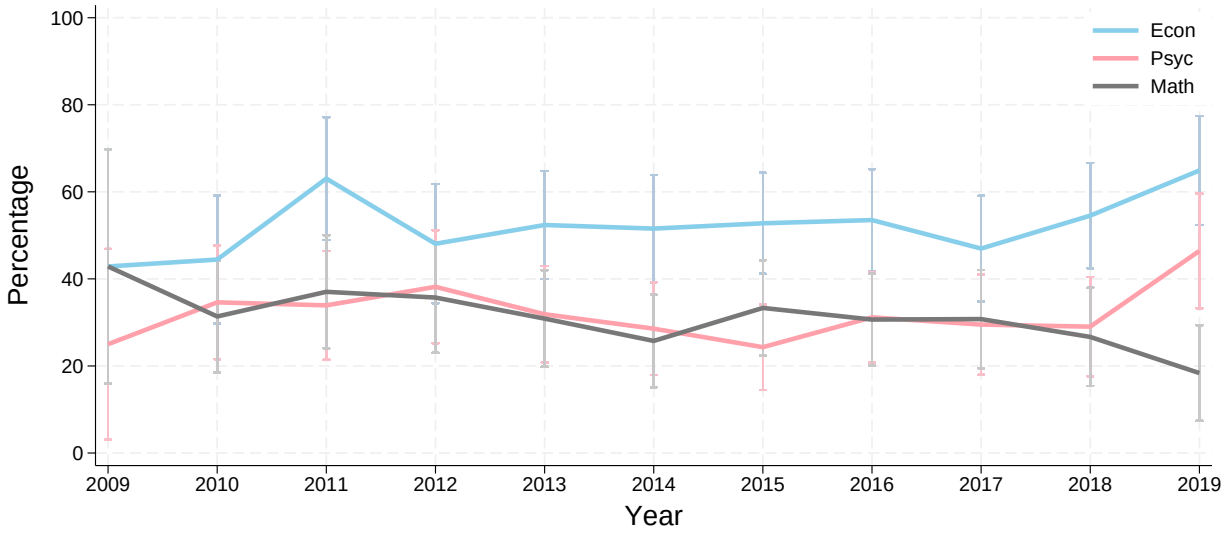
	(1) Rational -izable?	(2) Unobservable Cap ($\underline{\rho}$) Distance	(3) Dollar ($\underline{\delta}$) Distance
Full Professor	-.19***	-.093**	32,780***
	(.047)	(.037)	(6,171)
Associate	.052	.056	-4,407*
	(.038)	(.095)	(2,518)
Assistant Mean	.44	.21	16921.64
Observations	1,926	1,926	1,755
R^2	.15	.03	.32

Notes: This table reports the results of the descriptive Equation 18, which examines how pay rationalizability varies by cell characteristics. Assistant professors are the omitted category. Columns 2 and 3 include the full sample, including cells that could be rationalized. For these cells, $\underline{\rho} = \underline{\delta} = 0$. The outcome measures are defined in Section 5.2. All specifications include school, discipline, year, and title fixed effects. Standard errors clustered both at the school×year and the discipline×title levels (as shown in parentheses). Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In the appendix, we investigate differences over time and by discipline. Table D2 shows the most rationalizable firm structures are in economics, both in overall rationalizability and our distances from monotonicity metrics. Mathematics and psychology are approximately tied. We inspect this in more detail in Table D5; this finding is in part caused by the fact that economics cells are smaller (average of 6 researchers) than cells in math and psychology (averages of 11 and 9, respectively), connecting to Proposition 2. Tables D3 and D6 suggest that there has been little change in rationalizability over the period of our analysis; if anything, the distance to monotonicity has increased over time. Figure 3 summarizes these findings visually.

[29] A careful reader may note that we are finding a "disagreement" between the two "distances to monotonicity" approaches, of the form described in Section 4.3.

Figure 3: **Percentage Rationalizable Over Time (by Discipline)**



Notes: This figure plots the percentage of school× title×year cells whose pay structure can be rationalized as monotonic over time.

Heterogeneous Equal Pay Laws. While we have presented evidence that many firms likely do not comply with the equal pay principle across all workers, it is possible that this "non-compliance" stems from firms complying with only the weaker state-level restrictions. These impose equal pay restrictions only for workers of different demographic groups. Because of limited temporal variation in equal pay restrictions within state during our data window, we lack a clean natural experiment to answer this question.

However, we can study this hypothesis by seeing whether firms comply with their state-level equal pay restriction using the gender-specific version of our method in Section B3. This shrinks the the system of inequalities used to test monotonicity by removing pairs of the same gender. In Table 6 we report some basic descriptive evidence. 86% of cells were in states covered by a gender-specific EPEW, wherein the restriction applies only for pairs of workers of different genders. The remaining states de facto impose the monotonicity restriction across all pairs of workers.[30] Table 6 investigates whether firm structures are rationalizable considering all worker

[30]That is, the remaining states also impose the equal pay restriction across pairs of workers from different

comparisons and only cross-gender comparisons, by whether or not the state in question has gender-only EPEW. Universities in states where all comparisons are included are slightly more likely to be rationalizable, both overall and for the gender-rationalizability metric (described in Section B3).

Table 6: **Rationalizability by Equal Pay Laws**

Type of Pay Law	**Overall Rationalizable**	**Gender Rationalizable**
All Comparisons	42%	58%
Only Gender Inequality	38%	54%

Notes: Table 6 reports the share of school×discipline× rank×year cells whose pay structures can be rationalized as monotonic in observed productivity ("Overall") and under the gender-restricted test ("Gender"), stratified by the type of equal-pay law in each state. "All Comparisons" denotes pay-equity laws covering every pairing of protected classes; "Only Gender Inequality" denotes laws covering gender only.

6.1 Suggestive Shift–Share Evidence on Scale–Fairness Trade-Offs

Our final empirical exercise studies whether the difficulty of maintaining a monotone wage schedule is linked to the size of the decision-making unit. As described above, we instrument cell size with the school×title×year predictor $B_{sjt} = z_{sj0} g_{st}$, which interacts common school-level cumulative staffing growth with baseline title shares. We interpret this section as suggestive evidence consistent with Proposition 2 rather than as a stand-alone clean external-shock design.

Table 7 reports the corresponding IV estimates for the wage-fairness outcomes. Columns (1)–(2) show that each additional faculty member in a cell lowers the probability that pay can be rationalized as monotone by 3.1–3.6 percentage points. Columns (3)–(4) show that the implied effect on the unobservable-cap outcome $\bar{\rho}$ is small without fixed effects and about 0.02 with fixed effects. Columns (5)–(6) indicate that each additional faculty member raises the salary-gap tolerance needed to restore monotonicity by about \$4.4–\$4.9 thousand.

demographic groups, but define demographic groups as the intersection of on average 8 different characteristics (e.g. race, gender, religion, country of origin, etc). Even were each characteristic binary, we would anticipate $2^8 = 256$ different demographic groups. Given the comparatively small size of each cell in our data, it is likely that very few workers fall into the same demographic group, which implies that legal compliance is approximated by our standard monotonicity test.

Relative to a median assistant-professor salary of $92k, this represents a 4%–5% widening of the acceptable wage band after a single standard-deviation staffing increase. Under the IV magnitudes in Table 7, moving from an eight-person to a twelve-person faculty group corresponds to a 12–14 percentage-point lower probability of monotone compliance and a roughly $18,000–$20,000 larger salary gap that must be ignored to restore monotonicity.

Table 7: **Shift Share Results**

	(1) Rational -izable?	(2) Rational -izable?	(3) Unobservable Cap ($\underline{\rho}$) Distance	(4) Unobservable Cap ($\underline{\rho}$) Distance	(5) Dollar ($\underline{\delta}$) Distance	(6) Dollar ($\underline{\delta}$) Distance
Cell Size (Faculty)	-.031***	-.036**	-.0024	.02*	4,446***	4,905***
	(.005)	(.014)	(.005)	(.0095)	(410)	(714)
Fixed Effects	-	Y	-	Y	-	Y
F-Statistic	17	28	17	28	17	30
Observations	1,926	1,926	1,926	1,926	1,755	1,755

Notes: This table reports 2SLS estimates of the effect of department size on pay-fairness outcomes, instrumenting the number of research faculty with a discipline–rank–year shift–share predictor of predicted hires. The first stage is Equation 21, the second stage is Equation 19, and the IV definition is Equation 20. Significance levels: * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

The IV estimates are consistent with a complexity cost: as departments grow, administrators face a steeper trade-off between rewarding multidimensional performance and preserving a monotonic link between pay and output. In settings where expansion is inevitable—new degree programs, mergers, or mission-driven hiring sprees—routine monotonicity auditing and simplified evaluation rubrics may be a useful guardrail.

Our results about academic are intended as proof-of-concept demonstration of our method in a real world setting. However, our results suggest that monotonicity violations may be common and relatively large, and can rationalized only by large, well-placed weights on "unobservable" dimensions of productivity. Our results also provide suggestive evidence on the scale/fairness tradeoff discussed in Proposition 2.

7 Discussion and Conclusion

This paper offers a unified theoretical and empirical treatment of a firm's relative wages according to familiar notions of equality in the law, culture, and prior literature. We show that US-style EPEW laws which operate at the individual level impose a restriction that pay is monotonic in productivity.

Monotonicity is elegant but demanding, and is harder to satisfy as the firm grows in size. There are additional measurement challenges in determining a composite measure of productivity when output is multidimensional. To solve this issue, we take a conservative approach. We develop a linear-programming framework that asks whether there exists any non-negative weighting of the observed productivity metrics that can rationalize the observed pay as monotone. When no such vector of weights exists we derive two continuous "distance-to-fairness" measures: (i) the minimum wage gap that must be ignored to restore monotonicity and (ii) the minimum importance that must be assigned to an unobserved performance dimension to restore monotonicity. Both metrics are portable to other settings.

Applying the method to 3,000 faculty members in mathematics, economics, and psychology at 32 U.S. public universities (2008–2019) yields four main results :

1. **Prevalence of violations.** Roughly two-thirds of *school* $\times$ *discipline* $\times$ *rank* $\times$ *year* cells fail monotonicity. In the median failing cell, a gap of \$48,000 must be ignored before a monotone rationalization becomes possible.

2. **Role of unobservables.** Alternatively, the largest unobserved term would need to equal $\underline{\rho} \approx 30\%$ of the most heavily weighted observable metric — an implausibly large premium in many disciplines.

3. **Heterogeneity across disciplines and ranks.** Economics departments exhibit the fewest violations, owing partly to smaller group size and a more concentrated journal set. Full professors show the largest shortfalls, suggesting that non-research duties or opaque

retention offers may complicate compliance.

4. **Firm size and compliance.** Suggestive shift–share evidence is consistent with larger cells being less likely to admit a monotone pay structure and requiring a larger wage-gap tolerance to restore monotonicity. In the IV estimates, each additional faculty member is associated with about a 3–4 percentage-point lower probability of monotone compliance and about \$4.4–\$4.9 thousand more salary gap that must be ignored to restore monotonicity. indicates that adding four researchers (one standard deviation) reduces the probability of a compliant pay structure by 3–4 percentage points and widens the necessary wage-gap tolerance by \$6–7,000. Scale therefore magnifies the challenge posed by monotonicity.

Like any audit study, our exercise relies on the quality and breadth of the productivity metrics at hand; teaching and service are difficult to observe, and our generous treatment of them already favors compliance. The sample is also confined to research universities, a labor market where output is measured unusually well. Extending the methodology to settings with noisier productivity signals, richer nonlinear reward systems, or dynamic renegotiation of pay are promising avenues for future research.

From a normative perspective, we study alternative relative wage restrictions. Existing equal-pay statutes typically police EPEW but not EWEP. As EWEP is a common fairness notion, this gap leaves firms room to satisfy the letter of the law while violating an intuitive notion of individual fairness. More broadly, we discuss why it may be desirable to insist that fairness rules symmetrically treat wages and productivity.

However, individual-level EPEW restrictions "usually" satisfy EWEP, a point which we make both theoreically and empirically. In other words, EPEW nearly treats wage and productivity symmetrically. The same cannot be said about group-level, regression based approaches to EPEW and EWEP — theoretically and empirically, we show that these measures can differ, and interventions intended to increase pay fairness can potentially backfire.

Our results suggest three takeaways. First, group-level equal pay restrictions used in some countries may be less desirable that individual-level restrictions in the US. Second, firms may

wish to subject wage bills—not just hiring and promotion decisions—to internal monotonicity audits, especially after growth or mergers. Third, because multidimensional performance creates genuine ambiguity, transparency about the *weights* that tie pay to output can prevent post-hoc rationalizations and improve morale.

In sum, the *Equal Pay for Equal Work (for Equal Pay?)* paradox is real, quantifiable, and potentially widespread. Monotonicity offers a principled benchmark that aligns legal intuition with economic rigor, while our empirical framework shows how that benchmark can be studied and operationalized in practice.

References

Adao, Rodrigo, Michal Kolesár, and Eduardo Morales, "Shift-share designs: Theory and inference," *The Quarterly Journal of Economics*, 2019, *134* (4), 1949–2010.

Anderson, David, Margrét V Bjarnadóttir, Cristian L Dezső, and David Gaddis Ross, "On a firm's optimal response to pressure for gender pay equity," *Organization Science*, 2019, *30* (1), 214–231.

Angrist, Joshua D, "Estimating the labor market impact of voluntary military service using social security data on military applicants," *Econometrica*, 1998, *66* (2), 249–288.

Arnold, David, Will Dobbie, and Crystal S Yang, "Racial bias in bail decisions," *The Quarterly Journal of Economics*, 2018, *133* (4), 1885–1932.

Ayres, Ian, "Outcome tests of racial disparities in police practices," *Justice research and Policy*, 2002, *4* (1-2), 131–142.

Baker, Michael, Yosh Halberstam, Kory Kroft, Alexandre Mas, and Derek Messacar, "Pay transparency and the gender gap," *American Economic Journal: Applied Economics*, 2023, *15* (2), 157–183.

Bayer, Patrick and Kerwin Kofi Charles, "Divergent paths: A new perspective on earnings differences between black and white men since 1940," *The Quarterly Journal of Economics*, 2018, *133* (3), 1459–1501.

Becker, Gary S, *The economics of discrimination*, University of Chicago press, 1957.

_ , "Nobel lecture: The economic way of looking at behavior," *Journal of Political Economy*, 1993, *101* (3), 385–409.

Benson, Alan, Danielle Li, and Kelly Shue, "Promotions and the peter principle," *The Quarterly Journal of Economics*, 2019, *134* (4), 2085–2134.

Bergemann, Dirk and Stephen Morris, "Bayes correlated equilibrium and the comparison of information structures in games," *Theoretical Economics*, 2016, *11* (2), 487–522.

_ and _ , "Information design: A unified perspective," *Journal of Economic Literature*, 2019, *57* (1), 44–95.

_ , Benjamin Brooks, and Stephen Morris, "Counterfactuals with latent information," *American Economic Review*, 2022, *112* (1), 343–368.

Bertrand, Marianne and Sendhil Mullainathan, "Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination," *American Economic Review*, 2004, *94* (4), 991–1013.

_ , Claudia Goldin, and Lawrence F Katz, "Dynamics of the gender gap for young professionals in the financial and corporate sectors," *American economic journal: applied economics*, 2010, *2* (3), 228–255.

Blau, Francine D and Lawrence M Kahn, "The gender wage gap: Extent, trends, and explanations," *Journal of economic literature*, 2017, *55* (3), 789–865.

Bohren, J Aislinn, Kareem Haggag, Alex Imas, and Devin G Pope, "Inaccurate statistical discrimination: An identification problem," *Review of Economics and Statistics*, 2023, pp. 1–45.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel, "Quasi-experimental shift-share research designs," *The Review of economic studies*, 2022, *89* (1), 181–213.

Budig, Michelle J and Paula England, "The wage penalty for motherhood," *American sociological review*, 2001, *66* (2), 204–225.

Canay, Ivan A, Magne Mogstad, and Jack Mountjoy, "On the use of outcome tests for detecting bias in decision making," *Review of Economic Studies*, 2023, p. rdad082.

Card, David, Stefano DellaVigna, Patricia Funk, and Nagore Iriberri, "Gender differences in peer recognition by economists," *Econometrica*, 2022, *90* (5), 1937–1971.

Carpenter, Christopher S, "Revisiting the income penalty for behaviorally gay men: Evidence from NHANES III," *Labour economics*, 2007, *14* (1), 25–34.

Case, Anne and Christina Paxson, "Stature and status: Height, ability, and labor market outcomes," *Journal of political Economy*, 2008, *116* (3), 499–532.

Collins, Patricia Hill, "Black feminist thought," *Knowledge, Cons*, 1990.

Cook, Cody, Rebecca Diamond, Jonathan V Hall, John A List, and Paul Oyer, "The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers," *The Review of Economic Studies*, 2021, *88* (5), 2210–2238.

Cowgill, Bo, Fabrizio Dell'Acqua, Samuel Deng, Daniel Hsu, Nakul Verma, and Augustin Chaintreau, "Biased programmers? Or biased data? A field experiment in operationalizing AI ethics," in "Proceedings of the 21st ACM Conference on Economics and Computation" 2020, pp. 679–681.

Cullen, Zoë and Bobak Pakzad-Hurson, "Equilibrium Effects of Pay Transparency," *Econometrica*, 2023, *91* (3), 765–802.

_ , _ , and Ricardo Perez-Truglia, "Pushing the Envelope: The Effects of Salary Negotiations," Working Paper 33903, National Bureau of Economic Research 2025.

Dale, Stacy Berg and Alan B Krueger, "Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables," *The Quarterly Journal of Economics*, 2002, *117* (4), 1491–1527.

Dobbie, Will, Andres Liberman, Daniel Paravisini, and Vikram Pathania, "Measuring bias in consumer lending," *The Review of Economic Studies*, 2021, *88* (6), 2799–2832.

Garfinkel, Jon A, Mosab Hammoudeh, and James Weston, "Academic publishing behavior and pay across business fields," *Financial management*, 2024, *53* (1), 31–58.

Gibson, John, David L Anderson, and John Tressler, "Citations or journal quality: Which is rewarded more in the academic labor market?," *Economic Inquiry*, 2017, *55* (4), 1945–1965.

Ginther, Donna K and Kathy J Hayes, "Gender differences in salary and promotion in the humanities," *American Economic Review*, 1999, *89* (2), 397–402.

Goldin, Claudia, "A grand gender convergence: Its last chapter," *American economic review*, 2014, *104* (4), 1091–1119.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift, "Bartik instruments: What, when, why, and how," *American Economic Review*, 2020, *110* (8), 2586–2624.

_ , Peter Hull, and Michal Kolesár, "Contamination bias in linear regressions," Technical Report, National Bureau of Economic Research 2024.

Gomez-Mejia, Luis R and David B Balkin, "Determinants of faculty pay: An agency theory perspective," *Academy of Management journal*, 1992, *35* (5), 921–955.

Gordon, Nancy M, Thomas E Morton, and Ina C Braden, "Faculty salaries: Is there discrimination by sex, race, and discipline?," *The American Economic Review*, 1974, *64* (3), 419–427.

Gualdani, Cristina and Shruti Sinha, "Identification in discrete choice models with imperfect information," *arXiv preprint arXiv:1911.04529*, 2019.

Halac, Marina, Elliot Lipnowski, and Daniel Rappoport, "Rank Uncertainty in Organizations," *American Economic Review*, 2021, *111* (3), 757–86.

Hamermesh, Daniel S and Jeff Biddle, "Beauty and the labor market," 1993.

Harris, Richard and Mariluz Maté-Sánchez-Val, "Gender pay and productivity in UK universities: Evidence from research-intensive Business Schools," *Economics Letters*, 2022, *218*, 110738.

Hilmer, Michael J, Michael R Ransom, and Christiana E Hilmer, "Fame and the fortune of academic economists: How the market rewards influential research in economics," *Southern Economic Journal*, 2015, *82* (2), 430–452.

Hirsch, Barry T and Karen Leppel, "Sex discrimination in faculty salaries: Evidence from a historically women's university," *The American Economic Review*, 1982, *72* (4), 829–835.

Hoffman, Emily P, "Faculty salaries: Is there discrimination by sex, race, and discipline? Additional evidence," *The American Economic Review*, 1976, *66* (1), 196–198.

Holmstrom, Bengt and Paul Milgrom, "Multitask principal–agent analyses: Incentive contracts, asset ownership, and job design," *The Journal of Law, Economics, and Organization*, 1991, *7* (special_issue), 24–52.

Huang, Ruidi, Erik J Mayer, and Darius P Miller, "Gender bias in promotions: Evidence from financial institutions," *The Review of Financial Studies*, 2024, *37* (5), 1685–1728.

Hull, Peter, "What marginal outcome tests can tell us about racially biased decision-making," Technical Report, National Bureau of Economic Research 2021.

Jr, Arthur M Diamond, "What is a citation worth?," *Journal of Human Resources*, 1986, *21* (2), 200–215.

Katz, David A, "Faculty salaries, promotions, and productivity at a large university," *The American Economic Review*, 1973, *63* (3), 469–477.

Kim, Lanu, Bas Hofstra, and Sebastian Munoz-Najar Galvez, "A persistent gender pay gap among faculty in a public university system," *Scientific Reports*, 2024, *14* (1), 22212.

Kleibergen, Frank and Richard Paap, "Generalized reduced rank tests using the singular value decomposition," *Journal of Econometrics*, 2006, *133* (1), 97–126.

Li, Diyi and Cory Koedel, "Representation and salary gaps by race-ethnicity and gender at selective public universities," *Educational researcher*, 2017, *46* (7), 343–354.

Lyu, Yifei and Alexis Akira Toda, "Publications, citations, position, and compensation of economics professors," *Econ Journal Watch*, 2019, *16* (2), 239–257.

Magnolfi, Lorenzo and Camilla Roncoroni, "Estimation of discrete games with weak assumptions on information," *The Review of Economic Studies*, 2023, *90* (4), 2006–2041.

Mittal, Vikas, Lawrence Feick, and Feisal Murshed, "Publish and prosper: The financial impact of publishing by marketing faculty," *Marketing Science*, 2008, *27* (3), 430–442.

Montiel-Olea, José Luis and Carolin Pflueger, "A robust test for weak instruments," *Journal of Business & Economic Statistics*, 2013, *31* (3), 358–369.

Moore, William J, Robert J Newman, and Geoffrey K Turnbull, "Reputational capital and academic pay," *Economic Inquiry*, 2001, *39* (4), 663–671.

Mortensen, Dale T. and Christopher A. Pissarides, "Job Creation and Job Destruction in the Theory of Unemployment," *Review of Economic Studies*, 1994, *61* (3), 397–415.

Neal, Derek A and William R Johnson, "The role of premarket factors in black-white wage differences," *Journal of political Economy*, 1996, *104* (5), 869–895.

Obloj, Tomasz and Todd Zenger, "The influence of pay transparency on (gender) inequity, inequality and the performance basis of pay," *Nature Human Behaviour*, 2022, *6* (5), 646–655.

O'Keefe, Suzanne and Ta-Chen Wang, "Publishing pays: Economists' salaries reflect productivity," *The Social Science Journal*, 2013, *50* (1), 45–54.

Pager, Devah, "The mark of a criminal record," *American journal of sociology*, 2003, *108* (5), 937–975.

Passaro, Diego Gentile, Fuhito Kojima, and Bobak Pakzad-Hurson, "Equal Pay for *Similar* Work," *arXiv preprint arXiv:2306.17111*, 2024.

Peterson, Richard L, "An investigation of sex discrimination in commercial banks' direct consumer lending," *The Bell Journal of Economics*, 1981, pp. 547–561.

Postel-Vinay, Fabien and Jean-Marc Robin, "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity," *Econometrica*, 2002, *70* (6), 2295–2350.

Rambachan, Ashesh, "Identifying prediction mistakes in observational data," *The Quarterly Journal of Economics*, 2024, p. qjae013.

Sauer, Raymond D, "Estimates of the returns to quality and coauthorship in economic academia," *Journal of Political Economy*, 1988, *96* (4), 855–866.

Stock, James H and Motohiro Yogo, "Testing for Weak Instruments in Linear IV Regression," *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, 2005.

Swidler, Steve and Elizabeth Goldreyer, "The value of a finance journal publication," *The Journal of Finance*, 1998, *53* (1), 351–363.

Syrgkanis, Vasilis, Elie Tamer, and Juba Ziani, "Inference on auctions with weak assumptions on information," *arXiv preprint arXiv:1710.03830*, 2017.
The Pulse of Black America

The Pulse of Black America, *Technical Report* https://soba.iamempowered.com/sites/soba.iamempowered.com/files/PulseofBlackAmerica_4-10%20FINAL.pdf, *National Urban League/Benson Strategy Group 2022.*

Toutkoushian, Robert K, Marcia L Bellas, and John V Moore, *"The interaction effects of gender, race, and marital status on faculty salaries," The Journal of Higher Education, 2007, 78 (5), 572–601.*

Tuckman, Howard P and Jack Leahey, *"What is an article worth?," Journal of Political Economy, 1975, 83 (5), 951–967.*

Winter, Eyal, *"Incentives and Discrimination," American Economic Review, 2004, 94 (3), 764–773.*

Yildirim, Huseyin, *"Free Riding and Unequal Pay in Symmetric Teams," American Economic Journal: Microeconomics, 2025, 17 (3), 289–323.*

Appendix: For Online Publication Only

Table of Contents

A Proofs and Theoretical Extensions **a1**
- A1 Proof of Proposition 1 . . . a1
- A2 Proof of Proposition 2 . . . a3
- A3 Proof of Proposition 3 . . . a4
- A4 Proof of Proposition 4 . . . a6
- A5 Common limitations of EPEW restrictions . . . a8
- A6 New Group-Level Equality Measure . . . a11

B Methodology Details **a11**
- B1 Weak Pairwise Inequalities . . . a11
- B2 Minimal Strict Cap $\bar{\rho}$ via Charnes-Cooper and Lexicographic Search . . . a12
- B3 Gender Rationalizability . . . a17

C Dataset Construction **a18**
- C1 Salary Sample Selection . . . a18
- C2 Salary-Publication Merge . . . a19
  - C2.1 Data Preprocessing . . . a19
  - C2.2 Probabilistic Linkage . . . a20
  - C2.3 Gender Imputation . . . a21
- C3 Citation Record Inclusion . . . a21

D Additional Empirical Results **a22**
- D1 Main Specifications (without Controls) . . . a22
- D2 With Additional Controls . . . a23

E Group EPEW Symmetry Violations: Empirical Examples **a25**
- E1 Treatment Effects on Equal Work and Pay . . . a29

A Proofs and Theoretical Extensions

A1 Proof of Proposition 1

Proof. The desired result is implied by the following lemma.

Lemma A1. *Let S be a firm structure. The following statements are equivalent:*

(a) *The firm structure satisfies EPEW,*

(b) *For any two workers $i,j \in I$ (where without loss of generality $\theta_i \geq \theta_j$), $w_i \geq w_j$ with equality if $\theta_i = \theta_j$,*

(c) *there exists a non-decreasing function $\phi : [\min_{i \in I} \theta_i, \max_{i \in I} \theta_i] \to [0,1]$ such that any worker i of productivity θ receives wage $w_i = \phi(\theta)$.*

Proof of Lemma A1. We proceed by showing that **(a)** and **(b)** are equivalent, and then that items **(b)** and **(c)** are equivalent.

(a) $\Leftrightarrow$ (b) To see that **(a)** implies **(b)**, suppose the firm exhibits EPEW. We must show that for any i and j such that $\theta_i > \theta_j$ it is the case that $w_i \geq w_j$. Suppose for the sake of contradiction that for some i and j such that $\theta_i > \theta_j$ but that $w_i < w_j$. Then this violates EPEW, which is a contradiction.

To see that **(b)** implies **(a)**, suppose that for any two workers $i,j \in I$ (where without loss of generality $\theta_i \geq \theta_j$), $w_i \geq w_j$ with equality if $\theta_i = \theta_j$. If $\theta_i = \theta_j$, then by construction $w_i = w_j$, implying no violation of EPEW. Since $w_i \geq w_j$ is assumed, if $\theta_i > \theta_j$ then there is similarly no violation of EPEW.

(b) $\Leftrightarrow$ (c) It is easy to see that if all workers $i \in I$ have the same type then both **(b)** and **(c)** are satisfied. Therefore, we consider the case in which not all workers have the same type, and suppose that for any two workers $i,j \in I$ (where without loss of generality $\theta_i \geq \theta_j$), $w_i \geq w_j$ with equality if $\theta_i = \theta_j$. The requirement that for any two workers i,j such that $\theta_i = \theta_j$, it must be the case that $w_i = w_j$ implies that there exists some function $\phi : [\min_{i \in I} \theta_i, \max_{i \in I} \theta_i] \to [0,1]$ such that any worker i of productivity θ receives wage $w_i = \phi(\theta)$. We construct such a function ϕ that is non-decreasing. Without loss of generality, let $\theta_1 \leq \theta_2 \leq ... \leq \theta_{|I|}$, where if $i < j$ and $\theta_i = \theta_j$ then $w_j \geq w_i$. Define ϕ to be the step function containing points $(\theta_1, w_1), (\theta_2, w_2), ... (\theta_{|J|}, w_{|I|})$, where for each $\theta \in [\theta_1, \theta_{|I|}]$ we define $\phi(\theta) = \phi(\theta_i)$ for the unique i such that $\theta \in [\theta_i, \theta_{i+1})$. Then ϕ is clearly non-decreasing as desired.

To see that **(c)** implies **(b)**, suppose there exists a non-decreasing function $\phi : [\min_{i \in I} \theta_i, \max_{i \in I} \theta_i] \to [0,1]$ such that any worker i of productivity θ receives wage $w_i = \phi(\theta)$. That **(b)** is satisfied follows from two points. First, for any j such that $\theta_j = \theta$, $w_j = \phi(\theta) = w_i$. Second, for any j such that $\theta_j < \theta$ then $w_j = \phi(\theta_j) \leq \phi(\theta) = w_i$, where the inequality follows from the fact that ϕ is non-decreasing. □

□

A2 Proof of Proposition 2

Proof. Item 1 follows directly from Item (b) of Lemma A1. We therefore present a proof only for Item 2.

Fix $\epsilon > 0$ and let $\delta \in (0, \epsilon/4]$. Suppose that $S := (I, (\theta_i, w_i)_{i \in I})$ is a firm structure such that $\mathcal{F}_L(S) = 1$ and that, without loss of generality, for any $\theta \in [0,1]$ there exists a worker $i \in I$ such that $|\theta_i - \theta| < \delta$. The logic is unchanged if instead it is the case that for any $w \in [0,1]$ there exists a worker $i \in I$ such that $|w_i - w| < \delta$. To avoid repetition, we refer to both of the suppositions about S in the previous sentence as "the premise."

Relabel workers as necessary such that $\theta_1 \leq ... \leq \theta_{|I|}$. For convenience, we also introduce indices 0 and $|I|+1$ as "fake" workers, where $\theta_0 := 0$, $w_0 := 0 + w_1 \mathbb{1}_{\theta_1 = 0}$, $\theta_{|I|+1} := 1$, and $w_{|I|+1} := \mathbb{1}_{\theta_{|I|} < 1} + w_{|I|} \mathbb{1}_{\theta_{|I|} = 1}$. In words, "worker" 0 is the least productive and lowest paid worker, and "worker" $|I|+1$ is the most productive and highest paid worker. In what follows, we consider the case in which $w_0 = 0$ and $w_{|I|+1} = 1$; the complementary cases when $w_0 > 0$ or $w_{|I|} < 1$ follow from an essentially unchanged argument.

By the premise, it must be that for each $i \in \{1,...,|I|+1\}$:

$$\theta_i - \theta_{i-1} \in [0, 2\delta) \tag{22}$$

Consider the addition of new worker j to form S^+. It must be that $\theta_j \in [\theta_{i-1}, \theta_i]$ for precisely one $i \in \{1,...,|I|+1\}$. For such an $i \in \{1,...,|I|+1\}$, Item (b) of Lemma A1 implies that $\mathcal{F}_L(F^+)$ if and only if $w_j \in [w_{i-1}, w_i]$ where equality with w_{i-1} (w_i) must occur if $\theta_j = \theta_{i-1}$ ($\theta_j = \theta_i$). Let V_i denote the set of types for worker j with $\theta_j \in [\theta_{i-1}, \theta_i]$ that can be added to form S^+ such that $\mathcal{F}_L(S^+) = 1$. Then the preceding argument implies that $V_i = [\theta_{i-1}, \theta_i] \times [w_{i-1}, w_i]$. Because V_i and $V_{i'}$ are non-overlapping for $i \neq i'$,

$$|V| = |\bigcup_{i \in \{1,...,|I|+1\}} V_i| = \sum_{i \in \{1,...,|I|+1\}} |V_i|. \tag{23}$$

Note that by Item 1, $|V^+| \leq |V|$, where $|V^+|$ is defined as in Equation 23 for firm structure S^+, i.e. that "adding" a worker to a firm structure makes the firm structure no more likely to be "fair."

We seek to show that $|V| < \epsilon$. We proceed, given a firm structure S satisfying the premise, to create a firm structure S^M which removes workers whose productivities are too close together:

Step 1: Let i_1 be the smallest $i \in I$ such that $\theta_{i_1} - \theta_{i_1-1} < 2\delta$.[31] Define $S^1 := (I \setminus \{i_1\}, (\theta_i^1, w_i^1)_{i \in I \setminus \{i_1\}})$, where $(\theta_i^1, w_i^1) = (\theta_i, w_i)$ for all $i \in I \setminus \{i_1\}$, and continue on in the algorithm.

Step $k > 1$**:** Let i_k be the smallest $i \in I^{k-1}$ such that $\theta_{i_k}^{k-1} - \theta_{i'}^{k-1} < 2\delta$, where $i' \in I^{k-1}$ is the highest index worker in I^{k-1} that with index strictly lower than i_k, if such an i_k exists. If such an i_k exists, define $S^k := (I^{k-1} \setminus \{i_k\}, (\theta_i^k, w_i^k)_{i \in I^{k-1} \setminus \{i_k\}})$, where $(\theta_i^k, w_i^k) = (\theta_i^{k-1}, w_i^{k-1})$ for all $i \in I^{k-1} \setminus \{i_k\}$. Otherwise, let $S^k := F^{k-1}$ and terminate the algorithm.

As defined above, the algorithm terminates at the first step $k^* > 1$ such that $I^{k^*} = I^{k^*+1}$. Note that $k^* \leq |I| + 1$ by construction, i.e. the algorithm always terminates in finite time. For such k^*, define $S^M := (I^M, (\theta_i^M, w_i^M)_{i \in I^M}) = F^{k^*}$. Without loss of generality, relabel $i \in I^M$ such that $I^M = \{0, 1, ..., |I^M|\}$, where $\theta_0 < \theta_1 < ... < \theta_{|I^M|}$.

By the argument preceding the construction of S^M, it follows that $|V| \leq |V^M|$, where $|V^M|$ corresponds to Equation 23 for firm structure S^M. Therefore, the proof is complete if we show that $|V^M| < \epsilon$.

To see that $|V^M| < \epsilon$, notice that $\theta_i - \theta_{i-1} < 4\delta$ for each $i \in \{1, ..., |I^M|\}$, because otherwise, the algorithm constructing S^M would have terminated at an earlier step give that the productivity difference between consecutive workers in the initial firm structure S is no greater than 2δ. Also recall that for each worker $i \in I^M$ the wage $w_i \in [0,1]$. These two facts combine to imply that,

$$|V^M| < \sum_{i=1}^{|I^M|} 4\delta \cdot (w_i - w_{i-1}) \leq 4\delta(-0 + w_1 - w_1 + w_2 - w_2 + ... + 1) = 4\delta \leq \epsilon,$$

where the final inequality follows from the maintained assumption that $\delta \in (0, \epsilon/4]$.

□

A3 Proof of Proposition 3

Proof of Item 1. By construction $\bar{\mathcal{F}}_L$ is clearly stronger than $\mathcal{F}_L$. Therefore, $\bar{\mathcal{F}}_L(S) = 1$ only if $\mathcal{F}_L(S) = 1$. By Item 1, this implies that if $\bar{\mathcal{F}}_L(S) = 1$ then there exists a non-decreasing function $\phi : [\min_{i \in I} \theta_i, \max_{i \in I} \theta_i] \to [0,1]$ such that any worker i of productivity θ receives wage $w_i = \phi(\theta)$.

In the case that all $i \in I$ have the same type (θ, w), it is clear that $\bar{\mathcal{F}}_L(S) = 1$. Therefore, to complete the argument, we suppose that $\theta_i \neq \theta_j$ for some $i, j \in I$. Suppose ϕ is strictly increasing over $[\min_{i \in I} \theta_i, \max_{i \in I} \theta_i]$. Then there do not exist workers $i, j \in I$ with $w_i = w_j$ and $\theta_i \neq \theta_j$. Therefore, EWEP is satisfied in addition to EPEW, implying that $\bar{\mathcal{F}}_L(S) = 1$.

[31] By the premise, such an i_1 exists.

Now suppose $\bar{\mathcal{F}}_L(S)=1$. We construct the desired, strictly increasing function ψ as follows. Without loss of generality, let $\theta_1 \leq \theta_2 \leq ... \leq \theta_{|I|}$. By Item 1, if $\theta_i=\theta_j$ for any $i \neq j$ then $w_i=w_j$. Consider a maximal subset of workers J such that $\theta_1<...<\theta_{|J|}$, where we potentially relabel workers such that $\theta_1<\theta_2<...<\theta_{|J|}$. Then define ψ as a function containing points $(\theta_1,w_1),(\theta_2,w_2),...(\theta_{|J|},w_{|J|})$, and taking value $\psi(\theta)=w_i+\frac{w_{i+1}-w_i}{\theta_{i+1}-\theta_i}(\theta-\theta_i)$ for any $\theta \in (\theta_i,\theta_{i+1})$ for any $i \in \{1,...,|J|\}$. Then ψ satisfies the desired conditions.

□

Proof of Item 2. We first argue that $\bar{\mathcal{F}}_L$ is symmetric. Fix any firm structure $S:=(I,(\theta_i,w_i)_{i\in I})$. By an argument similar to that in Lemma A1, if $\bar{\mathcal{F}}_L(S)=1$ then for any two workers $i,j \in I$ (where without loss of generality $\theta_i \geq \theta_j$), $w_i \geq w_j$, where $w_i \geq w_j$ if and only if $\theta_i=\theta_j$. Therefore, for any two workers $i,j \in I$, either $(\theta_i,w_i)=(\theta_j,w_j)$, or both $\theta_i>\theta_j$ and $w_i>w_j$. Therefore, for the reflection S' of S it is the case that $\bar{\mathcal{F}}_L(S')=1$. Similarly, if $\bar{\mathcal{F}}_L(S)=0$ then $\bar{\mathcal{F}}_L(S')=0$. Therefore, $\bar{\mathcal{F}}_L$ is symmetric.

Next, we show that $\mathcal{F}_L$ is nearly symmetric, by appealing to our characterizations of $\mathcal{F}_L$ and $\bar{\mathcal{F}}_L$ in Proposition 1 and Item 1 of the current proposition, respectively. For any firm structure S such that there exists a strictly increasing function $\psi:[\min_{i\in I}\theta_i,\max_{i\in I}\theta_i] \to [0,1]$ such that for each worker i of productivity $\theta \in [0,1]$, $w_i=\psi(\theta)$, then we have established in Item 1 that $\mathcal{F}_L(S)=\mathcal{F}_L(S^r)=\bar{\mathcal{F}}_L=1$, where S^r is the reflection of S. Therefore, it remains to consider firm structures S such that there does not exist such a strictly increasing function ψ. There are two (exhaustive) cases:

- First, let S be such that there exists a non-decreasing function $\phi:[\min_{i\in I}\theta_i,\max_{i\in I}\theta_i] \to [0,1]$ such that any worker i of productivity θ receives wage $w_i=\phi(\theta)$. Then $\mathcal{F}_L(S)=1$. For any $\epsilon>0$, we construct a firm structure $S_\epsilon=(I,(\theta_i^\epsilon,w_i^\epsilon)_{i\in I})$ that admits a strictly increasing function $\psi:[\min_{i\in I}\theta_i,\max_{i\in I}\theta_i] \to [0,1]$ such that for each worker i of productivity $\theta^\epsilon \in [0,1]$, $w_i^\epsilon=\psi(\theta)$. To do so, let $\theta_i=\theta_i^\epsilon$ for all $i \in I$, and let $\gamma:=\min\left\{\min_{i\in I}\{1-w_i:w_i<1\} \cup \min_{i,j\in I}\{|w_i-w_j|:w_i\neq w_j\}\right\}$ be the minimum of the difference between any worker who receives a wage strictly lower than 1, and the smallest wage gap between workers who receive different wages. Let $J_i:=\{j\in I\setminus\{i\}:w_j=w_i \text{ and } \theta_j<\theta_i\}$ be the set of workers who receive the same wage as i and are strictly less productive. Let $K_i:=\{k\in I\setminus\{i\}:w_j=w_i \text{ and } \theta_k>\theta_i\}$. Let

$$w_i^\epsilon:=\begin{cases} w_i\cdot(1+\frac{\gamma}{4\cdot|I|}\cdot\epsilon|J_i|) & \text{if } w_i<1, \\ w_i\cdot(1-\frac{\gamma}{4\cdot|I|}\cdot\epsilon|K_i|) & \text{if } w_i=1. \end{cases}$$

Finally, $w_i^\epsilon \in [0,1]$ and $|w_i^\epsilon - w_i| < \epsilon$ by construction for all i. Therefore, S_ϵ satisfies all of the requirements of a firm structure, and is within ϵ "distance" of S. Furthermore, our construction implies that no workers receive the same wage unless they have the same productivity. Also, for any two workers i,j with $\theta_i > \theta_j$, $w_i^\epsilon > w_j^\epsilon$. Therefore, Item 1 and Proposition 1 imply that both S_ϵ and its reflection S_ϵ^r satisfy the strict monotonicity property, meaning that $\mathcal{F}_L(S) = \mathcal{F}_L(S_\epsilon^r) = \mathcal{F}_L(S_\epsilon) = 1$.

- Second, let S be such that there does not exist a non-decreasing function $\phi : [\min_{i\in I} \theta_i, \max_{i\in I} \theta_i] \to [0,1]$ such that any worker i of productivity θ receives wage $w_i = \phi(\theta)$. Then $\mathcal{F}_L(S) = 0$ by Proposition 1. Moreover, the proof of Proposition 1 implies that there exist distinct $i,j \in I$ such that $w_j > w_i$ and $\theta_j \leq \theta_i$. If there exist such i and j with $\theta_j < \theta_i$, then letting $S_\epsilon = S$ for any $\epsilon > 0$ implies that $\mathcal{F}_L(S) = \mathcal{F}_L(S_\epsilon^r) = \mathcal{F}_L(S_\epsilon) = 0$. If there does not exist such i and j with $\theta_j < \theta_i$, then for any $\epsilon > 0$ we can construct S_ϵ^r such that $\mathcal{F}_L(S) = \mathcal{F}_L(S_\epsilon^r) = \mathcal{F}_L(S_\epsilon) = 0$ in a similar fashion as in the previous cases we consider.

In both of these exhaustive cases, we have shown the desired result. Therefore, we conclude that $\mathcal{F}_L$ is nearly symmetric.

Finally, note that from Proposition 1 and Item 1 of the current proposition, it is clear that $\bar{\mathcal{F}}_L$ is stronger than $\mathcal{F}_L$. Consider any symmetric fairness rule $\mathcal{F}$ that is stronger than $\mathcal{F}_L$. We show that this $\mathcal{F}$ is also stronger than $\bar{\mathcal{F}}_L$. To this end, consider $S := (I, (\theta_i, w_i)_{i\in I})$ such that $\mathcal{F}_L(S) = \mathcal{F}(S) = 1$. By Proposition 1, there exists a non-decreasing function $\phi : [\min_{i\in I} \theta_i, \max_{i\in I} \theta_i] \to [0,1]$ such that any worker i of productivity θ receives wage $w_i = \phi(\theta)$. The combination of $\mathcal{F}_L(S) = 1$ and the symmetry restriction implies that there are no two workers $i,j \in I$ such that $\theta_i \neq \theta_j$ and $w_i = w_j$, because otherwise this would contradict that $\mathcal{F}(S) = 1$. But then by the proof of Item 1 of the current proposition, there exists a strictly increasing function $\psi : [\min_{i\in I} \theta_i, \max_{i\in I} \theta_i] \to [0,1]$ such that for each worker i of productivity $\theta \in [0,1]$, $w_i = \psi(\theta)$. This implies that $\bar{\mathcal{F}}_L(S) = 1$, completing the proof.

□

A4 Proof of Proposition 4

Proof. We prove the result by considering each of the following three claims.

Symmetry: We begin by demonstrating the claim regarding symmetry. We say that $g \in \{A,B\}$ experiences a *strong group-level violation of EPEW* compared to group $g' \neq g$ if the average pay of g-group members is lower than the average pay of g'-group members across all iso-productivity

bands: for each $\theta \in \hat{\Theta}$,

$$\frac{\sum_{i \in g^{\theta}} w_i}{|g^{\theta}|} \leq \frac{\sum_{i \in g'^{\theta}} w_i}{|g'^{\theta}|}, \tag{24}$$

where the inequality is strict for at least one $\theta \in \hat{\Theta}$ and where g^{θ} and g'^{θ} indicate the set of workers of group g and g', respectively, who have productivity θ. We similarly say that $g \in \{A,B\}$ experiences a *strong group-level violation of EWEP* if the average productivity of $g-$group members is higher than the average productivity of $g'-$group members across all iso-wage bands.

By construction, if a firm structure S admits a strong group-level violation of EPEW (EWEP) for group B, then $\beta^{isoprod} < 0$ and $\beta^{contprod} < 0$ ($\beta^{isowage} > 0$ and $\beta^{contwage} > 0$), implying that $\mathcal{F}^{isoprod}(S) = \mathcal{F}^{contprod}(S) = 0$ ($\mathcal{F}^{isowage}(S) = \mathcal{F}^{contwage}(S) = 0$). Moreover, the inequalities in the previous sentence are reversed if S admits a strong group-level violation of EPEW (EWEP) for group A. Therefore, the desired claim is established via the following remark.

Remark 5. *A firm structure $S := (I,(\theta_i,w_i)_{i \in I})$ with partition A,B of workers can simultaneously exhibit the following features:*

- *S does not satisfy EPEW or EWEP, and*
- *Group A experiences a strong group-level violation of EWEP while group B experiences a strong group-level violation of EPEW.*

These claims follow from the example firm structure presented in Table 1. For each iso-wage band at the firm (rows), group A workers are strictly more productive on average. However, for each iso-productivity band (columns), group B workers are strictly lower paid on average. Therefore, group A experiences a strong group-level violation of EWEP while group B experiences a strong group-level violation of EPEW.

Distinct fairness rules: Next, we demonstrate that $\mathcal{F}^{isoprod}$,$\mathcal{F}^{contprod}$,$\mathcal{F}^{isowage}$, and $\mathcal{F}^{contwage}$ are distinct fairness rules. That $\mathcal{F}^{isoprod} \neq \mathcal{F}^{isowage}$ and $\mathcal{F}^{contprod} \neq \mathcal{F}^{contwage}$ follows from Remark 5 and the argument immediately preceding it, which jointly demonstrate the existence of a firm structure S for which $\mathcal{F}^{isoprod} = 0 = \mathcal{F}^{contprod}$ and $\mathcal{F}^{isowage} = 1 = \mathcal{F}^{contwage}$. To see that $\mathcal{F}^{isoprod} \neq \mathcal{F}^{contprod}$ and $\mathcal{F}^{isowage} \neq \mathcal{F}^{contwage}$, note that for any firm structure $S = (I,(\theta_i,w_i)_{i \in I})$ in which $w_i \neq w_j$ and $\theta_i \neq \theta_j$ for all distinct i and j, $\mathcal{F}^{isoprod}(S) = \mathcal{F}^{isowage}(S) = 1$ by construction. However, note that $\beta^{contprod}$ and $\beta^{contwage}$ are continuous in worker types holding fixed the set of workers. Formally, for any firm structure $S = (I,(\theta_i,w_i)_{i \in I})$ and $\epsilon > 0$, there exists $\delta > 0$ such that any firm structure $S_\delta = (I,(\theta_i^\delta,w_i^\delta)_{i \in I})$ where $|\theta_i^\delta - \theta_i| < \delta$ and $|w_i^\delta - w_i| < \delta$, it is the case that $|\beta_S^{contprod} - \beta_{S_\delta}^{contprod}| < \epsilon$ and $|\beta_S^{contwage} - \beta_{S_\delta}^{contwage}| < \epsilon$, where we index the regression coefficients by the firm structure. In light of this statement, consider the firm structure S

presented in Table 1. For such a "nearby" firm structure S_δ for sufficiently small δ in which all workers have distinct wages and all workers have distinct productivities, it is the case that $\mathcal{F}^{isoprod}(S_\delta)=1$, $\mathcal{F}^{contprod}(S_\delta)=0$, $\mathcal{F}^{isowage}(S_\delta)=1$, and $\mathcal{F}^{contwage}(S_\delta)=0$. Thus, our arguments in this paragraph have established there exist firm structures for which any two of the fairness rules discussed differ in their output.

Near symmetry: Finally, we demonstrate the claim regarding near symmetry. First, we consider $\mathcal{F}^{isoprod}$. Consider a firm structure $S=(I,(\theta_i,w_i)_{i\in I})$ for which $\beta_S^{isoprod}<0$ and in which all workers have different wages, $w_i\neq w_j$ for all distinct $i,j\in I$. Then $\mathcal{F}^{isoprod}(S)=0$ and $\mathcal{F}^{isoprod}(S^r)=1$ by construction, where S^r is the reflection of S. Also, for sufficiently small δ, any nearby (i.e., defined as in the proof of distinctness of the four fairness rules) firm structure $S_\delta=(I,(\theta_i^\delta,w_i^\delta)_{i\in I})$ is such that $w_i^\delta\neq w_j^\delta$ for all distinct $i,j\in I$. Therefore, as argued previously, $\mathcal{F}^{isoprod}(S_\delta^r)=1$ for the reflection of any such firm structure S_δ given sufficiently small δ. Therefore, $\mathcal{F}^{isoprod}$ is not nearly symmetric because $\mathcal{F}^{isoprod}(S_\delta^r)\neq\mathcal{F}^{isoprod}(S)=0$. A nearly identical argument in which wages and productivity are interchanged establishes that $\mathcal{F}^{isowage}$ is not nearly symmetric.

Second, we consider $\mathcal{F}^{contprod}$. Consider a firm structure $S=(I,(\theta_i,w_i)_{i\in I})$ for which $\beta_S^{contprod}<0$ and $\beta_{S^r}^{contprod}>0$, as in Table 1. Then $\mathcal{F}^{contprod}(S)=0$ and $\mathcal{F}^{contprod}(S^r)=1$. But as previously argued, $\mathcal{F}^{contprod}$ is continuous in worker types, that is, for any "nearby" firm structure $S_\delta=(I,(\theta_i^\delta,w_i^\delta)_{i\in I})$ for sufficiently small δ, it is still the case that $\mathcal{F}^{contprod}(S_\delta)=0$ and $\mathcal{F}^{contprod}(S_\delta^r)=1$. Therefore, $\mathcal{F}^{contprod}$ is not nearly symmetric because $\mathcal{F}^{contprod}(S_\delta)=0\neq1=\mathcal{F}^{contprod}(S_\delta^r)$. A nearly identical argument in which wages and productivity are interchanged establishes that $\mathcal{F}^{contwage}$ is not nearly symmetric.

□

A5 Common limitations of EPEW restrictions

Our analysis thus far has interpreted EPEW broadly, binding across all workers within the firm. In this section, we consider two real-world limitations that restrict the set of constraints that EPEW often imposes.

Formally, given a set of workers I, let $c:I\rightrightarrows I$ be a correspondence that tracks for each worker $i\in I$ which set of workers $I\setminus\{i\}$ cannot "violate" i by the definition of EPEW. That is, given a firm structure $S=(I,(\theta_i,w_i)_{i\in I})$ and such a "comparability correspondence" $c(\cdot)$, the *(limited by c) EPEW condition* prohibits S from including i and $j\in c(i)$ such that $\theta_j\leq\theta_i$ and $w_j>w_i$. We assume that $c(\cdot)$ is symmetric: for $i,j\in I$, $j\in c(i)$ if and only if $i\in c(j)$.

We now discuss two common restrictions on $c(\cdot)$:

Job-based restrictions: Some firms employ workers for jobs that are qualitatively different from others, the set of comparable workers are those doing the same job. To capture this, we assume that $c(\cdot)$ is transitive: for $i,j,k \in I$, for any $\{j,k\} \subset c(i)$ then $j \in c(k)$ (and $k \in c(j)$). With such a comparability correspondence $c(\cdot)$, all of our earlier results can be interpreted as being "within job." That is, our conclusions hold within a particular set of workers $J \subset I$ such that all members $i \in J$ are comparable to one another and not comparable to any workers outside of J.

Demographic-based restrictions: Some EPEW explicitly bind across, and not within, demographic groups. For example, Kansas recognizes two demographic groups, men and women, and EPEW restrictions bind only across genders (Kan. Stat. Ann. § 44-1205 (2025)):

> *"[N]o employer having employees of both sexes shall discriminate, within any establishment in which such employees are employed, between employees on the basis of sex by paying wages to employees in such establishment at a rate less than the rate of wages paid to employees of the opposite sex in such establishment for equal work on jobs..."*

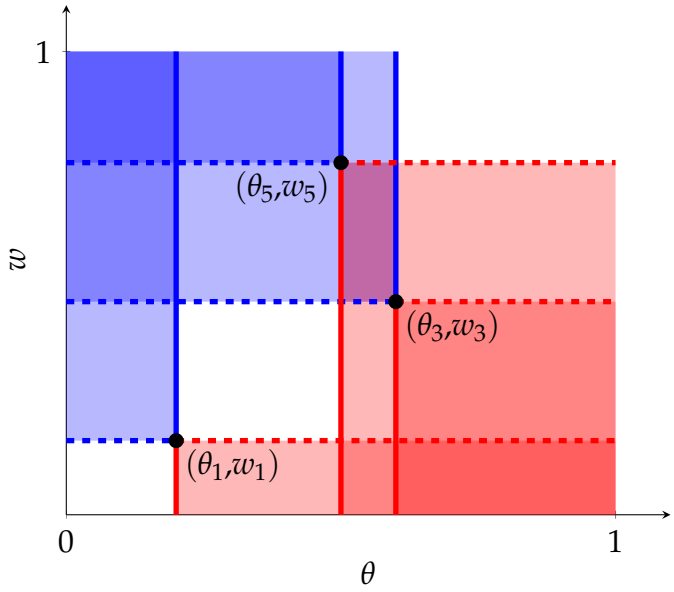
Formally, we assume that $c(\cdot)$ is associated with a partition $\{P_1,P_2,...,P_m\}$ of set I where $m \leq |I|$. For any $i \neq j \in I$, $j \in c(i)$ (and, by symmetry, $i \in c(j)$) if and only if $i \in P_\ell$ and $j \in P_k$ for $k \in \{1,...,m\}$ and $\ell \in \{1,...,m\} \setminus \{k\}$.

This restriction to across-demographic comparison groups is necessarily weaker than the all-worker comparison group we have studied heretofore. To see this visually, consider Figure 4. This figure displays a firm structure with six workers, numbered 1 through 6. Odd-numbered workers are members of the first demographic group (partition member P_1) and even-numbered workers are members of the second demographic group (partition member P_2). Plotted are the types of workers in demographic group 1. To accord with a demographic-based EPEW with two groups defined as above, workers in group 2 can have types anywhere in the white regions and on any of the segments of the dotted lines adjacent to white regions. Clearly, pay monotonicity is not satisfied: worker 5 is more productive and lower paid than is worker 3.

Remark 6. *Having displayed the difference between the "all comparisons" EPEW and the demographic-based EPEW restriction, we make some observations about conditions under which they converge:*

- *Holding fixed a firm structure S and considering two comparability correspondences $c(\cdot)$ and $\hat{c}(\cdot)$ such that the partition of workers induced by $\hat{c}(\cdot)$ is finer than the partition induced by $c(\cdot)$, then $\hat{c}(\cdot)$ the firm structure satisfies demographic EPEW with respect to $\hat{c}(\cdot)$ only if it satisfies demographic EPEW with respect to $c(\cdot)$. In the limit that $c(\cdot)$ produces the trivial partition (all workers are of the same demographic group) then EPEW with respect to $c(\cdot)$ imposes no restriction. In the limit that $\hat{c}(\cdot)$ induces the finest possible partition (each worker is of a distinct demographic group) then EPEW with respect to $\hat{c}(\cdot)$ is equivalent to pay monotonicity (our Proposition 1).*

Figure 4: **Demographic-Based EPEW Restriction**



Notes: This figure represents demographic-based EPEW restrictions for a particular firm structure and comparability correspondence. There are six workers, and odd-numbered workers are of one demographic group and even-numbered workers are of a second. Even-numbered workers can have types anywhere in the white regions and on any of the segments of the dotted lines adjacent to white regions, but pay monotonicity is necessarily violated (see workers 3 and 5).

- *Even if the comparability correspondence does $c(\cdot)$ does not induce a partition such that each worker is in their own partition member, EPEW with respect to $c(\cdot)$ will be equivalent to pay monotonicity if for each $i \in I$ of type (θ_i, w_i) either: there exists no distinct $j \in I$ with productivity θ_j such that $\theta_j \geq \theta_i$, or there exists a distinct worker $j \in I$ with productivity θ_j such that there exists no distinct worker $k \in I$ with productivity θ_k such that $\theta_j > \theta_k \geq \theta_i$ and where j is of a different partition element from i. Similar logic as the proof of Proposition 1 binds, leading to a violation of EPEW with respect to $c(\cdot)$ if pay monotonicity is not satisfied. The "next most productive" worker j (or one of the "next most productive" workers) is of a different demographic group, implying that j is paid no less than i. If there is another worker k of the same demographic group as i and the productivity as j cannot be paid strictly less than j, which, by transitivity, means that k cannot be paid strictly less than i.*

- *Given a demographic-based comparability correspondence $c(\cdot)$, an analogue of Proposition 2 holds with minor modifications. First, adding a new worker to the firm structure (holding fixed the demographic groups of all existing workers) again makes EPEW with respect to $c(\cdot)$ (the analogue of point 1 of Proposition 2). Second, the "flexibility" result in part 2 of Proposition 2 holds if*

there exists a worker of each demographic group induced by $c(\cdot)$ within either a δ productivity neighborhood or a δ wage neighborhood of any value in $[0,1]$.

A6 New Group-Level Equality Measure

Although our main paper highlights complications with regression based notions of EWEP and EPEW, there exist different group-level measures of fairness that satisfy our axioms. Given that the approaches above control for productivity when considering wages (or vice versa), it seems intuitive that relevant notions of fairness may take into account both wage and productivity discrepancies across groups. For example, one could define a fairness rule $\mathcal{F}$ that deems a firm structure to be unfair if and only if $\beta^{contprod} \leq 0$ and $\beta^{contwage} \leq 0$. Alternatively, consider the following fairness rule:

Group-Aggregated Distance to Monotonicity. Let S be a firm structure, and consider a partition of workers represented in S into groups A and B. For each $i \in A$ we consider the *violated* set of workers $U_i := \{j \in B | \theta_j \geq \theta_i, w_j \leq w_i\} \setminus \{j \in B | \theta_i = \theta_j, w_i = w_j\}$ which are the workers in B who are more productive and paid less than i (with at least one comparison strict). Similarly, consider the *violating* set of workers $L_i := \{j \in B | \theta_j \leq \theta_i, w_j \geq w_i\} \setminus \{j \in B | \theta_i = \theta_j, w_i = w_j\}$. For any pair of workers i and j, let $d_{ij} := (w_i - w_j)^2 + (\theta_i - \theta_j)^2$ be the Euclidean distance (in the two-dimensional plane specifying wage and productivity) between i and j. We define

$$\beta^{dist} := \sum_{i \in A} \sum_{j \in L_i} d_{ij} - \sum_{i \in A} \sum_{j \in U_i} d_{ij} \tag{25}$$

which is the total "overpayment given productivity" (first term) minus the total "underpayment given productivity" of group A. If and only if the total overpayments weakly exceed the total underpayments is $\beta^{dist} \leq 0$. A fairness rule $\mathcal{F}$ that renders a firm structure S fair if and only if $\beta^{dist} \leq 0$ is clearly symmetric due to the fact that the Euclidean distance of any two points in $\mathbb{R}^2$ is invariant to reflection along the 45 degree line.

B Methodology Details

B1 Weak Pairwise Inequalities

Our main linear program for rationalizability tests where the followingset is non-empty:

$$\mathcal{B} \equiv \left\{ \vec{\beta} \in \mathbb{R}^K : w_i > w_j \Longrightarrow \vec{\beta} \cdot \vec{\theta}^i > \vec{\beta} \cdot \vec{\theta}^j, \forall i \neq j \right\}.$$

We can reformulate this as follows.

$$\begin{aligned}
&\text{maximize} \quad \delta \\
&\quad \vec{\beta},\delta \\
&\text{subject to} \quad \vec{\beta}\cdot(\vec{\theta}^i-\vec{\theta}^j)\geq\delta, \qquad \forall i\neq j: w_i>w_j \qquad (1), \\
&\qquad\qquad\qquad \textstyle\sum\beta_k=1 \qquad (2), \\
&\qquad\qquad\qquad \beta_k\geq 0 \qquad (3), \\
&\qquad\qquad\qquad \delta\geq 0 \qquad (4)
\end{aligned} \tag{26}$$

Interpretation. Constraint (1) says that if $w_i > w_j$, then i's productivity must exceed j's by a common slack term δ. Constraints (2) and (3) are scale normalization. Without some normalization, the LP becomes scale-invariant and unbounded, and the objective (maximize δ) would blow up to $+\infty$. Constraint (4) keeps δ positive. Then the solver finds the *largest common* slack δ that fits *every* pairwise inequality with the *same* direction as the observed wage ordering. We call this term $\delta^\star$. This produces a clear certificate.

$\delta^\star > 0 \iff$ the strict system is feasible (all inequalities can be made strict).

$\delta^\star = 0 \iff$ the strict system is infeasible (at least one inequality can only hold weakly).

Empty $\delta^\star \iff$ the system is infeasible (cannot even hold weakly).

Note that this approach can be combined with other constraints. In our empirical application, we also require that all β_k terms be weakly positive, and that at least one β be strictly positive.

B2 Minimal Strict Cap $\bar{\rho}$ via Charnes-Cooper and Lexicographic Search

This appendix gives a procedure to compute

$$\bar{\rho}_{\text{strict}} = \inf\left\{\rho\geq 0\colon \exists(\beta,\eta) \text{ with } (\theta_i-\theta_j)\cdot\beta+\eta_i-\eta_j>0 \ \ \forall(i,j): w_i>w_j \text{ and } \rho\max_k\beta_k\geq\max_i\eta_i\right\},$$

that is, the minimal cap on unobserved productivity, normalized by $\max_k\beta_k$, that admits a strictly monotone rationalization of the observed pay ordering. Because (β,η) are only identified up to a common positive scaling, this normalization is substantive. Throughout, both $\bar{\rho}_{\text{strict}}$ in this appendix and $\bar{\rho}$ in the main text are defined relative to $\max_k\beta_k$, i.e. the most heavily weighted observable metric in the feasible solution; they are not separate ratios to an ex ante "top metric" anchor.

The method uses a lexicographic two-stage scheme: an outer search over ρ and, at each ρ,

a small family of linear programs (one per candidate maximizer of the observable weights) that certifies that the feasible set is non-empty with strict pairwise inequalities.

Setup. Fix a cell (school × discipline × rank × year). For each worker i let

$$\theta_i = (\theta_{i1},\ldots,\theta_{iK}) \in \mathbb{R}^K_+$$

collect observable productivity metrics, and write latent productivity as

$$y_i = \theta_i \cdot \beta + \eta_i,$$

with unknown weights $\beta = (\beta_1,\ldots,\beta_K) \in \mathbb{R}^K_+$ and nonnegative unobservables $\eta_i \geq 0$. If $w_i > w_j$, strict consistency of pay with latent productivity requires

$$(\theta_i - \theta_j)\cdot\beta + \eta_i - \eta_j > 0 \qquad \text{for all } (i,j) \text{ with } w_i > w_j. \tag{27}$$

In a series of upcoming steps, this appendix bounds the size of unobservables relative to the strongest metric weight by

$$\rho \cdot \max_k \beta_k \geq \max_i \eta_i, \tag{28}$$

and seeks the smallest ρ that permits a *strict* solution to (27). As discussed in Section 4.2, solving this system of equations yields a bound on the relative importance of unobserved productivity compared to observed productivity metrics.

Step 1: Charnes-Cooper scaling (1962) makes the cap linear. The ratio in (28) is nonlinear. Let

$$B := \max_k \beta_k, \qquad t := \frac{1}{B}, \qquad \tilde{\beta} := t\beta, \quad \tilde{\eta} := t\eta, \quad \tilde{H} := t \cdot \max_i \eta_i.$$

We restrict attention to solutions with $B > 0$ (equivalently, at least one $\beta_k > 0$), which is without loss for our computation because the inner LP normalization $\tilde{\beta}_k = 1$ enforces $B > 0$. Then $0 \leq \tilde{\beta}_k \leq 1$ for all k. By construction, $\max_k \tilde{\beta}_k = 1$. Operationally, we implement this in the inner step by solving one LP for each $k \in \{1,\ldots,K\}$ with the normalization $\tilde{\beta}_k = 1$ (and $0 \leq \tilde{\beta}_\ell \leq 1$ for all ℓ), and then taking the best (largest) strictness margin across k. Multiplying (27) by t yields the *scaled* strict inequalities

$$(\theta_i - \theta_j)\cdot\tilde{\beta} + \tilde{\eta}_i - \tilde{\eta}_j > 0 \qquad \text{for all } (i,j) \text{ with } w_i > w_j. \tag{29}$$

Similarly, multiplying (28) by t yields the scaled bound

$$\tilde{H} = t\max_i \eta_i = \max_i \tilde{\eta}_i \le \rho. \tag{30}$$

Because LPs cannot encode strict inequality relationships directly, we introduce a *common margin* $\delta \ge 0$ and require

$$(\theta_i - \theta_j)\cdot\tilde{\beta} + \tilde{\eta}_i - \tilde{\eta}_j \ge \delta \qquad \text{for all } (i,j) \text{ with } w_i > w_j, \tag{31}$$

so that $\delta > 0$ certifies strict feasibility. In the primitive (unscaled) variables, the common margin equals δ/t.

Why do K LPs suffice? After the scaling, the nonlinear condition $\max_k \tilde{\beta}_k = 1$ is equivalent to the union over k of the linear events $\{\tilde{\beta}_k = 1,\ 0 \le \tilde{\beta}_\ell \le 1\ \forall \ell\}$. Any feasible solution must belong to at least one such polyhedron; if several metrics tie for the maximum weight, the same solution belongs to several of them. Hence solving one LP per candidate maximizer k and taking the largest common slack exactly recovers the feasible set.

Step 2: Inner LP at a fixed cap ρ: maximize the common margin. For a *given* $\rho \ge 0$ and each $k \in \{1,\dots,K\}$, define the LP

$$\begin{aligned}
\phi_k(\rho) := \max_{\tilde{\beta},\tilde{\eta},\tilde{H},\delta \ge 0} \quad & \delta && && (32)\\
\text{s.t.} \quad & (\theta_i - \theta_j)\cdot\tilde{\beta} + \tilde{\eta}_i - \tilde{\eta}_j \ge \delta, \forall (i,j) : w_i > w_j, && (33)\\
& 0 \le \tilde{\beta}_\ell \le 1, && \forall \ell, && (34)\\
& \tilde{\beta}_k = 1, && && (35)\\
& 0 \le \tilde{\eta}_i \le \tilde{H}, && \forall i, && (36)\\
& \tilde{H} \le \rho. && && (37)
\end{aligned}$$

Note that the constraints $0 \le \tilde{\eta}_i \le \tilde{H}$ implement $\max_i \tilde{\eta}_i \le \tilde{H}$, and hence are equivalent to imposing $\max_i \tilde{\eta}_i \le \rho$ together with an auxiliary variable $\tilde{H}$. Define the strictness certificate at ρ as $\phi(\rho) := \max_{k \in \{1,\dots,K\}} \phi_k(\rho)$.

This is the same device used to guarantee strict inequalities as in Appendix B1. The normalization $\tilde{\beta}_k = 1$ pins down the overall scale of $(\tilde{\beta},\tilde{\eta},\delta)$ so that the objective δ is well-defined and ρ retains the interpretation of bounding $\max_i \eta_i$ relative to the most heavily weighted observable metric.

Interpretation.

- If $\phi(\rho) > 0$, then there exists a solution with *strict* pairwise inequalities at that cap ρ.
- If $\phi(\rho) = 0$, only weak inequalities (or infeasibility) can be attained at ρ.

Because the feasible set of observed and unobserved productivity weights expands as ρ increases, $\phi(\rho)$ is nondecreasing in ρ. Consequently, there exists a threshold $\bar{\rho}_{\text{strict}}$ such that $\phi(\rho) > 0$ (i.e., strict feasibility holds) if and only if $\rho \in [\bar{\rho}_{\text{strict}}, \infty)$.

Step 3 (Outer search): bisection for the minimal strict cap. We now locate the smallest cap ρ that admits strict feasibility. Because our procedure uses numerical rather than symbolic optimization, the algorithm approaches $\bar{\rho}_{\text{strict}}$ iteratively by evaluating a sequence of candidate values of ρ. We describe the procedure for selecting and testing candidate values immediately below (Step 3.1).

To decide when this iterative search has converged (Step 3.2), we introduce a small positive constant $\tau_{\text{conv}} > 0$, called the *convergence tolerance*. This quantity sets the numerical precision of the search: once successive candidate values of ρ differ by less than τ_{conv}, the procedure halts and reports the corresponding approximation to $\bar{\rho}_{\text{strict}}$ (Step 3.3). The convergence tolerance is purely computational—it governs the stopping rule for the numerical search and does not affect the theoretical model itself.

Numerical implementation. Because the inner problems are solved numerically, we distinguish the solver tolerance from the outer bisection tolerance. Let $\tau_{LP} > 0$ denote a small inner tolerance. We classify a candidate ρ as strictly feasible only if the solver returns feasibility and the optimal common slack satisfies $\phi(\rho) > \tau_{LP}$; values in $[0, \tau_{LP}]$ are treated as numerically zero. The outer search then stops once $U - L < \tau_{conv}$, with $\tau_{LP} \ll \tau_{conv}$. This keeps the strict-feasibility classification from being driven by machine precision.

Step 3.1: Bracketing. We begin by calculating the end points for our search over candidate values for ρ. We compute a weak lower bound

$$\bar{\rho}_{\text{weak}} = \min_{k \in \{1,\ldots,K\}} \min_{\tilde{\beta}, \tilde{\eta}, \tilde{H}} \tilde{H} \quad \text{s.t. } (\theta_i - \theta_j) \cdot \tilde{\beta} + \tilde{\eta}_i - \tilde{\eta}_j \geq 0 \ \forall (i,j) : w_i > w_j, 0 \leq \tilde{\beta}_\ell \leq 1 \ \forall \ell, \tilde{\beta}_k = 1, 0 \leq \tilde{\eta}_i \leq \tilde{H} \ \forall i.$$

Set $L := \bar{\rho}_{\text{weak}}$. This value represents the smallest cap on unobserved productivity for which the *weak* system (with $\delta = 0$) is feasible; for any $\rho < L$, the model cannot satisfy even the non-strict inequalities. We then check $\phi(L)$: if $\phi(L) > 0$, then we have found that L is already the strict cap and can proceed to Step 3.3 (completion and report).

Otherwise, we next find an upper bound for our interval search. We grow an upper bound

geometrically until strictness appears: set the variable U equal to $\max\{1,2L\}$.[32] We then keep doubling U until $\phi(U) > 0$.[33]

The upper bound U found in this step is only an upper bound for the search. It is not the minimal strict cap itself. The geometric doubling in Step 3.1 is only used to locate an initial bracket (that is, end points for our search). This consists of a lower bound L with $\phi(L) = 0$ and an upper bound U with $\phi(U) > 0$. This ensures that the true minimal strict cap $\bar{\rho}_{\text{strict}}$ lies somewhere in $(L,U]$. In the step below, we locate the cap within this interval.

Step 3.2: Bisection (if $L < U$). We now proceed with a *binary search* over the bracketing interval (Knuth, 1997; Cormen et al., 2009, described in detail next). This requires a monotonicity condition that we show is met at the end of this appendix (page a17).

If the width of the search interval exceeds the convergence tolerance ($U - L > \tau_{\text{conv}}$), then we compute the midpoint $M = (L+U)/2$ and solve the family of inner LPs (32)–(37) at $\rho = M$ (one for each $k \in \{1,\ldots,K\}$) to evaluate

$$\phi(M) = \max_{k \in \{1,\ldots,K\}} \phi_k(M).$$

This determines whether the system is strictly feasible at $\rho = M$.

Because M lies between L and U, monotonicity implies that either $\phi(M) = 0$ (in which case the strict threshold exceeds M) or $\phi(M) > 0$ (in which case the strict threshold is at most M). In the former case, we know that the minimum cap lies in the interval $(M,U]$. In the latter case, we know the minimum cap lies in the interval $(L,M]$. Therefore, in the former case we can repeat this bisection stage recursively over the interval $(M,U]$. In the latter case we can repeat this bisection stage recursively over the interval $(L,M]$. This procedure can be iterated arbitrarily many times to give a tighter upper and lower bound of the strict cap.

Notationally, at each iteration we maintain the convention that L denotes the largest value of ρ tested so far that *fails* strict feasibility ($\phi(L) = 0$) and U denotes the smallest value of ρ that *succeeds* ($\phi(U) > 0$). The midpoint M is simply the next test value in the iteration. If $\phi(M) > 0$, we update the upper bound $U \leftarrow M$; otherwise, we update the lower bound $L \leftarrow M$. By construction, the feasible interval shrinks by half each iteration. The procedure continues until the interval width $U - L$ is smaller than the numerical tolerance τ_{conv}.

[32] The expression $\max\{1,2L\}$ is a numerical safeguard that ensures the initial upper bound U starts from a positive, nontrivial value even if L is zero or extremely small. Once $L > 0$, the "max" term has no effect—the doubling loop proceeds in exactly the same way and still locates a finite U for which $\phi(U) > 0$, whenever such a value exists.

[33] Because strict inequalities are imposed only on pairs with $w_i > w_j$, the binding comparisons define an acyclic order over distinct wage levels; ties generate no constraint. Assigning sufficiently separated η-values across those wage levels therefore yields strict feasibility once ρ is large enough.

At that point, the interval endpoints satisfy $\phi(L) = 0$ and $\phi(U) > 0$, with width $U - L < \tau_{\text{conv}}$. By monotonicity, the true threshold $\bar{\rho}_{\text{strict}}$ lies in $(L, U]$, so reporting U approximates $\bar{\rho}_{\text{strict}}$ to within τ_{conv}.

Step 3.3: Completion and Report. After the bisection procedure terminates (i.e., once $U - L < \tau_{\text{conv}}$), the interval endpoints satisfy $\phi(L) = 0$ and $\phi(U) > 0$, with $U - L < \tau_{\text{conv}}$. Thus, by monotonicity, $\bar{\rho}_{\text{strict}} \in (L, U]$, and we report the approximation

$$\bar{\rho}_{\text{strict}} = U, \qquad \text{with strict common margin } \delta^\star(U) = \phi(U) > 0.$$

Because only the relative scale of (β, η) matters for the inequalities and the cap, the inner LP identifies these objects only up to a positive multiplicative factor. The solution $(\tilde{\beta}^\star, \tilde{\eta}^\star)$, normalized by $\tilde{\beta}^\star_{k^\star} = 1$ for some maximizer $k^\star \in \operatorname{argmax}_k \tilde{\beta}^\star_k$, provides one convenient representative. Any positive rescaling of (β, η) leaves the inequalities and the minimal strict cap $\bar{\rho}_{\text{strict}}$ unchanged.

Note on Monotonicity. During the bisection search, the current candidate values of the cap always lie in an interval $I = (L, U]$ whose width is eventually driven below the convergence tolerance τ_{conv}. Let ρ_1 and ρ_2 be any two values in this interval with $\rho_1 \le \rho_2$. Because increasing ρ relaxes the linear constraint $\tilde{H} \le \rho$ in each LP (32)–(37), every solution feasible for $\phi_k(\rho_1)$ remains feasible for $\phi_k(\rho_2)$ whenever $\rho_1 \le \rho_2$. Hence each $\phi_k(\rho)$ is nondecreasing in ρ, and therefore their maximum $\phi(\rho) = \max_k \phi_k(\rho)$ is also nondecreasing in ρ. This monotonicity property is what makes bisection on ρ valid: once we find an interval $(L, U]$ with $\phi(L) = 0$ and $\phi(U) > 0$, the true threshold lies inside this interval, and bisection shrinks it to length no more than τ_{conv}.

B3 Gender Rationalizability

In our setup thus far, we have included all pairwise comparisons between workers (Definition 6). Example 1 included 3 workers and $\binom{|3|}{2} = 3$ pairwise comparisons. However in certain jurisdictions, equal pay laws regulate only *certain* pairwise comparisons. A common relaxation of EPEW is to require no worker receives strictly lower pay than a worker *of the opposite gender* if the worker is weakly more productive. For example, the California Equal Pay Act (adopted 1949) implemented this form of equal pay regulation before broadening in 2015 to include a larger set of pairwise comparisons.

Our monotonicity test can be easily adjusted to such scenarios by dropping all productivity inequalities involving two workers of the same gender.

Definition B1. *Let $g(i)$ represent the gender of worker i. The pairwise inequalities of differently-gendered workers define a feasible set for the weight vector $\vec{\beta}$:*

$$\mathcal{B}^G \equiv \left\{ \vec{\beta} \in \mathbb{R}^K : w_i > w_j \implies \vec{\beta} \cdot \vec{\theta}^i > \vec{\beta} \cdot \vec{\theta}^j, \forall i \neq j, g(i) \neq g(j) \right\}.$$

A wage structure is gender rationalizable *if the feasible set $\mathcal{B}^G$ is non-empty.*

"Gender rationalizability" can be generalized to arbitrary characteristics beyond gender (i.e., race or other characteristic). In the next two sections, we measure "distances to monotonicity" for wage distributions that are not rationalizable. Each distance metric is based on pairwise inequalities, and can be specialized to a gender-specific version by similarly dropping same-gender inequalities.

C Dataset Construction

Below, we detail the procedure used to construct our main dataset.

C1 Salary Sample Selection

We began by selecting the top 100-ranked economics graduate programs from US News and World Report's "Best Economics Schools" (2022) as sample candidates. From this pool, we identified 63 public universities where faculty salaries are presumed to be public record (legal obligation) and conducted our search for salaries in two passes. In the first, using search terms "[university] faculty salary database," we retrieved salaries for 18 universities, often sourced through their institutional resource offices. For another 31, using search terms "[state] salary database," we obtained salaries from the corresponding state database. Data was unavailable for the remaining 14 universities. Certain universities in the top 100 list belonged to a larger, statewide university system, with all salaries consolidated into a single file (e.g., University of California System). In such cases, we included all universities in the system if they operate as non-branch institutions and offer an economics graduate program. This added four universities, bringing our sample to 53.

We excluded 21 universities, reducing our final sample to 32. Nearly half of these exclusions can be attributed to lack of coverage of the Card et al. (2022) publication data and our scraped salary data in the same year. Other shortcomings in state-published salary data contributed to another significant portion of exclusions. Pennsylvania's PennWATCH, for instance, lacks a university identifier, while Virginia's Data Point does not provide faculty names. Our final sample spans from 2008 to 2019 and includes an average of eight years of salaries per university.

C2 Salary-Publication Merge

Card et al. (2022) compile annual and cumulative publications for active NAS and AAAS researchers in economics, mathematics, and psychology across 42 top-ranked journals. We use STATA's reclink2 function to probabilistically link authors' salaries with their publications, and supplement this with a manual review to eliminate false positive matches.

C2.1 Data Preprocessing

Before matching, we replicate the name standardization procedure outlined in Card et al. (2022). Accented or non-English characters are replaced with the closest standard English characters and special characters, suffixes, and honorifics are removed. We convert names to a standard case format, such that the first letter of each name is capitalized, middle names are reduced to initials, and periods are added after initials (e.g., Janet Louise Yellen becomes Janet L. Yellen). If a name includes an initial succeeded by a fully spelled out middle name, for example, A. James Robinson, the order is swapped to become James A. Robinson. Fuzzy matching on names is preferable in our situation because it allows us to reconcile individuals that may appear under different naming conventions across datasets or over time (e.g., alternative spellings or shortened names like Steve vs. Steven).

In addition to standardizing names, we standardize titles to establish a clear hierarchy among assistant, associate, and full professors, while excluding non-tenure-track positions. Since titles are often abbreviated (e.g., "asst." for assistant), we standardize them to account for inconsistencies in representation across files. We remove part-time (e.g., visiting professors), emeritus, and other non-tenure-track positions (e.g., clinical professors). Table C1 lists all keywords (and their associated abbreviations) used to filter faculty titles.

Table C1: **Title Classification and Exclusion**

Position	**Keywords**
Classification	
Assistant Professor	assistant, assist, assnt, asst, ast
Associate Professor	associate, assoc, asoc
Full Professor	professor, prof
Exclusion	
Administrative roles	dean, chair, head, director (dir), chancellor, provost, president (pres)
Part-time faculty	visiting (vis, vst), adjunct (adj), acting (act), affiliated (affil), emeritus (emer)
Other non-research	lecturer, teach (tch), clinical (clin), practitioner (prac), extension (ext)

Note: Assistant and associate professor classifications also require "professor" and "prof."

We also address variation in how universities report salaries, as well as part-year employment (e.g., resulting from sabbaticals or furloughs). Specifically, 17% of universities separately report base pay and "bonus" pay which captures additional compensation such as summer salary, while 83% report only total pay. We therefore use total pay for our analysis. If one's salary falls by more than 25% from one year to the next at the same university-department, that observation is flagged as part-year employment and dropped. We also drop salaries below the federal minimum wage ($14,500) for similar reasons.

C2.2 Probabilistic Linkage

A challenge at the linkage stage is isolating economics, mathematics, and psychology faculty from the broader pool of staff, given the absence of a department identifier in the salary data. To facilitate within-department matching between salaries and publications, we use the Wayback Machine to retrieve faculty rosters for each university-department combination, drawing from a one-year window around the corresponding salary period. Fuzzy matching between salaries and rosters is then performed on names, excluding middle initials, using a modified bigram string comparator.[34] We omit middle initials because they are not reported consistently, however, a final validation step removes cases where middle initials, if present in both datasets, do not match. University and year matches must be exact, and we impose a minimum match score of 0.95 for a match to be accepted.

Since rosters are unavailable for 10% of university-years in economics, 11% in mathematics, and 9% in psychology, we use data from adjacent years to fill 71% of the gaps in economics, 53% in mathematics, and 56% in psychology. If a faculty member is present in the salary data but absent from the roster for a given university-year, we check whether they appear in both the salary data and the roster (i.e., they match) in either the preceding or following year. When this condition is met, we retain these observations as matches.

Now, within-department, we link salaries with publications, applying the same fuzzy matching approach as before, yet require only an exact match on year (no university identifier in the publication data). We also retain faculty who matched in the previous merge though do not appear in the Card et al. (2022) data. These individuals are classified as "rookie" faculty, with their publications set to zero for missing years. If a faculty member is a rookie in some years but not others, we adjust their publication metrics accordingly.

Not all salary or roster files provide faculty titles. When matching salaries with rosters, we retain titles from rosters to account for any missing or mismatched titles in the salary data. Still, some missing titles remain for the three Indiana universities in our sample, affecting a fraction

[34] See Wasi and Flaaen (2015) for additional details.

of author-year cells: 27% in economics, 66% in mathematics, and 52% in psychology. We drop these observations at the conclusion of the matching process.

C2.3 Gender Imputation

A limitation to the Card et al. (2022) data is the lack of gender indicators for all authors. Post-matching, ungendered authors represent 31% of economics, 31% of mathematics, and 32% of psychology. We infer these genders via the Genderize.io API, which predicts genders based on a dataset of one billion name-to-gender mappings spanning 220 countries. For a given name, the API returns the likely gender and an associated probability, reflecting the proportion of individuals in the dataset with that name who are classified as that gender. We set a threshold of 95% and impute the corresponding gender when the threshold is met or exceeded. Still, some names, like Andrea, whose gender varies by nationality remain ungendered. To resolve these cases, which involve 4% of economics authors, 11% of mathematics authors, and 4% of psychology authors, we conduct a manual review, consulting online sources, such as researchers' websites or university directories.

C3 Citation Record Inclusion

The Scopus citation and abstract database, maintained by the publisher Elsevier, indexes over 24,000 high-quality, peer-reviewed journals. We query this database to supplement the publication data from Card et al. (2022) with annual and cumulative citation records as additional productivity measures. Scopus is particularly advantageous to us because its references date back to 1966, aligning well with the timeframe of the Card et al. (2022) data, which covers publications from 1960 to 2019. We detail our query procedure below.

First, we pass author names (first and last) along with their university affiliations and department codes (e.g., "ECON" for economics) into the Author Search API, which returns IDs linking authors to their publications. If a call returns multiple IDs for a single author, we retain all IDs and handled these duplicate entries in the subsequent step. Next, we feed author IDs into the Author Retrieval API, receiving a list of publication IDs. To avoid double-counting publications caused by multiple IDs assigned to a given author, we remove duplicate publication IDs before proceeding with the final step. We then pass each author's publication IDs into the Citation Overview API with a date range of 1996 to 2019 and extract annual citations per publication. From here, we calculate two citation measures: annual citations per author, aggregated across all journals, and cumulative citations per author, aggregated across all journals and all prior years up to and including the current year. Since Scopus catalogs a much wider range of journals compared to the 42 high-impact ones selected by Card et al. (2022), an implication is

that some (primarily) rookie faculty accumulate citations despite never publishing in any of those journals. We leverage this feature to also compute variables for annual and cumulative non-Card et al. publications.

D Additional Empirical Results

D1 Main Specifications (without Controls)

Table D2: **Differences by Discipline**

	(1) Rational -izable?	(2) Unobservable Cap ($\underline{\rho}$) Distance	(3) Dollar ($\underline{\delta}$) Distance
Economics	.24***	.028	-14,366***
	(.048)	(.074)	(3,808)
Psychology	.019	.006	1,939
	(.054)	(.053)	(4,446)
Math Mean	.31	.17	31147.52
Observations	1,926	1,926	1,755
R^2	.15	.03	.32

Notes: This table reports the results of Equation 18, which examines how pay rationalizability varies by cell characteristics. Math professors are omitted. The outcome measures are defined in Section 5.2. All specifications include school, discipline, year, and title fixed effects. Robust standard errors clustered both at the school×year and the discipline×title levels (as shown in parentheses). Significance levels: * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table D3: **Changes over Time**

	(1) Rational -izable?	(2) Unobservable Cap ($\underline{\rho}$) Distance	(3) Dollar ($\underline{\delta}$) Distance
Year	-.0075*	.012	504**
	(.0041)	(.013)	(225)
Math Mean	.31	.17	31147.52
Observations	1,926	1,926	1,755
R^2	.15	.025	.31

Notes: This table reports the results of Equation 18, which examines how pay rationalizability varies by cell characteristics. The outcome measures are defined in Section 5.2. All specifications include school, discipline, year, and title fixed effects. Robust standard errors clustered both at the school×year and the discipline×title levels (as shown in parentheses). Significance levels: * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

D2 With Additional Controls

For our descriptive results, there are a few cell characteristics that make monotonicity easier to satisfy. For example, it is possible that non-monotonicity is higher among Full Professors simply because there are often more of them in a given department. This creates more inequalities in the linear system. All other things being equal, this narrows the space of monotonic β weights.

Similarly, disciplines or groups that publish in a wider variety of journals will find it easier to satisfy monotonicity. Suppose the highest-paid professor in a group has only one publication in their entire career. However, no other person in the group has ever published in that journal. Monotonicity could be rationalized by placing high weight on that one journal. For these reasons, we present our descriptive result both with and without these Controls_{sdjt}. To control for these patterns flexibly, we use fixed effects for the size of the cell and the total number of journals (and other productivity metrics).

Table D4: **Differences by Seniority (with Controls)**

	Rationalizable?			Unobservable Cap Distance ($\underline{\rho}$)			Dollar Distance ($\underline{\delta}$)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Full Professor	.21***	.025	.11**	-.11**	.045	.036	-15	11,930***	3,611
	(.045)	(.055)	(.045)	(.044)	(.037)	(.042)	(2,999)	(3,645)	(2,923)
Associate	.11***	.08*	.08**	.068	.072	.082	-5,707**	-6,204**	-4,907**
	(.033)	(.041)	(.032)	(.097)	(.066)	(.069)	(2,367)	(2,803)	(2,292)
Assistant Mean	.44	.44	.44	.21	.21	.21	16921.64	16921.64	16921.64
Controls	Size	#J	Both	Size	#J	Both	Size	#J	Both
Observations	1,926	1,926	1,926	1,926	1,926	1,926	1,755	1,755	1,755
R^2	.41	.2	.45	.03	.77	.78	.58	.43	.6

Notes: This table reports the results of the descriptive Equation 18, which examines how pay rationalizability varies by cell characteristics. Assistant professors are omitted. "Size" controls refers to fixed effects for the number of researchers in each cell, and "#J" refers to the number of distinct journals or other publication metrics observed in this cell. The outcome measures are defined in Section 5.2. All specifications include school, discipline, year, and title fixed effects, plus controls for cell size and number of journals. Robust standard errors clustered both at the school×year and the discipline×title levels (as shown in parentheses). Significance levels: * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table D5: **Differences by Discipline (With Controls)**

	Rationalizable?			Unobservable Cap Distance ($\underline{\rho}$)			Dollar Distance ($\underline{\delta}$)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Economics	.048	.17***	.076*	.045	-.1**	-.088*	2,759	-8,693***	1,484
	(.037)	(.042)	(.038)	(.075)	(.046)	(.046)	(3,592)	(2,967)	(3,032)
Psychology	.0016	.0063	.016	.021	-.02	-.021	6,367**	2,847	5,826*
	(.037)	(.048)	(.038)	(.053)	(.046)	(.047)	(2,809)	(3,774)	(2,941)
Math Mean	.31	.31	.31	.17	.17	.17	31147.52	31147.52	31147.52
Controls	Size	#J	Both	Size	#J	Both	Size	#J	Both
Observations	1,926	1,926	1,926	1,926	1,926	1,926	1,755	1,755	1,755
R^2	.41	.2	.45	.03	.77	.78	.58	.43	.6

Notes: This table reports the results of Equation 18, which examines how pay rationalizability varies by cell characteristics. Math professors are omitted. "Size" controls refers to fixed effects for the number of researchers in each cell, and "#J" refers to the number of distinct journals or other publication metrics observed in this cell. The outcome measures are defined in Section 5.2. All specifications include school, discipline, year, and title fixed effects, plus controls for cell size and number of journals where labeled. Robust standard errors clustered both at the school×year and the discipline×title levels (as shown in parentheses). Significance levels: * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table D6: **Changes over Time (with Controls)**

	Rationalizable?			Unobservable Cap Distance ($\underline{\rho}$)			Dollar Distance ($\underline{\delta}$)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Year	-.0056	-.006	-.0059	.014	.0021	.0026	680***	539**	765***
	(.0041)	(.0038)	(.0038)	(.015)	(.002)	(.0021)	(209)	(224)	(229)
Math Mean	.31	.31	.31	.17	.17	.17	31147.52	31147.52	31147.52
Controls	Size	#J	Both	Size	#J	Both	Size	#J	Both
Observations	1,926	1,926	1,926	1,926	1,926	1,926	1,755	1,755	1,755
R^2	.41	.2	.45	.032	.77	.78	.58	.44	.6

Notes: This table reports the results of Equation 18, which examines how pay rationalizability varies by cell characteristics. "Size" controls refers to fixed effects for the number of researchers in each cell, and "#J" refers to the number of distinct journals or other publication metrics observed in this cell. The outcome measures are defined in Section 5.2. All specifications include school, discipline, year, and title fixed effects. Robust standard errors clustered both at the school×year and the discipline×title levels (as shown in parentheses). Significance levels: * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

E Group EPEW Symmetry Violations: Empirical Examples

In the settings below, we empirically measure group-level EPEW and EWEP violations and show that a violation of one of these fairness notions does not imply a violation of the other, as we previously theoretically showed in Table 1. In all settings below, we estimate the following four variants of (2) and (4). In each equation, i indexes workers, and j indexes employers. The first two are group-regression versions of EPEW.

$$w_{ij} = \beta_1 \text{Female}_i + \gamma\theta_{ij} + \delta_j + \epsilon_i, \tag{38}$$

where where w_{ij} indicates worker i's wage at employer j, θ_{ij} is the worker i's productivity at firm j, and δ_j are firm fixed effects. $\text{Female}_i = 1$ for women, and $=0$ for men (all datasets below code only for two genders). In each of the applied settings below, we state how θ_{ij} is measured. This specification is the realized version of the linear productivity approach (4).

We also estimate the iso-productivity approach to EPEW:

$$w_{ij} = \beta_1 \text{Female}_i + \underbrace{\sum_{\theta \in \hat{\Theta}} \gamma_\theta \mathbb{1}(\theta_{ij} = \theta)}_{\substack{\text{Iso-productivity band} \\ \text{fixed effects}}} + \delta_j + \epsilon_i, \tag{39}$$

where the variables have the same meaning. The γ_θ terms refer to iso-productivity fixed effects

(whose implementation we clarify below). We also estimate the equivalent linear and iso/FE versions for EWEP:

$$\theta_{ij} = \beta_1 \text{Female}_i + \gamma w_{ij} + \delta_j + \epsilon_i, \tag{40}$$

and

$$\theta_{ij} = \beta_1 \text{Female}_i + \underbrace{\sum_{w \in \hat{W}} \gamma_w \mathbb{1}(w_{ij} = w)}_{\substack{\text{Iso-wage band} \\ \text{fixed effects}}} + \delta_j + \epsilon_i. \tag{41}$$

The γ coefficient in these is now attached to the wage effects (linearly and with FEs).

For all four regressions above, we use Poisson pseudo-maximum likelihood regressions (PPML, Davies and Guy, 1987; Silva and Tenreyro, 2006; Correia et al., 2020) to facilitate interpretation as percentage increases. We also limit the sample to avoid singletons and keep the same sample across both types of analysis (although the qualitative results are robust to relaxing this limitation). Finally, we cluster standard errors at the evaluator level.

Motivating Example 1 (Pay in Math Jobs) A well-known paper by Bohren et al. (2023, *Review of Economics and Statistics*) collect data on workers' math ability using a standardized test. They use this as a measure of productivity for jobs whose tasks require skills tested on the exam. The authors then collect data from subjects assuming the role of employers, asking them employers to predict workers' performance on the exam, and to propose a wage for each worker.

Table E7 shows patterns using the data from Bohren et al. (2023) similar to Table 1. Bohren et al. (2023) contains a within-subject information treatment. Here, we examine only the pre-treatment data (we examine the effects of the intervention separately in Appendix E1). In columns 1 and 2, we see that women are paid less on average, controlling for their levels of productivity (i.e., test scores). However, in columns three and four, we see that female subjects were less productive (i.e. lower scores) compared to men paid the same.

Table E7: **Equal Work and Pay in Bohren et al. (2023) by Gender**

	(1) Wage	(2) Wage	(3) Productivity	(4) Productivity
Female Worker	-.027***	-.026***	-.087***	-.087***
	(.0062)	(.0065)	(.0045)	(.0045)
Control	Productivity	Productivity (FEs)	Wage	Wage (FEs)
Observations	11,507	11,507	11,507	11,507

Notes: This table contains the results from specifications (38) through (41) applied to the replication data from Bohren et al. (2023). Columns 1 and 2 predict wages conditional on productivity, and columns 3 and 4 prediction productivity conditional on wages. Standard errors in parentheses, clustered at the evaluator level. All columns include fixed effects for the "employer ID" (i.e., the subject id of the subject performing the evaluation). * p<0.1, ** p<0.05, *** p<0.01.

Motivating Example 2 (Algorithmic Pay) Cowgill et al. (2020) conduct a similar exercise as Bohren et al. (2023). Instead of administering a math test to new subjects, the authors use the results of the PIAAC standardized exam administered to 200K subjects by OPEC.

Rather asking employers propose wages subjectively, the authors work with 400 machine learning engineers to develop a predictive model for 20K test subjects. Engineers were given test results from other subjects as training data, along with rich biographical details about each subject in the training set. Rather than subjectively predicting their math scores, the engineers developed prediciton models from the biographical covariates. We call the engineers' predictions the "pay index."

Table E8 presents group EPEW and group EWEP comparisons for the highest quartile of math subjects. We document group EPEW violations in the pay index (on average) favoring men, and group EWEP violations (on average) favoring women.

Table E8: **Equal Work and Pay in Cowgill et al. (2020) by Gender**

	(1) Pay Index	(2) Pay Index	(3) Productivity	(4) Productivity
Female Worker	-13***	-13***	-.056***	-.033***
	(.87)	(.86)	(.000011)	(.001)
Control	Productivity	Productivity (Quintile FEs)	Pay Index	Pay Index (Quintile FEs)
Observations	897,242	897,242	897,242	897,242

Notes: This table contains the results from specifications (38) through (41) applied to the data from Cowgill et al. (2020). Columns 1 and 2 predict wages conditional on productivity, and columns 3 and 4 prediction productivity conditional on wages. Standard errors in parentheses, clustered at the evaluator level. All columns include fixed effects for the engineer ID (i.e., an identifier for the engineer performing the evaluation). * p<0.1, ** p<0.05, *** p<0.01.

Motivating Example 3 (Economics Faculty Research) In Table E9, we examine the salaries and research productivity of economists using data compiled for this paper. Research productivity metrics were complied by the authors based on Card et al. (2022) and other sources (described in Section 5), and merged with publicly-available salary information following the procedure in Appendix C. We use cumulative citations as the productivity metric. For the employers, we use school $\times$ year fixed effects. For the iso-productivity fixed effects, we use within-cell deciles.

The two regressions use the same data, but reach different conclusions depending on which variable is tested for equality, and which is used as a conditioning control variable. In Columns 1 and 2 of Table E9, we show that female faculty are paid less than men with the same number of cumulative citations. If cumulative citations were the key productivity metric, women as would have a group-level EPEW complaint. Columns 3 and 4 show that female faculty during this period on average have fewer cumulative citations compared to men with the same salary. If cumulative citations were the key productivity metric, men would have a group-level EWEP complaint. □

Table E9: **Equal Work and Pay in Academic Citation Data, by Gender**

	(1) Salary	(2) Salary	(3) Cumulative Citations	(4) Cumulative Citations
Female Professor	-.02*	-.018*	-.36***	-.27***
	(.012)	(.011)	(.054)	(.055)
Control	Cumulative Citations	Cumulative Citations (Decile FEs)	Salary (Linear)	Salary (Percentile FEs)
Observations	3,167	3,167	3,167	3,167

Notes: This table contains the results from specifications (38) through (41) applied to the data assembled for this paper in Section 5 and Appendix C, based on Card et al. (2022). Columns 1 and 2 predict wages conditional on productivity, and Columns 3 and 4 prediction productivity conditional on wages. Standard errors in parentheses, clustered at the evaluator level. All columns include fixed effects for of the employer (school) $\times$ year. * p<0.1, ** p<0.05, *** p<0.01.

E1 Treatment Effects on Equal Work and Pay

Remark 2 uses a counterexample to show that interventions aiming to improve group EPEW could possibly have the an opposite effect on group EWEP (and vice versa). In this section we study an experimental interventions taking place in the Bohren et al. (2023) and Cowgill et al. (2020) data. In both cases, the interventions increase violations of group EPEW. By contrast, we find no detectable improvement in group EWEP; if anything, the point estimates weakly suggest the opposite.

Pay in Math Jobs. In the Bohren et al. (2023) study of math tests, subjects in the experiment were given an informational intervention: They were shown data summarizing how average performance varied by age, gender, and nationality. The authors find that employers significantly changed their wage offers on the basis of gender and nationality. We now show how this had different effects on group EPEW and group EWEP.

In Table E10, we run specifications (38) through (41) with the Bohren et al. (2023) data. We add an interaction between Female_i and the intervention. In Columns 1 and 2, the interaction is negative and statistically significant, so the intervention makes the female wage coefficient more negative and therefore increases violations of group EPEW. In Columns 3 and 4, the interaction is positive but imprecisely estimated, so we find no detectable improvement in group EWEP; if anything, the point estimates suggest a slight worsening.

Table E10: **Treatment Effects on Equal Work and Pay (Bohren et al., 2023, by Gender)**

	(1) Wage	(2) Wage	(3) Productivity	(4) Productivity
Female Worker	-.029***	-.029***	-.088***	-.088***
	(.0062)	(.0063)	(.0044)	(.0044)
Female Worker × Post-Intervention	-.021**	-.02**	.0092	.0089
	(.0087)	(.0087)	(.0081)	(.008)
Control	Productivity	Productivity (FEs)	Wage	Wage (FEs)
Observations	17,280	17,280	17,280	17,280

Notes: This table contains the results from specifications (38) through (41) applied to the replication data from Bohren et al. (2023) (with treatment interactions). Columns 1 and 2 predict wages conditional on productivity, and Columns 3 and 4 prediction productivity conditional on wages. Standard errors in parentheses, clustered at the evaluator level. All columns include fixed effects for the "employer ID" (i.e., the subject id of the subject performing the evaluation). * p<0.1, ** p<0.05, *** p<0.01.

Algorithmic Pay. The aforementioned Cowgill et al. (2020) study of programmers also contained an intervention. In the control group, programmers were given data featuring a natural sample section problem: The training data consisted of people who were already employed in math-intensive jobs. This mimics real-world problems in resume screening algorithms, where the only resumes with productivity labels are the non-random subsample able to gain employment at the target firm.

As an experimental intervention, some programmers were given a representative sample (rather than a selected sample). This was made possible by PIAAC's careful representative sampling and weighting. There were also two other experimental interventions within the "selected sample" condition, attempting to educate and inform the programmer about the potential for algorithmic bias. Neither of these had any effect, and thus we merge the with the control in the results that follow.

Table E11 reports the results of this randomized intervention using specifications (38) through (41) with the Cowgill et al. (2020) data. We add an interaction between Female_i and the intervention. Again, we find a large effect negative effect on group EPEW, but no detectable effect on group EWEP; if anything, the point estimates weakly suggest the opposite.

Table E11: **Treatment Effects on Equal Work and Pay in Cowgill et al. (2020) by Gender**

	(1) Pay Index	(2) Pay Index	(3) Productivity	(4) Productivity
Female Worker	-.29***	-.18***	-.056***	-.033***
	(.0079)	(.0079)	(.000011)	(.0011)
Female Worker x Intervention	-14***	-14***	.000032	.00085
	(.85)	(.84)	(.000035)	(.0015)
Control	Productivity	Productivity (Quintile FEs)	Pay Index	Pay Index (Quintile FEs)
Observations	897,242	897,242	897,242	897,242

Notes: This table contains the results from specifications (38) through (41) applied to the data from Cowgill et al. (2020). Columns 1 and 2 predict wages conditional on productivity, and Columns 3 and 4 prediction productivity conditional on wages. Standard errors in parentheses, clustered at the evaluator level. All columns include fixed effects for the engineer ID (i.e., an identifier for the engineer performing the evaluation). * p<0.1, ** p<0.05, *** p<0.01.

References

Bohren, J Aislinn, Kareem Haggag, Alex Imas, and Devin G Pope, "Inaccurate statistical discrimination: An identification problem," *Review of Economics and Statistics*, 2023, pp. 1–45.

Card, David, Stefano DellaVigna, Patricia Funk, and Nagore Iriberri, "Gender differences in peer recognition by economists," *Econometrica*, 2022, *90* (5), 1937–1971.

Charnes, Abraham and William W Cooper, "Programming with linear fractional functionals," *Naval Research logistics quarterly*, 1962, *9* (3-4), 181–186.

Cormen, T, C Leiserson, R Rivest, and Clifford Stein, "introduction to algorithms," *algorithms*, 2009, *1* (1L9K1), 1L9K1.

Correia, Sergio, Paulo Guimarães, and Tom Zylkin, "Fast Poisson estimation with high-dimensional fixed effects," *The Stata Journal*, 2020, *20* (1), 95–115.

Cowgill, Bo, Fabrizio Dell'Acqua, Samuel Deng, Daniel Hsu, Nakul Verma, and Augustin Chaintreau, "Biased programmers? Or biased data? A field experiment in operationalizing AI ethics," in "Proceedings of the 21st ACM Conference on Economics and Computation" 2020, pp. 679–681.

Davies, Richard B and Clifford M Guy, "The statistical modeling of flow data when the Poisson assumption is violated," *Geographical Analysis*, 1987, *19* (4), 300–314.

Knuth, Donald Ervin, *The art of computer programming*, Vol. 3, Pearson Education, 1997.

Silva, JMC Santos and Silvana Tenreyro, "The log of gravity," *The Review of Economics and statistics*, 2006, pp. 641–658.

Wasi, Nada and Aaron Flaaen, "Record linkage using Stata: Preprocessing, linking, and reviewing utilities," *The Stata Journal*, 2015, *15* (3), 672–697.