

Testing for Discrimination in Law Firms: An Outcome Test Approach*

Emma Chiu[†]

Stéphane Mechoulan[‡]

October 13, 2026

Preliminary and incomplete. Please do not circulate.

Abstract

We investigate whether discrimination could account for the well-documented underrepresentation of certain groups among law firm partners. Leveraging the JD dataset and an observational variant of the outcome test, we examine disparities in promotion rates by gender as well as age, parenthood, and law school rank. We find that marginally promoted female partners are less productive than their male peers in terms of billable hours; under the assumptions of the outcome test framework, this observation is inconsistent with a higher promotion threshold for women. Women nevertheless report substantially more workplace discrimination across the legal profession and at each point in the survey. The test also suggests bias in favor of older lawyers along the dimension of new matters brought to the firm in a majority of specifications. Other comparisons are inconclusive.

JEL codes: J16, J44, K40.

Keywords: Lawyers; Discrimination; Promotion; Outcome Test.

*We thank Dino Gerardi, Nicola Persico, Nayoung Rim, Ricardo Perez-Truglia, conference participants at the Southern Economic Association Conference, the Canadian Law and Economics Association Conference, the French Law and Economics Association conference, and seminar participants at IDHEAP Lausanne, Université Paris I Panthéon Sorbonne, Université Paris X Nanterre, Queens University, University of Tokyo, the International University of Japan, Hitotsubashi University, BITS Pilani Hyderabad, NIPFP Delhi, and Kyoto University for helpful comments. We are grateful to Herman Lange, who provided excellent research assistance. All errors and omissions are ours.

[†]Department of Economics, Northwestern University. Email: emmachiu2024@u.northwestern.edu.

[‡]Department of Public and International Affairs and Department of Finance, Dalhousie University. Email: s.Mechoulan@dal.ca.

1 Introduction

The existence of barriers preventing women and minorities from advancing to managerial and executive positions has been acknowledged for decades. In 1995, the Glass Ceiling Commission released a report detailing the types of obstacles faced by women and minorities, as well as ways to correct the situation (Commission, 1995). According to Haile et al. (2016), for example, women’s career paths are barraged with predicaments such as organizational culture, gender stereotypes, failure to capitalize on female talents, differences in national culture, sexism, relegation to human resources positions, gender differences in pay, and political powers. This pattern is particularly persistent in the legal profession, where women and minorities are found in substantially lower proportions at the partner level than what could be expected from their early career representation (National Association for Law Placement, 2019; England et al., 2020), despite industry-level commitments to systemic change (Lee and Liu, 1999).

This paper examines whether discrimination could explain this phenomenon. To that end, we use a rich source of information, the After the JD (AJD) dataset, which tracks more than four thousand lawyers over the first two decades of their careers. Our primary objective is to detect whether certain lawyers are held to a different standard when the decision of promotion from associate to partner is being made, and if so, to what extent. For that purpose, we turn to a variant of the outcome test methodology pioneered by Becker (1957). This test is meant to measure animus, or taste-based discrimination. Specifically, we apply the theoretical framework developed in Persico (2009), which addresses the problem of infra-marginality that threatens the validity of outcome tests. That is, observed average productivity differences could be accounted for by genuine differences between groups, undetectable when we cannot isolate individuals treated at the margin. This can lead to misleading conclusions regarding prejudice.

Our work extends an abundant literature that studies obstacles and systemic barriers for women and minorities that not only slow down their early career progress but also create a precedent for bigger gaps as careers advance (Bertrand and Hallock, 2001; Blau and Kahn,

2017; Adams and Funk, 2012).¹ The branch of this literature that our work connects with most closely concerns the representation and roles of women and minorities in the legal profession and predominantly draws from AJD data.

Chanow and Sterling (2021) and Sterling and Reichman (2012) use AJD data to show how systemic barriers and firm culture make women and minorities leave before reaching the partner level. Likewise, Payne-Pikus et al. (2010) and Park and Rim (2020) analyze how gender, maternity status, and racial factors within law firms affect the promotion rates and personal lives of lawyers. Antecol et al. (2014) expand on these studies by investigating the disconnect between reported experiences of bias and the conventional measures of earnings discrimination in the legal profession. They suggest that traditional measures may not fully capture the racial and gender biases that new lawyers report. By exploring lawyers' productivity, mainly captured by hours billed, our study shows a more nuanced understanding of discrimination beyond mere pay gaps.

Three recent studies have made use of the AJD to assess the interplay between gender and career performance. Azmat and Ferrer (2017) examine the performance differences between female and male lawyers and attribute a significant part of the gap to the differential impacts of having young children and variation in partnership aspirations. In subsequent work, Azmat et al. (2020) show how career aspirations at the associate level play an important role in the observed differences in partnership promotions between genders. Lehmann (2011) offers a different perspective by showing that black lawyers are more likely than similarly credentialed white peers to be hired into top law firms, but also more likely to be assigned tasks that are unlikely to lead to promotion.

We make two key contributions within this literature. First, we directly build on Lehmann (2011), which uses the first two waves of the AJD. By the end of the second wave, few associates had been evaluated for promotion. However, Kay and Gorman (2012) show that minority presence at the partner level depends on the duration of the partner evaluation process. Our use of the third wave of AJD data provides a more comprehensive picture of

¹See Bertrand (2011) for a review on the broader spectrum of gender disparities in the workforce, connecting the dots between pay gaps, leadership positions, and discrimination.

the promotion process in the legal profession.

Second, we complement [Azmat et al.](#)’s work by leveraging the outcome test approach to conduct an objective performance assessment of discrimination in law firms. This technique has been used in the context of drug searches ([Ayres, 2002](#); [Knowles et al., 2001](#); [Anwar and Fang, 2006](#)), parole board decisions ([Mechoulan and Sahuguet, 2015](#)), bail decisions ([Arnold et al., 2018](#); [Canay et al., 2024](#); [Grau and Vergara, 2025](#)), and scholarly achievements [Johannesen and Muchardt \(2026\)](#), to cite a few. Some of these tests leverage an institutional feature or game-theoretical argument to establish that the average outcome test can be construed as a marginal outcome test. We employ a novel empirical approach to solving the infra-marginality problem: the Prediction-Based Outcome Test (P-BOT) ([Grau and Vergara, 2025](#)).

Applying this methodology to our context involves ranking promoted associates according to their predicted probabilities of promotion to define samples of marginal partners, and then comparing group-specific productivity measures for individuals near the threshold of promotion. Unlike approaches that recover the margin using quasi-experimental variation, the P-BOT uses observed promotion probabilities to construct the marginal sample and permits dependence between observed and latent determinants of selection.

The dataset that we use provides an excellent chance of overcoming the infra-marginality challenge for two main reasons. First, the legal profession, unlike many other high-skilled domains, has established quantified ways to measure productivity — mainly, billable hours and number of new matters or clients brought to the firm — which directly determine law firms’ revenues. Second, given the 6-to-9 year duration of the partnership evaluation process, all or nearly all partner-track lawyers should have gone through the promotion (typically up or out) process by Wave III of the AJD. Additionally, the discrimination literature has for the most part examined a single outcome, usually binary, whereas the AJD enables us to consider multiple, continuous outcomes. Due to the small sample sizes in the AJD, in the benchmark model, we define marginal individuals coarsely as those who are in the bottom 30% of predicted promotion propensity among those promoted.

Our selection equation regression results confirm that number of billable hours and number of new matters brought to the firm are the strongest predictors of promotion, alongside desire for promotion. In addition, while being a woman is, *ceteris paribus*, significantly negatively correlated with promotion in some specifications of the model, being a mother is significantly positively correlated with promotion. Other immutable attributes (race/ethnicity, year of birth, and law school rank) are less clearly related to promotion. These results are not interpreted as evidence of discrimination or lack thereof.

Even with low power, the P-BOT reveals, at the margin, evidence consistent with a lower effective promotion threshold for women along the dimension of partner billable hours, as well as for older lawyers along the dimension of new matters brought to the firm. Comparisons along other dimensions yield no consistent, significant differences.²

Bias, in the context of gender, could be interpreted in a number of ways: an accommodation for an expected temporary lower productivity consistent with the early years of maternity, anticipated or current, or a preference for a diverse workforce, neither of which we can test. Regardless, we find no clear evidence of taste-based discrimination against women in the narrow context of the promotion to partner stage.

An important qualification is that the outcome test requires post-promotion performance to be informative of partners' underlying productivity. If promoted partners faced group-specific post-promotion treatment that systematically depressed their productivity, such as through the allocation of clients, assignments, or other revenue-generating opportunities, results may be misleading. We conduct several checks designed to assess this possibility. First, we re-estimate the P-BOT among partners who report experiencing no discrimination and obtain similar or larger gender gaps in billable hours. We also find that reported discrimination is not concentrated among women near the promotion margin, implying that any post-promotion suppression should be applied to women similarly; however, the gender gap is smaller, rather than larger, away from the margin, inconsistent with broad post-promotion suppression. Moreover, within propensity-score quintiles, women do not experience a signifi-

²We could not run a P-BOT comparing performances between racial groups given the too small number of minority partners at the margin. Similarly, we could not compare the performance of mothers against non-mothers within the subsample of marginally promoted lawyers.

cantly different change in billable hours from before to after promotion than men, providing additional evidence against a targeted post-promotion suppression. Lastly, based on examinations of total hours worked, pro-bono work, and law firm characteristics, we find no evidence to suggest differences in post-promotion responsibilities between men and women. While these tests cannot rule out all forms of post-promotion differential treatment, they provide evidence against such treatment being the primary explanation for our findings.

Our main result is also robust to marginal sample selection criteria, though conclusions of some tests are limited due to small sample sizes. We first perform a series of rank stability diagnostic tests to determine whether the ordering of individuals — especially those identified as “near the margin” — remains consistent when we account for the uncertainty in the estimated propensity scores. We examine sample stability and monotonicity tests, and we run a joint confidence interval test for rankings as suggested in [Mogstad et al. \(2022\)](#). While we find support from the first two tests, the latter is uninformative given our sample size. Second, we perform a simulation-based perturbation test, which is designed to measure the sensitivity of our estimates to potential misclassifications of marginal partners arising from uncertainty in the propensity score estimates. We find support for our conclusions from this test as well. Non-marginal outcome tests, based on larger swathes of the distribution of partners up to the entire sample, produce stronger results concurring with our main estimates. Lower marginal sample cutoffs, such as 20 and 10 percent, return coefficients of similar sign and roughly similar magnitude; however, estimates become noisy due to small sample sizes. We obtain estimates of similar direction and magnitude when using gender-specific selection equations and definitions of marginal partners. Thus, cumulatively, from different measures of productivity; from marginal and non-marginal outcome tests; and from different sets of assumptions over the promotion cutoff, our tests suggest that the evidence is consistent with some bias in favor of women and older lawyers, though none related to parental status or law school prestige.

The outcome test does not rule out the possibility that discrimination occurred prior to the promotion decision, but such prior discrimination would not invalidate the test. Never-

theless, to explore this possibility, we run two additional tests. First, we determine whether perceptions of discrimination are correlated with desire for promotion. Second, we determine what factors are more likely to lead associates to leave the private law firm environment. One cannot draw unambiguous conclusions based on those tests, but if one category of associates perceives discrimination relatively more, and perception of discrimination leads to lower desire for promotion or more early departures, this could suggest (unless perceptions have no basis in reality) that associates vying for promotion are a self-selected sample that may have survived a discriminatory environment. Thus, absence of evidence of prejudice among promotion-eligible associates would not imply absence of prejudice for all.

Desire for promotion and perception of discrimination are not significantly correlated. In addition, we find no statistically significant association between reported discrimination and subsequent departure from the private law firm environment, reducing concerns about endogenous attrition. Finally, we examine whether lawyers' perceptions of discrimination line up with the conclusions of an 'objective' analysis and whether promotion impacts satisfaction over workplace diversity. Generally, women tend to perceive more discrimination across all types of legal employment. Moreover, the private law firm environment is more conducive to perceptions of discrimination than other areas of law. However, we do not find evidence that promotion leads to a reduced perception of discrimination or positive changes in one's satisfaction over workplace diversity.

The remainder of the paper proceeds as follows: Section 2 provides further details on the AJD dataset, Section 3 outlines the conceptual framework underpinning our model, and Section 4 describes our empirical methods. Section 5 presents our results, and Section 6 concludes.

2 Data

We use data from the After the JD (AJD) project developed by the American Bar Foundation. The AJD has gathered a rich set of data covering the experiences of a nationally representative cross-section of U.S. law graduates admitted to the bar in 2000 over the first twelve years of

their careers. The AJD’s three publicly accessible survey waves (2004, 2009, 2014) contain time-invariant characteristics at the baseline survey (e.g., race, gender, law school rank, undergraduate and law school GPAs, etc.) plus time-varying attributes (e.g., presence of children, marital status, firm size and location, number of positions held, length of stay at a position, number of new matters or clients brought to the firm, etc.), allowing us to track the career trajectories of a large sample of lawyers with diverse background characteristics.³

We impose a few sample restrictions. In our main specifications, we restrict eligibility to lawyers who were associates in Wave II because we want to make sure that they have had the opportunity to apply for promotion.⁴ We also drop individuals for whom we do not observe Wave III position and relevant productivity data, as well as individuals with missing data on other key characteristics.

In our primary analysis, we measure post-promotion productivity with annual billable hours, which we label as the ‘Billable Hours Sample’. We additionally measure productivity by total hours worked in the past week (‘Total Hours Sample’) and annual new matters or clients brought to the firm (‘New Matters Sample’). Imposing these data restrictions, our samples contain 269, 470, and 288 individuals, respectively, with variations coming from differing rates of missingness in productivity metrics. We refer to these samples as ‘full sample’ in the text and define ‘promoted samples’ of those who become partners by Wave III; promoted sample sizes are 169, 196, and 170, respectively. We further define a ‘marginal sample’ that consists of promoted individuals whose estimated promotion propensity lies in the bottom 30% of the propensity score distribution.⁵ In other words, these are the partners who ‘barely made the cut’. Table 1 provides descriptive statistics of various baseline characteristics for the full sample, promoted sample, and the marginal sample created for

³The dataset does not allow for distinctions with respect to gender identities beyond the female-male binary, and the sample size appears too small to extend the analysis to sexual orientation. In addition, for our purposes, the sample size of racial and ethnic minority law firm partners is small.

⁴A lawyer is considered promoted if they move from associate to an equity or a non-equity partner position. Unfortunately, because our sample size is already small for partners, we cannot exclude non-equity partners (the sample would be almost halved). We include in the sample associates who declare a low desire to be promoted, almost all of whom were not promoted. Our results are not sensitive to the exclusion of these individuals.

⁵Table 3 presents the marginal effects of a probit model for the determinants of promotion using our full sample.

the purpose of analyzing the billable hours outcome.⁶ We observe that about 40 percent of lawyers are female. Lawyers in the marginal sample tend to be older and less racially diverse, and they have fewer billable hours and new matters than in both the full and promoted samples.

3 Conceptual Framework and Testable Hypotheses

3.1 Promotion decisions and group-specific thresholds

Our empirical approach draws on the outcome-test framework of [Persico \(2009\)](#) and the prediction-based implementation developed by [Grau and Vergara \(2025\)](#). We adapt these ideas to partnership decisions in law firms, where our object of interest is whether the standard for promotion differs systematically across groups.

The following summarizes the model described in [Appendix E](#). We model a law firm as choosing which associates to promote subject to a limited number of available partnership positions. The firm evaluates associates using a richer set of information than the econometrician observes, including both observed and unobserved characteristics related to future value to the firm. Let an associate’s group membership capture the characteristic along which discrimination may operate (e.g., gender or race), and let post-promotion performance proxy for the firm’s legitimate productivity-related objective. If promotion decisions are made solely on the basis of expected future value, then the associates who are just barely promoted should, on average, have similar subsequent performance across groups. By contrast, if one group faces a stricter effective promotion threshold than another, then the associates from that group who are promoted at the margin should exhibit stronger ex-post performance.

This intuition can be expressed thus. For each associate, the firm compares the expected value of promotion to an objective promotion threshold. Promotion occurs when expected value exceeds that threshold. In this framework, prejudice operates by shifting the threshold applied to members of a given group, so that two associates with the same expected value may

⁶Summary statistics split by gender and for the total hours and new matters samples are presented in [Table 2](#), [Table A1](#), and [Table A2](#), respectively.

face different promotion decisions depending on group membership. The relevant comparison is therefore not between the typical partners belonging to two groups, but rather between partners whose promotion decisions lay *at or near the margin*.

This distinction is important because bias is only behaviorally relevant for individuals whose treatment status can be changed by the threshold. Associates who would be promoted under any plausible standard are not informative about discriminatory treatment, nor are associates who would never be promoted. The outcome test is therefore informative only to the extent that it isolates individuals whose promotion decision was genuinely close.

3.2 The inframarginality problem

This observation leads to the main identification challenge: *inframarginality*. In practice, the econometrician does not observe which promoted associates were barely promoted and which were comfortably above the threshold. As a result, comparing average post-promotion outcomes among all promoted associates may be misleading.

To see why, suppose one group has a thicker upper tail of very strong candidates. That group may thus exhibit higher average post-promotion performance even if the firm applies identical promotion thresholds to all groups. Conversely, even if one group faces a stricter promotion standard, its average post-promotion productivity could be lower – for example, if it has fewer extremely productive members – which would lead a simple comparison of means to fail to detect discrimination or even suggest the wrong conclusion.

What matters instead is the performance of promoted associates who were close to the promotion cutoff. Under the null hypothesis of no group-specific prejudice, the expected post-promotion performance of these marginally promoted associates should be similar across groups. Under differential treatment, the group held to a stricter standard should display stronger subsequent performance at the margin. The empirical challenge is thus to identify, within the sample of promoted associates, those who are most likely to have been promoted narrowly.

3.3 Prediction-based identification of marginal partners

We follow the observational implementation proposed by [Grau and Vergara \(2025\)](#) to approximate this unobserved margin. Rather than attempting to recover the firm’s actual cutoff, the econometrician can use pre-promotion observables to estimate how likely each associate was to be promoted. Associates who subsequently become partners can then be ranked according to predicted promotion probability to approximate proximity to the margin.

Formally, let promotion be determined by a latent index:

$$R_i = \mathbb{1}\{R_i^* \geq 0\}, \quad (1)$$

where $R_i = 1$ denotes the promotion of associate i to partner, and R_i^* captures the difference between the associate’s expected value to the firm and the effective threshold applied by the firm. Associates with $R_i^* = 0$ are marginal in the sense that a small change in the firm’s evaluation or threshold would reverse the promotion decision.

Because R_i^* is unobserved, we estimate instead the promotion propensity

$$p(X_i) = \mathbb{E}[R_i \mid X_i], \quad (2)$$

where X_i contains pre-promotion observables available to the econometrician. The intuition is straightforward: among associates who were promoted, those with relatively low predicted probabilities of promotion are more likely to have been close to the threshold than those whose promotion was highly predictable. Restricting attention to promoted associates in the lower tail of the estimated promotion-probability distribution therefore provides an observational approximation to the set of marginally promoted associates.

Rather than relying on experimental or quasi-experimental variation, this prediction-based strategy’s identifying content comes from ranking promoted associates by predicted proximity to the threshold and comparing subsequent outcomes within that near-margin sample. The approach remains valid even when the econometrician does not observe all variables used by firms in promotion decisions, provided the estimated promotion propensity

is informative about relative closeness to the threshold.⁷

This logic yields the core empirical prediction of the paper. Under the null of no group-specific prejudice in promotion thresholds, promoted associates who are observationally close to the margin should exhibit similar post-promotion performance across groups. If, instead, one group is held to a stricter standard, then among those promoted near the margin, that group should outperform the comparison group after promotion. In our application, we implement this test by estimating promotion propensities using pre-promotion characteristics, identifying promoted associates with the lowest predicted promotion probabilities, and comparing their subsequent performance across groups.

Appendix E presents the formal framework underlying this argument, including the treatment-allocation problem, the mapping from group-specific thresholds to expected outcomes at the margin, and the assumptions under which the prediction-based outcome test provides an informative approximation to the marginal outcome test.

4 Empirical Methodology

4.1 Prediction model

To identify marginal individuals we follow [Grau and Vergara \(2025\)](#)’s ranking procedure. That is, we run a `Probit` model with promotion to partner as the outcome of interest, as a function of lawyers’ observable characteristics. Specifically, we consider a model with the following underlying function:

$$\text{Promotion}_i = \Phi(\alpha_0 + \alpha_1 \text{Prod}_i + \alpha_2 X_i + \alpha_3 Z_i) \quad (3)$$

⁷In particular, associates classified as close to the margin based on observables must provide a valid approximation to truly marginal associates, the promotion rule must admit the required separability and monotonicity structure linking observables and unobservables, and the unobserved component must satisfy a homoskedasticity restriction or a weaker monotonicity restriction permitted under heteroskedasticity. Under these conditions, lower predicted promotion probabilities among promoted associates can be used to rank individuals by their proximity to the promotion threshold. We state these assumptions formally in Appendix E.

where Promotion_i is the probability that a lawyer i is promoted from associate to partner, Prod_i is a vector of productivity measures (billable hours over the past year, new matters brought to the firm, and total hours worked over the past week), X_i is a vector of lawyer characteristics, such as gender dummy, age dummy (born before or after the median sample birth year of 1972), race dummy (white or Asian vs. others), law school prestige dummy (top 20 vs. not), parent dummy, and parent-female interaction dummy, and Z_i are law firm geographical locations.⁸

The **Probit** serves as our benchmark prediction model, but we examine whether the resulting ranking depends materially on choice of model. Alternative specifications include a linear probability model, a Lasso-based linear specification (considering interactions and squared terms), and a heteroskedastic **Probit**. The substantive results are similar across these approaches, so the discussion in the main text focuses on the baseline specification; Appendix B reports the alternatives.

In constructing this selection model, we are aware that women’s professional achievements may not be properly measured. Specifically, the literature (public choice, experiments) consistently shows that women spend more time in collegial endeavors or in the pursuit of collective goals (perhaps less valued by firms) (see, e.g., [Croson and Gneezy \(2009\)](#)) when our main productivity measures are individual-based. They also disproportionately contribute to low-promotability tasks ([Babcock et al., 2017](#)). However, while we cannot rule out that this pattern is at play ahead of the promotion process, our test is not meant to capture it and would not be invalidated by it. The propensity score model is solely intended to produce an informative ordering of associates by promotion likelihood; it should identify those close to the margin but need not explain why.

Second, the primary outcome test assumes that female partners are not prevented from working at their full potential. This is analogous to the no-discrimination assumption by the police and parole officers in the parole release context ([Mechoulan and Sahuguet, 2015](#)). In particular, we posit that the number of hours billed by partners cannot be systematically

⁸We omit variables with too many missing observations that would only marginally increase the fit of the selection equation (3) (law school GPA, area of law, law firm size, number of attempts at the bar exam, etc.) and would make our sample size drop further.

affected by group-specific treatment, in which case an observed productivity difference could reflect post-promotion discrimination rather than discrimination in the promotion threshold itself.

To address objections to this non-directly testable hypothesis, we conduct multiple complementary checks. First, we run the analysis on the subsample of partners who report no discrimination. If some partners felt that their number of billable hours had been arbitrarily reduced or if their work assignments were designed to prevent them from working their desired number of billable hours, or that their work was otherwise undervalued, we would expect restricting the sample to exclude these individuals would attenuate a gap generated by such treatment. This approach can also be substantiated by other subjective measures related to personal satisfaction on the job. Second, we examine the relation between reported discrimination and promotion propensity and repeat the analysis on the full sample of promoted partners to assess whether gaps are consistent with a group-specific post-promotion suppression narrative. We also compare changes in billable hours pre- and post-promotion across genders and propensity quintiles; if post-promotion suppression disproportionately reduced women’s productivity, we would expect women to exhibit a larger decline, or smaller increase, in billable hours following promotion. Finally, we investigate whether women’s post-promotion tasks may differ: we repeat the analysis using total number of hours worked in the past week, as workplace discrimination could impact recorded billable hours through the allocation or valuation of work without necessarily reducing the total amount of time a partner actually works, and also explore pro-bono work and law firm characteristics.

In the same vein, it is worth noting the coefficients in equation (3) should not themselves be used to gauge discrimination because the estimation of α_2 is susceptible to omitted-variable bias. An apparent significant effect of gender or race, for example, may be accounted for by variables (unobserved by the researcher) that are omitted from the regression and that positively correlate with gender or race, such as parental status, socioeconomic status, professional network, and so forth. Thus, the coefficients only serve to determine the relative weights associated with each predictor and generate a ranking of promotion probabilities.

The validity of the resulting ordering rests on the equivalent of the assumptions outlined in [Grau and Vergara \(2025\)](#) and further detailed in [Appendix E](#). Roughly, post-promotion outcomes must change smoothly as associates approach the true promotion threshold from above; the selection rule must admit an additively separable relationship between observed and unobserved determinants of promotion; and the residual variation must satisfy the conditional-variance restrictions needed for lower propensity scores to convey useful information about proximity to the boundary. We examine empirically whether the behavior of our prediction model is consistent with the implications of these conditions.

Propensity score fit. [Table 3](#) reports the baseline **Probit** estimates. We evaluate the resulting accuracy and stability of this model, as a better fit limits the scope of local rank reversals and, thus, inframarginality bias. Using (0.5) as a classification rule, at least 75.1% of observations in the billable hours sample are classified correctly, with corresponding rates of 78.2% for women and 72.7% for men. Predictions are slightly better among those promoted. Performance is similar in the new matters sample and slightly worse among the total hours sample. An out-of-sample cross-validation exercise produces comparable classification rates.⁹

Which observables matter the most? In order to gauge each covariate’s impact on the model’s conditional variance and examine the scope for inframarginality bias, we next ask whether our ranking is robust to the removal of predictors from the initial **Probit** model. That is, we recompute promotion propensities after removing particular predictors from X_i and examine the change in composition of the marginal sample. The composition of the marginal sample is generally stable; most lawyers classified as marginal under the baseline specification remain in that set across specifications. This result is likely driven by a high correlation between the elements of X_i , which is reassuring: no single observed covariate appears responsible for defining the marginal sample. We report the complete exercise in [Appendix D](#).

⁹For cross-validation exercise, we repeatedly estimate the **Probit** using a randomly selected 90% of the sample and use the remaining 10% for evaluation. Across 50 repetitions, 76.5% of the billable hours sample cases, 68.9% of the total hours sample cases, and 72.7% of the new matters sample cases are classified correctly. Additionally, to check the stability of the marginal sample, we apply the rank-inference procedure of [Mogstad et al. \(2022\)](#); however, due to sample limitations, we cannot determine marginal rankings with sufficient significance. The resulting bounds are wide enough that we are unable to draw conclusions.

Testing for monotonicity. The separability condition described in Appendix E implies restrictions on how observables should relate to promotion and expected productivity. We assess this implication by comparing estimated relationships across subsamples designed to differ in their unobserved characteristics. For each of the three outcome samples, we compare the coefficients of regressions of R_i on observables, as well as post-promotion productivity on observables, across subsamples that are likely to differ in their unobserved characteristics. For each of our three P-BOT samples, we divide the data according to eight criteria and, within each subsample, examine the relationship between various pre-promotion period productivity indicators and promotion propensity, as well as post-promotion productivity.

These cross-subsample comparisons are displayed in Tables D4 and D5, where each column represents a different subsample, and each row reports the regression coefficient of the dependent variable on the specified covariate. The estimates show general directional consistency across subsamples: for a given covariate, coefficients are all statistically indistinguishable from zero (insignificant); insignificant or significant and positive; or insignificant or significant and negative. There are no covariates for which some estimates are significant and positive while others are significant and negative. For instance, Panel A of Table D4 examines the relationship between promotion and billable hours, new matters, and promotion desire in our billable hours sample. Billable hours is positively and significantly correlated with promotion across demographic subsamples defined by gender, age, law school rank, and parental status, with the exception of insignificant coefficients (though high variances) for older partners and top law school attendees. Similarly, new matters is positively and significantly correlated with promotion among all subgroups, as is promotion desire.¹⁰ We interpret this stability as suggestive evidence in favor of monotonicity.

¹⁰While there is some variation in magnitude between subgroups, we find that our P-BOT estimates are robust to the use of separate selection equations by gender, which we describe in further detail in our results section.

4.2 Outcome equation

Having estimated promotion propensities, we order promoted associates (i.e., partners) from lowest to highest predicted probability of promotion. The outcome test is then carried out within subsets drawn from the lower end of this distribution, following the procedure described in further detail in Appendix E. We apply the same procedure to comparisons by gender, parental status, age, and attendance at a top-ranked law school.

Prior to running the formal P-BOT, we first examine graphically how simple group comparisons evolve as progressively more meritorious lawyers are admitted to the estimation sample. We first calculate the post-promotion average productivity indicators (billable hours, total hours, and number of new matters brought to the firm) for female and male partners in the lowest quintile of predicted promotion probability. We then expand the sample to include the first two quintiles, the first three, and so forth until all promoted lawyers are included. We present similar analyses split by parental status, age, and top law school attendance as well. These cumulative comparisons show whether the estimated outcome gap is concentrated among partners who appear closest to the promotion boundary or instead persists throughout the promoted population.

Our primary P-BOT analytical specification defines the marginal sample using the lower 30% of the propensity-score distribution among promoted associates. The relatively coarse cutoff reflects the limited number of partners available in our sample: narrower definitions would more closely target the theoretical margin but would sharply reduce precision.

We complement those estimates with the perturbation test developed by [Grau and Vergara \(2025\)](#) and described in Appendix E. This exercise aims to assess the scope for infra-marginality bias driven by local rank reversals. That is, the test looks at whether uncertainty in the estimation of propensity scores can materially alter the classification of marginal partners and, therefore, the outcome test results. To implement the test, for each promoted lawyer, we simulate 500 realizations from a standardized normal distribution to randomly ‘jitter’ R_i^* . We then recompute the ranking, redefine the estimation sample, and re-estimate the outcome test. Finally, we plot the distribution of the estimates from these simulations.

5 Results

5.1 Determinants of promotion propensities

Table 3 shows that billable hours and new matters brought to the firm have strong predictive power – in fact, the strongest – on the likelihood of being promoted. This is how associates contribute to the success of their firm. Therefore, it was expected that performance along those dimensions would attract the reward of promotion to partner. However, even when factored in, the adjusted $\bar{R}^2$ is low: in the Billable Hours sample, it is only 22.5%. For the Total Hours and New Matters samples, respectively, it is 11.3% and 22.8%. In both the Billable Hours and New Matters samples, we construct our selection equation using both pre-promotion decision billable hours and new matters as productivity measures; in the Total Hours sample, we use total hours and new matters.

Each sample constructs propensity scores including a combination of demographics and productivity measures. While in theory nothing restricts us from constructing propensity scores via a combination of billable hours, total hours, and new matters, restricting to non-missingness across all three significantly reduces our sample sizes.

Being female is significantly negatively correlated with promotion in the Billable Hours and New Matters samples, and being white or Asian is significantly positively correlated with promotion in the Total Hours sample.¹¹ We note a proclivity towards the promotion of mothers, which we did not expect; this could be consistent with firms’ accommodation of motherhood.¹² The finding could also reflect survivor selection — only highly productive mothers remaining in large firms long enough to be promotion candidates. We do not investigate the race and motherhood results in further detail due to the small sample sizes of mothers and non-white/Asian individuals among marginal partners.

We were also able to test the ‘rat race’ hypothesis presented by [Landers et al. \(1996\)](#) in

¹¹These results, however, are not stable over minimal changes in sample definition. Removing individuals with the lowest promotion desire would make the gender coefficients insignificant, as seen in Table A5.

¹²This result is not necessarily in conflict with the assertion that women are more likely than men to delay their first child until after the promotion decision is resolved ([Park and Rim, 2020](#)). Here, we control for gender and parenthood.

the context of promotion in law firms. If ‘working hard’ is conducive to promotion independently of objective productivity, we should see a statistically significant coefficient associated with total hours after controlling for billable hours. We do not see evidence of this; in unreported regressions, including both billable and total hours in the selection equation yields a significant coefficient on billable hours, yet an insignificant coefficient on total hours.

5.2 Outcome tests: gender

Figure 1 illustrates our main result with post-promotion period billable hours as the outcome. We highlight three implications from the figure.

First, in the first quintile of the distribution of promoted individuals, that is, by construction, the least promising among partners, men bill more hours than women on average, although the difference is not statistically significant at the 5% level using standard methods of inference.¹³ We therefore cannot detect evidence of discrimination against women on the basis of that test.¹⁴

Second, for males, average productivity is very much constant, whereas for females, the average productivity appears to increase as we consider partners with a higher probability of promotion. Given the standard errors, however, the observation is more suggestive than conclusive. We confirm this in the non-cumulative Figure A5, where we see no clear-cut evidence of monotonic increases or decreases in average productivity across quintiles.

The slightly narrowing gender gap as progressively higher-propensity partners enter the sample (Figure 1) indicates that composition away from the promotion boundary is unlikely to account for a large share of the near-margin estimate. Moreover, the relatively stable productivity across quintiles and persistence of a gender gap (Figure A5) reduces concerns about accurately identifying the marginal sample: inadvertently including inframarginal partners

¹³Figure A8a suggests that the pattern that we describe is present among the most marginal partners and seemingly more pronounced, thus reinforcing the presumption that the lack of significance observed in the figure results from small sample sizes.

¹⁴We are also able to analyze the productivity of the most able associates who did not get promoted. Here, bearing in mind that our results must be qualified by an even smaller sample size, productivity differences among the most successful associates who were not promoted also do not suggest taste-based discrimination (see Figure A1).

is less likely to bias the results of the outcome test if inframarginal partners are observably similar post-promotion to marginal partners.

Moving from a graphical, illustrative approach to our analytical framework, we evaluate the performance of female partners relative to their male peers by comparing the means of those in the bottom 30%. This percentage should be lower in principle, but the small sample size constrains us to select a coarser than ideal partition.¹⁵ Following [Grau and Vergara \(2025\)](#), we implement the P-BOT using two methods: a simple approach and a non-parametric approach. The simple approach selects a sample of promoted lawyers close to the margin based on the quantiles of the predicted propensity score of promotion and performs a linear regression on gender post-promotion productivity. On the other hand, the non-parametric approach estimates the point difference by subtracting the female and non-female conditional expectations for lawyer’s productivity, calculated non-parametrically at the 30th percentile of the estimated promotion probability distribution. Regardless of the approach, a finding that female partners are more productive (i.e., work more hours or bring in more matters) at the margin would suggest discrimination under the outcome test framework, as it would be consistent with a higher standard being applied to them at the promotion decision stage.

Table 4 Panel A presents the results of the implementation of the P-BOT in our main Billable Hours sample. In the first column of Panel A, we implement the simple approach, where the point estimate is obtained from a linear regression of billable hours on a gender indicator in a sample of promoted partners with low predicted propensity scores.¹⁶ In the second column of Panel A, we implement the non-parametric version. In each case, 95% confidence intervals are derived from a percentile bootstrap with 500 repetitions.¹⁷ Notably, this inference method differs from Figure 1, which relies on standard, non-bootstrapped

¹⁵While we continue to see some consistent results as we narrow the marginal cutoff to 20% and 10%, sample sizes reach single digits, leading to highly variable results across specifications and inflated standard errors; we are hesitant to draw conclusions about robustness given these small samples.

¹⁶Note that the samples chosen here for the marginal test do not perfectly align with the bottom quintile shown graphically in Figure 1. Moreover, standard error computations differ: the figure provides simple 95% confidence intervals around each group’s mean, while in the P-BOT, we use a bootstrap.

¹⁷We follow [Grau and Vergara \(2025\)](#)’s methods for inference; as they note, the bootstrap may not be valid for propensity score-based procedures, but since our regressor that generates the distribution is derived from a continuous process, it is valid in the case of the P-BOT.

confidence intervals around each group’s mean rather than a single bootstrapped interval around the difference.

Point estimates are both negative and statistically significant at the 5% level, implying a lower effective promotion threshold for women relative to men under the P-BOT assumptions. Based on the simple estimate, marginally promoted female partners are billing about 250 fewer hours annually relative to marginal male partners (and about 500 fewer from the non-parametric estimate¹⁸).

Alternative specifications support our main result. Marginal male partners bill more hours when using an OLS model to estimate promotion probability (Table B2 Panel A), when using OLS with Lasso to select predictors (Table B3 Panel A), and when using a heteroskedastic probit model (Table B4 Panel A). Table B7 Panel A, which is the average outcome test and counterpart of the fifth quintile comparison in Figure 1, yields the same conclusion with slightly more precision: men bill more hours than women on average.¹⁹ Thus, the P-BOT fails to suggest any evidence of discrimination against women with respect to promotion to partner, and if anything, points to the opposite.

An additional concern is that a single promotion equation may overlook heterogeneity in the impacts of observables on promotion propensity across groups. Thus, as a robustness check, we re-estimate the selection equation separately by gender and rank promoted male and female partners by their group-specific promotion propensities. The resulting estimates, reported in Table B1, are similar to those from our baseline specification.

We also consider extended samples to test the sensitivity of these results to sample construction. First, we consider a sample where we look at those promoted either by Wave II or Wave III. That is, we expand the sample to lawyers who were associates in Wave I for our ranking procedure. However, we acknowledge that being promoted in Wave II is not a

¹⁸Aside from the fairly wide confidence intervals around each estimate, the discrepancy between the simple and non-parametric estimates largely stems from the triangular kernel weighting in the latter; when using a uniform kernel, we see more similar results.

¹⁹Again, the distinction between the regression results and the graph comes from the construction of the confidence intervals: in our regressions, those are bootstrapped, whereas in the graph, we construct standard 95% confidence intervals around each mean. Notably, checking for significance via overlap in graphical confidence intervals can be a more conservative test than necessary, as the individual means’ intervals can overlap yet the difference in means may still be significant (Greenland et al., 2016).

‘clean’ outcome, as when we rank likelihood of promotion, we include associates who may not yet have been considered or felt ready for promotion. In addition, as Wave I data lacks information on promotion desire, we exclude this information in Wave II for the sake of consistency. We also compute propensity scores using total hours across all samples, due to lack of recorded billable hours in Wave I (which changes the scoring procedure for the determination of promotion propensities). Hence, the results from our extended sample are not directly comparable to those stemming from the main sample. Nevertheless, we submit that this extended sample contains some additional information that is worth examining despite its inherent drawbacks. In our extended marginal sample of lawyers who were promoted in waves II or III (Table B8 Panel A), the parametric test yields a significant difference between males and females at the 5% level, while the non-parametric one does not.

Table B9 Panel A represents the largest sample, with no cutoff to isolate marginal partners, and combining partners promoted in waves II and III. In that sample, we find again that women bill fewer hours at the 5% (as well as 1%) level. As Ayres (2002) notes, inframarginality bias operates in a predictable direction: inframarginal individuals — those productive enough that the firm promotes them regardless of any threshold — dilute productivity gaps when included in the sample by raising the average. We therefore expect the full-sample estimate to be biased toward zero relative to the true marginal gap. A significant gender gap in the non-marginal sample is accordingly a conservative lower bound: detecting it despite this dilution strengthens the marginal result.

With respect to total hours in the past week, at the margin both tests fail to reveal a statistically significant difference between genders, but the sign of the difference suggests that marginal males work more hours than marginal females (Table 4 Panel A). The difference is also statistically insignificant using the simple P-BOT approach and marginally significant using the non-parametric approach in the marginal extended sample combining those promoted to partner in waves II and III (Table B8 Panel A). However, this difference is significant at the 5% level in the entire sample of partners promoted in Wave III (Table B7 Panel A), in line with the fifth quintile graphical comparison in Figure A3.

In our primary specification, women seem to bring more new matters to the firm, but the differences are again insignificant, and by a larger margin than for total hours (Table 4 Panel A). No statistically significant differences emerge in any of our samples with respect to new matters brought to the firm.

One possible explanation is that women may face a worse post-promotion environment (client allocation, mentorship, continued discrimination as partners) that suppresses their hours, e.g., they may be missing out on desirable assignments that would have allowed them to clock the necessary billable hours to equal men’s output. Or, they may be missing out on the opportunity to work on more intensive cases, even at the partner stage. We cannot directly test such propositions; however, we run a number of suggestive tests.

First, we run the P-BOT on the sample of individuals who declare not feeling discriminated against across any dimension (in Wave II or Wave III) and find similar, if not larger, point estimates (see Tables B5 and B6). To the extent that women accurately gauge and report post-promotion discrimination, if the post-promotion suppression narrative were true, we would expect the reverse.

Moreover, we do not see evidence that reporting discrimination is correlated with promotion propensity (Figure A6). That is, marginal women are no more likely than nonmarginal women to report discrimination, suggesting that, contingent on reports of discrimination reflecting reality, any post-promotion discrimination would not select on marginal candidates specifically. If discrimination were operating primarily through post-promotion mechanisms — client allocation, mentorship access, partner-track obstruction — one would therefore expect it to appear at least as strongly in the non-marginal sample, where the number of women is substantially larger and the estimates more precise. However, the full-sample tests show a post-promotion billable hours gap between women and men that is significant and negative yet meaningfully smaller in magnitude than the marginal gap. This is more consistent with a lower promotion threshold allowing through women who are genuinely less productive at the margin, rather than a uniform post-promotion ceiling depressing all women equally, as the post-promotion suppression narrative would suggest.

While one may argue that women further from the margin are simply productive enough to overcome a post-promotion penalty, examining changes in men’s and women’s billable hours from before to after promotion provides little support for this explanation. As shown in Figure A7, within each propensity quintile, the pre-to-post change in billable hours for women is not significantly different from the corresponding change for men. If women faced a productivity penalty that emerged or intensified after promotion, we would expect their billable hours to decline more, or increase less, following promotion relative to men. However, we do not see evidence of this across any of the quintiles.

We also find no evidence of a gender difference in post-promotion tasks. Evidence from the P-BOT on total hours worked in the past week highlights a similar directional gap, suggesting that the billable hours finding cannot be explained solely by women working comparable total hours but receiving fewer opportunities to translate those hours into billable work. We find no significant difference in men’s and women’s involvement in pro-bono work (Table B11). We do find that women tend to work in smaller firms than men, as well as specialize in financial law less frequently (as shown in Figure A9), which could drive differences in post-promotion expectations; still, χ^2 tests on these differences are statistically insignificant, and productivity metrics such as billable hours do not appear significantly correlated with characteristics such as firm size.

The data provide some evidence consistent with the accommodation interpretation. Among lawyers who become parents between Waves II and III, women experience a more negative change in hours than men who also become parents, relative to the corresponding difference among those who do not become parents. That is, women’s labor supply appears to be more strongly affected by parenthood than men’s, adding support to the idea that firms may accommodate an expected reduction in women’s labor supply associated with motherhood. Notably, the P-BOT comparing marginal non-mothers to all marginal men, despite a small sample size, does not return significantly different results than our main specification. This suggests that if results are driven by an accommodation for current or future motherhood, firms are adjusting to accommodate all women rather than solely existing mothers.

5.3 Outcome tests: additional characteristics

Figure A2 provides a similar analysis to Figure 1 but across different characteristics, comparing the performances of partners born before and after 1972 (median age), parents and non-parents, and graduates from top 20 and non-top 20 law schools.²⁰ Focusing on the marginal individuals in the first quintile, we do not see clear graphical evidence of discrimination across any dimension.

The analysis with respect to parenthood, in Table 4 Panel B, supports the graphical evidence. Whether with billable hours, total hours, or new matters, there is no statistical difference in performance between parents and non-parents at the margin.²¹ Table B7 Panel B, Table B8 Panel B, and Table B9 Panel B show a similar lack of significance at the 5% level for the samples without percentile restriction, for the samples including those promoted in Wave II, and for the samples without percentile restriction including those promoted in waves II and III, respectively.

The P-BOT applied to age (Table 4 Panel C) also generally fails to detect any difference in performance between those born before and after the median birth year at the margin. However, in the New Matters sample, we see that at the 1% significance level, marginally promoted partners born before 1972 bring in about 12 fewer new matters (simple method) or 6 fewer new matters (non-parametric method) than marginally promoted partners born after 1972. One possible explanation for this is that older lawyers may be expected to focus on different areas; for instance, there may be less emphasis on client acquisition and more on client retention. Notably, this result becomes insignificant in our three additional sample specifications (Table B7 Panel C, Table B8 Panel C, and Table B9 Panel C), but we do gain significance on our billable hours result, with point estimates reflecting older marginal partners billing between 96 and 186 fewer annual hours post-promotion than younger marginal partners. Thus, on the whole, the evidence therefore points towards no discrimination against

²⁰We graphically present the same comparisons in terms of total hours and new matters in Figure A3 and Figure A4, respectively.

²¹To reiterate, we could not run the analysis by gender and parenthood (in particular, to determine if there is a difference in performance between mothers and non-mothers) since the sample sizes would become too small.

older associates, and if anything, a bias towards them.

The P-BOT applied to law school rank also fails to detect a significant difference in productivity between marginal partners who attended vs. did not attend a top 20 law school (Table 4 Panel D), a result that holds both among the samples with no percentile cutoff (Table B7 Panel D) and among the samples with no percentile cutoff that combines Waves II and III (Table B9 Panel D). However, in the sample that maintains the percentile cutoff but combines Waves II and III (Table B8 Panel D), we see significance across the board: post-promotion, top law school marginal partners bill more hours (significant at the 5% level) and have more total hours (significant at the 10% level) than non-top law school marginal partners, but they also bring in fewer new matters (significant at the 1% level). These results are primarily driven by anomalies of the data: The majority of the marginal sample comes from Wave I, where a handful of the non-top law school marginal partners are unusually productive along the new matters dimension, while many top law school marginal partners do not bring in any new matters.

5.4 Outcome tests: Perturbation tests

Misclassifications can induce inframarginality bias if the outcome distribution differs between wrongly promoted and wrongly denied partners. To address this concern, we implement the perturbation test following the procedure outlined in Appendix E.

Figure 2 shows the results for the Billable Hours sample, while results for the Total Hours and New Matters samples can be found in Figure D1. These tests suggest limited scope for inframarginality bias in the Billable Hours and Total Hours samples, as the distribution of the perturbed outcome test includes very few zeros. Only in the worst case scenario is the conclusion of prejudice reversed. For the New Matters sample, the perturbation test reaffirms our null results, as the distribution is centered around zero.

5.5 Attrition

We also look into whether different categories of the population are more likely to drop from the law firm environment, as well as out of the sample entirely. This is not necessarily evidence of discrimination but can usefully complement the narrative, especially if there are differences in how productivity correlates with likelihood of departure (whether from private law or from our sample) across groups. About a third of the initial respondents have left the sample by Wave III (see Table C3); among respondents who stay in the sample, about half have left private law by Wave III (see Table C1). Among those who leave private law, the most common exits are other industry positions and solo practices, followed by government positions (see Figures C1 and C2).

Park and Rim (2020) find that attrition is not endogenous relative to fertility timing. We similarly test whether attrition is endogenous relative to desire for promotion, perception of discrimination, female status (and motherhood), age, or law school prestige. Table C2 suggests that older associates and those who attended top 20 law schools are more likely to drop out of private law, while parents, those promoted, and those with a higher desire to be promoted are less likely. We find no significant differences in private law attrition by gender. Moreover, when it comes to leaving the sample entirely, table C4 suggests that attrition is not linked with desire for promotion, perception of discrimination, female status (and motherhood) or law school prestige, though there is evidence of age-related attrition. These findings help rule out a scenario whereby discriminated and therefore discouraged categories of associates would drop from private law — or from the survey — before the promotion decision.

5.6 Promotion and perceptions

Next, we utilize the various subjective variables in the dataset to assess whether lawyers' perceptions of discrimination align with the conclusions of an 'objective' analysis and to evaluate the impact of promotion on satisfaction with workplace diversity.

Note that the survey allows us to differentiate between various forms of bias, including

verbal, assignment-based, client-based, and colleague-based.²² Extending [Lehmann](#)’s work, in particular, we examine whether perceptions of discrimination have any predictive power on promotion (holding all fixed and variable objective productivity indicators constant) and whether pre-promotion perceptions of discrimination are correlated with post-promotion productivity measures among partners. Given the panel nature of the dataset, we estimate an individual fixed effects model of probability of promotion on satisfaction over diversity of workplace and perceived discrimination of various forms.

Table [C5](#) shows that pre-promotion perception of discrimination has no bearing on our two main measures of productivity, billable hours and new matters, among partners. Results on total hours are mixed, with some forms of discrimination correlated with more hours and others with fewer; however, it is worth noting that total hours is inherently a less stable measure of overall productivity due to its shorter time frame. Further, Table [C6](#) shows that, after controlling for gender, the enumerated forms of discrimination are not significantly correlated with promotion itself.

We also checked how perception of discrimination correlates with desire to be promoted in Table [C7](#).²³ We find that there is no significant correlation between desire for promotion and an aggregate indicator of perception of any form of discrimination.

Finally, Table [5](#) also investigates whether promotion is correlated with satisfaction over diversity or different forms of perceived discrimination. Overall, the results suggest that promoted individuals between waves II and III are on the job more nor less likely to express satisfaction over diversity or (subjectively) experience discrimination; thus, we see a lack of evidence that there is survival bias at play. We do, however, note increased satisfaction over diversity among those promoted when combining all three waves (Table [C8](#)), suggesting that those promoted early may become more complacent with respect to diversity issues. Experiences of discrimination, on the other hand, are unaffected.

²²As shown in Figure [C3](#), across the three waves, nearly 40 percent of female lawyers in private law firms report experiencing some form of discrimination. This rate is more than three times higher than that of male lawyers. A similar pattern is observed in non-private law firms, although it is less pronounced.

²³Note that the sample is larger since it contains both promoted and non-promoted individuals.

6 Conclusion

We apply a prediction-based outcome test approach to detect the possible presence of bias against certain categories of lawyers in the promotion process within law firms using the three waves of the AJD dataset. At the margin — that is, for partners whom the model predicts to have been barely promoted — a higher performance by those with a given characteristic would suggest bias against those exhibiting that characteristic, the logic being that they would have had to overcome a higher threshold in order to become promoted.

To determine the marginal partners, the test ranks associates according to their predicted probabilities of promotion, based on relevant observable characteristics, including number of billable hours and number of new matters brought to the firm. We note that billable hours and new matters are the most important predictors of promotion. Unexpectedly, motherhood displays a positive correlation as well. Total hours play a weakly positive significant role.

Applying the marginal prediction-based outcome test, we find that men seem to be at a disadvantage with respect to billable hours. This pattern suggests different explanations, such as accommodation for lower productivity during early motherhood, or a ‘compensatory’ mechanism of trend reversal relative to a previously established practice of discrimination as found in [Card et al. \(2022\)](#). In the present context, it may also be construed as a compensatory mechanism for *current* wrongs, i.e., mistreatment at the associate level.

Our results are robust to a perturbation test that randomizes the error term used to calculate the promotion threshold. We also examine whether the gender difference in post-promotion billable hours could instead reflect differential treatment after partnership; however, the result persists across several diagnostic exercises.

When analyzing perceptions of discrimination, women are more likely to report discrimination in all legal settings, not just private law firms. We do find some evidence that promotion is correlated with increased satisfaction over diversity, but only among those who got promoted early. However, the law firm environment is indeed more conducive to perceptions of discrimination, particularly among those who do not end up being promoted, yet our findings caution against the notion that women face a stricter promotion threshold.

Our work has several limitations. One is the small number of individuals promoted to partner, which required a coarser than ideal partition to identify those at the margin. We focused on potential discrimination based on gender, parenthood, age, and law school rank, and had to forgo testing for racial discrimination due to the limited number of racial minorities in the marginal group. Similarly, there were not enough partners who were mothers at the margin.

Another important caveat is that partners may be evaluated based on other criteria, such as their ability to manage a team of associates and the (unobserved) productivity of such a team, rather than their own individual performance. However, since this quality is not easily observable by firms during the promotion process, it would not affect the ranking procedure used to determine who is marginal among those promoted. We also checked that female and male partners, both at the margin and in general, oversee a similar small number of associates on average, so it is unlikely that team productivity differences could stem from that dimension.

If a fourth wave of data is collected, we hope to gather questions specifically related to the senior partner experience. A fourth wave would also substantiate how large law firms have changed substantially in their diversity practices, remote work structures, and partnership tracks.

Finally, the outcome test performed in this analysis only considers one form of possible discrimination, namely taste-based discrimination at the promotion process stage. Even if this specific form is ruled out, it should not end the investigation. While we can use self-reported perceived discrimination to filter out other possible forms of discrimination at the associate stage, subtle barriers may still hamper the productivity of lawyers before they are considered for promotion. Discrimination could also be present in ways not measured by any outcome test, including marginal outcome tests. Those may still be underinclusive as tests of disparate treatment because they are not well structured to capture treatment motivated by rational statistical inference, that is, statistical discrimination, which would require different detection methods ([Bohren et al., 2022](#)).

References

- ADAMS, R. B. AND P. FUNK (2012): “Beyond the glass ceiling: Does gender matter?” *Management science*, 58, 219–235.
- ANTECOL, H., D. A. COBB-CLARK, AND E. HELLAND (2014): “Bias in the legal profession: Self-assessed versus statistical measures of discrimination,” *The Journal of Legal Studies*, 43, 323–357.
- ANWAR, S. AND H. FANG (2006): “An alternative test of racial prejudice in motor vehicle searches: Theory and evidence,” *American Economic Review*, 96, 127–151.
- ARNOLD, D., W. DOBBIE, AND P. HULL (2022): “Measuring Racial Discrimination in Bail Decisions,” *American Economic Review*, 112, 2992–3038.
- ARNOLD, D., W. DOBBIE, AND C. S. YANG (2018): “Racial bias in bail decisions,” *The Quarterly Journal of Economics*, 133, 1885–1932.
- AYRES, I. (2002): “Outcome tests of racial disparities in police practices,” *Justice Research and Policy*, 4, 131–142.
- AZMAT, G., V. CUÑAT, AND E. HENRY (2020): “Gender promotion gaps and Career aspirations,” *Management Science*, forthcoming, available at SSRN 3518420.
- AZMAT, G. AND R. FERRER (2017): “Gender gaps in performance: Evidence from young lawyers,” *Journal of Political Economy*, 125, 1306–1355.
- BABCOCK, L., M. P. RECALDE, L. VESTERLUND, AND L. WEINGART (2017): “Gender Differences in the Allocation of Low-Promotability Tasks: The Role of Backlash and Affective Responses,” *American Economic Review: Papers and Proceedings*, 107, 131–135.
- BECKER, G. (1957): *the Economics of Discrimination*, University of Chicago Press.
- BERTRAND, M. (2011): “New perspectives on gender,” in *Handbook of labor economics*, Elsevier, vol. 4, 1543–1590.

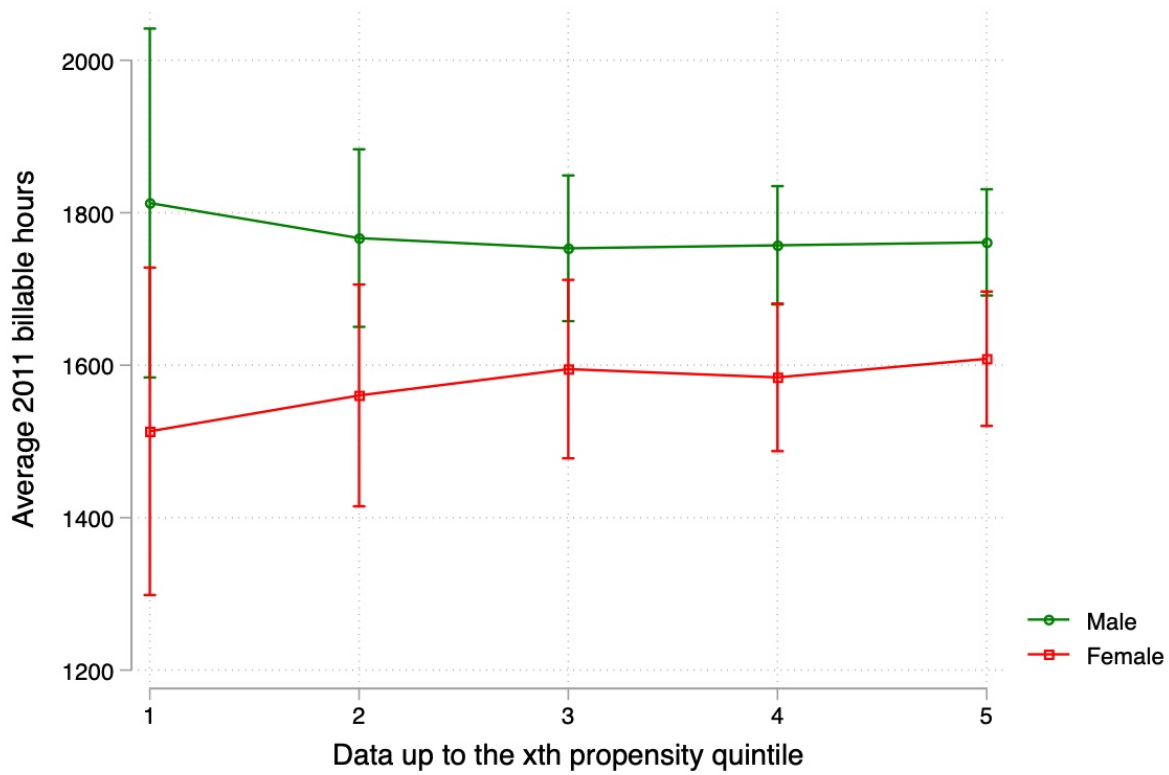
- BERTRAND, M. AND K. F. HALLOCK (2001): “The gender gap in top corporate jobs,” *ILR Review*, 55, 3–21.
- BLAU, F. D. AND L. M. KAHN (2017): “The gender wage gap: Extent, trends, and explanations,” *Journal of economic literature*, 55, 789–865.
- BOHREN, J. A., P. HULL, AND A. IMAS (2022): “Systemic discrimination: Theory and measurement,” *National Bureau of Economic Research*.
- CANAY, I. A., M. MOGSTAD, AND J. MOUNTJOY (2024): “On the Use of Outcome Tests for Detecting Bias in Decision Making,” *The Review of Economic Studies*, 91, 2135–2167.
- CARD, D., S. DELLA VIGNA, P. FUNK, AND N. IRIBERRI (2022): “Gender Differences in Peer Recognition by Economists,” *Econometrica*, 90, 1937–1971.
- CHANOW, L. AND J. S. STERLING (2021): “In Their Own Words: Experienced Women Lawyers Explain Why They Are Leaving Their Law Firms and the Profession,” *American Bar Association*, <https://www.americanbar.org/content/dam/aba/administrative/women/2021/intheirownwords-f-4-19-21-final.pdf>.
- COMMISSION, G. C. (1995): “Glass Ceiling Commission-Good for business: Making full use of the nation’s human capital,” <https://hdl.handle.net/1813/79348>.
- CROSON, R. AND U. GNEEZY (2009): “Gender Differences in Preferences,” *Journal of Economic Literature*, 47, 448–474.
- ENGLAND, P., A. LEVINE, AND E. MISHEL (2020): “Progress toward gender equality in the United States has slowed or stalled,” *Proceedings of the National Academy of Sciences*, 117, 6990–6997.
- GRAU, N. AND D. VERGARA (2022): “Combining discrimination diagnostics to identify sources of statistical discrimination,” *Economics Letters*, 215, 110294.

- (2025): “An observational implementation of the outcome test with an application to ethnic prejudice in pretrial detentions,” *Journal of the European Economic Association*, conditionally accepted.
- GREENLAND, S., S. SENN, K. ROTHMAN, J. CARLIN, C. POOLE, S. GOODMAN, AND A. DG (2016): “Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations,” *Eur J Epidemiol*, 31, 337–350.
- HAILE, S., T. EMMANUEL, AND A. DZATHOR (2016): “Barriers and Challenges Confronting Women for Leadership and Management Positions: Review and Analysis.” *International Journal of Business & Public Administration*, 13.
- HULL, P. (2021): “What marginal outcome tests can tell us about racially biased decision-making,” *National Bureau of Economic Research Working Paper*.
- JOHANNESSEN, N. AND S. MUCHARDT (2026): “Is the bar higher for female scholars? Evidence from career steps in economics,” *Journal of Public Economics*, 257, 105021.
- KAY, F. M. AND E. H. GORMAN (2012): “Developmental practices, organizational culture, and minority representation in organizational leadership: The case of partners in large US law firms,” *The Annals of the American Academy of Political and Social Science*, 639, 91–113.
- KNOWLES, J., N. PERSICO, AND P. TODD (2001): “Racial bias in motor vehicle searches: Theory and evidence,” *Journal of Political Economy*, 109, 203–229.
- LANDERS, R. M., J. B. REBITZER, AND L. J. TAYLOR (1996): “Rat Race Redux: Adverse Selection in the Determination of Work Hours in Law Firms,” *American Economic Review*, 86, 329–348.
- LEE, J. AND C. LIU (1999): “Measuring discrimination in the workplace: Strategies for lawyers and policymakers,” *U. Chi. L. Sch. Roundtable*, 6, 195.

- LEHMANN, J.-Y. (2011): “Job assignment and promotion under statistical discrimination: Evidence from the early careers of lawyers,” *mimeo*.
- MECHOULAN, S. AND N. SAHUGUET (2015): “Assessing racial disparities in parole release,” *The Journal of Legal Studies*, 44, 39–74.
- MOGSTAD, M., J. P. ROMANO, A. SHAIKH, AND D. WILHELM (2022): “Inference for Ranks with Applications to Mobility Across Neighborhoods and Academic Achievement Across Countries,” *The Review of Economic Studies*.
- NATIONAL ASSOCIATION FOR LAW PLACEMENT (2019): “Report on Diversity in US Law Firms,” https://www.nalp.org/uploads/2018NALPReportonDiversityinUSLawFirms_FINAL.pdf.
- PARK, K. H. AND N. RIM (2020): “The Disparate Impact of Up-or-Out Promotion Policy on Fertility Timing,” *American Law and Economics Review*, 22, 127–172.
- PAYNE-PIKUS, M. R., J. HAGAN, AND R. L. NELSON (2010): “Experiencing discrimination: Race and retention in America’s largest law firms,” *Law & Society Review*, 44, 553–584.
- PERSICO, N. (2009): “Racial Profiling? Detecting Bias Using Statistical Evidence,” *Annual Review of Economics*, 1, 229–254.
- STERLING, J. S. AND N. REICHMAN (2012): “Navigating the gap: Reflections on 20 years researching gender disparities in the legal profession,” *FIU L. Rev.*, 8, 515.
- VYTLACIL, E. (2002): “Independence, monotonicity, and latent index models: An equivalence result,” *Econometrica*, 70, 331–341.

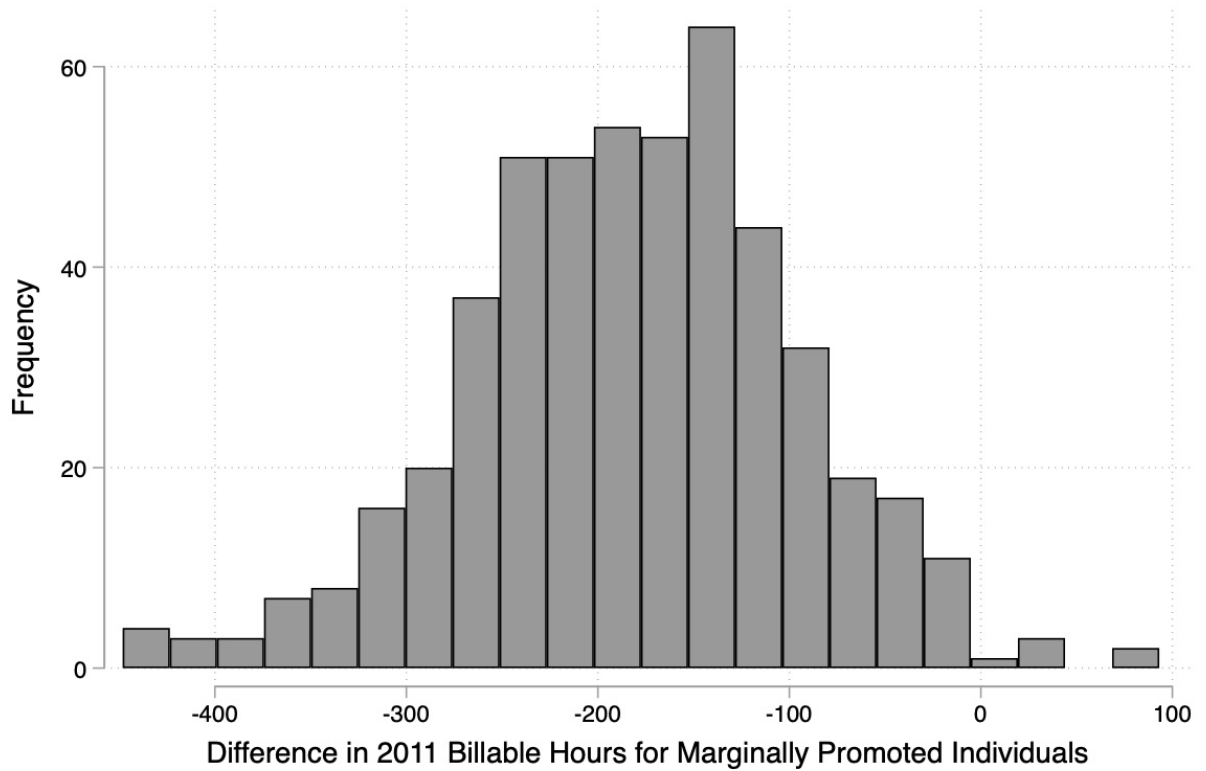
7 Tables and Figures

Figure 1: 2011 Billable Hours for Different Propensity Score Quintiles by Gender



Note: This figure plots average billable hours post-promotion (2011) among promoted men and women in different groups of predicted promotion probabilities (1: quintile 1; 2: quintiles 1-2; 3: quintiles 1-3; 4: quintiles 1-4; 5: full sample). Propensity scores estimated using a probit model. Standard 95% confidence intervals are computed around each group's mean; results are similar if bootstrapping.

Figure 2: Perturbation Test (Billable Hours)



Note: This figure displays the results of the perturbation test; that is, it shows the distribution of differences in means between the 2011 billable hours of marginally promoted women and marginally promoted men. To produce these results, we first estimate each individual's propensity score using a probit model. Then, for each promoted individual, we simulate 500 realizations from a standard normal distribution that are subsequently used to simulate propensity scores. We use these scores to redefine samples of marginal individuals and estimate the outcome test within each one.

Table 1: Descriptive Statistics (Billable Hours)

	Full Sample		Promoted Sample		Marginal Sample	
	Mean	SD	Mean	SD	Mean	SD
Promoted	0.628	0.484	1.000	0.000	1.000	0.000
2006 Billable Hours	1875.257	400.684	1939.941	381.049	1810.667	450.148
2011 Billable Hours	1648.491	442.256	1700.604	369.304	1670.196	423.807
2006 New Matters	3.513	13.047	4.935	16.193	0.686	1.225
Female	0.442	0.498	0.396	0.491	0.431	0.500
Born in 1972 or Earlier	0.428	0.496	0.396	0.491	0.510	0.505
Top 20 Law School	0.297	0.458	0.302	0.460	0.275	0.451
Parent	0.584	0.494	0.615	0.488	0.569	0.500
Mother	0.249	0.433	0.260	0.440	0.235	0.428
White or Asian	0.848	0.360	0.858	0.350	0.902	0.300
Promotion Desire	7.658	2.854	8.462	2.268	6.549	2.859
Observations	269		169		51	

Note: This table presents descriptive statistics for our primary yearly billable hours sample, as well as the promoted sample that results from selecting only promoted associates and the marginal sample that results from selecting observations who were promoted and had propensity scores in the bottom 30th percentile. Sample is restricted to Wave II associates.

Table 2: Descriptive Statistics (Billable Hours, Split by Gender)

	Full Sample		Promoted Sample		Marginal Sample	
	Mean	SD	Mean	SD	Mean	SD
Panel A: Male						
Promoted	0.680	0.468	1.000	0.000	1.000	0.000
2006 Billable Hours	1936.440	414.402	2000.333	364.014	1900.897	432.797
2011 Billable Hours	1739.320	439.016	1761.127	359.162	1779.448	399.640
2006 New Matters	3.227	8.690	4.127	10.225	0.655	1.317
Born in 1972 or Earlier	0.467	0.501	0.402	0.493	0.448	0.506
Top 20 Law School	0.293	0.457	0.284	0.453	0.207	0.412
Parent	0.600	0.492	0.588	0.495	0.586	0.501
White or Asian	0.847	0.362	0.853	0.356	0.897	0.310
Promotion Desire	8.307	2.388	8.559	2.306	6.310	2.929
Observations	150		102		29	
Panel B: Female						
Promoted	0.563	0.498	1.000	0.000	1.000	0.000
2006 Billable Hours	1798.134	370.178	1848.000	390.643	1691.727	454.584
2011 Billable Hours	1534.000	420.889	1608.463	368.075	1526.182	420.105
2006 New Matters	3.874	17.059	6.164	22.471	0.727	1.120
Born in 1972 or Earlier	0.378	0.487	0.388	0.491	0.591	0.503
Top 20 Law School	0.303	0.461	0.328	0.473	0.364	0.492
Parent	0.563	0.498	0.657	0.478	0.545	0.510
White or Asian	0.849	0.360	0.866	0.344	0.909	0.294
Promotion Desire	6.840	3.176	8.313	2.217	6.864	2.800
Observations	119		67		22	

Note: This table presents descriptive statistics split by gender for our primary yearly billable hours sample, as well as the promoted samples that result from selecting only promoted associates and the marginal samples that result from selecting observations who were promoted and had propensity scores in the bottom 30th percentile. Sample is restricted to Wave II associates.

Table 3: Determinants of Promotion Propensities Using a Probit Model

	Billable Hours Sample	Total Hours Sample	New Matters Sample
Billable Hours	0.106*** (0.028)		0.120*** (0.028)
Total Hours		0.061*** (0.023)	
New Matters	0.345*** (0.101)	0.041* (0.024)	0.336*** (0.086)
Promotion Desire	0.125*** (0.025)	0.123*** (0.022)	0.120*** (0.025)
Female	-0.141* (0.080)	-0.088 (0.066)	-0.168** (0.077)
Pre-72	-0.082 (0.053)	-0.047 (0.044)	-0.089* (0.052)
White or Asian	0.011 (0.074)	0.136** (0.056)	-0.017 (0.074)
Top 20 Law School	0.043 (0.057)	0.020 (0.050)	0.054 (0.057)
Parent	-0.029 (0.071)	-0.038 (0.060)	-0.054 (0.069)
Mother	0.235** (0.105)	0.143 (0.087)	0.288*** (0.102)
Middle Atlantic	0.034 (0.165)	0.085 (0.104)	0.005 (0.151)
East North Central	0.066 (0.166)	0.162 (0.101)	0.147 (0.153)
West North Central	0.245 (0.156)	0.252** (0.102)	0.308** (0.146)
South Atlantic	0.104 (0.153)	0.204** (0.096)	0.118 (0.143)
East South Central	0.151 (0.198)	0.161 (0.133)	0.177 (0.199)
West South Central	0.137 (0.166)	0.241** (0.115)	0.213 (0.155)
Mountain	0.057 (0.188)	0.232* (0.140)	0.128 (0.180)
Pacific	0.113 (0.154)	0.166* (0.096)	0.131 (0.143)
Observations	269	470	288
Adjusted / Pseudo R^2	0.225	0.113	0.228
Correctly classified (0.5 prob as threshold)	0.751	0.683	0.743
Correctly classified (prediction: promoted)	0.768	0.634	0.755
Correctly classified (prediction: nonpromoted)	0.709	0.712	0.720

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: This table presents the influences of various characteristics on likelihood of promotion within our billable hours, total hours, and new matters samples using a probit model to model promotion propensity. Columns report marginal effects, all continuous covariates (billable hours, total hours, new matters, promotion desire) are standardized, and data is restricted to those potentially promoted in Wave III.

Table 4: Main P-BOT Results Using Probit to Estimate Promotion Probability, Marginal 30% of Promoted Partners

	Billable Hours		Total Hours		New Matters	
	Simple Version	Non-Parametric	Simple Version	Non-Parametric	Simple Version	Non-Parametric
Panel A: Gender						
Point Estimate (F-M)	-253.266**	-493.342***	-4.839	-6.298	4.097	-0.428
Female Average	1526.182	1420.161	46.441	44.471	12.545	5.639
Male Average	1779.448	1913.503	51.280	50.769	8.448	6.067
95% C.I.	[-484.294, -2.765]	[-750.307, -202.177]	[-10.029, 1.913]	[-15.311, 3.008]	[-7.288, 14.792]	[-8.141, 5.475]
No. of Females	22	22	34	34	22	22
No. of Males	29	29	25	25	29	29
Panel B: Parental Status						
Point Estimate (P-NP)	-136.108	-75.989	-3.385	-0.340	5.126	3.657
Parent Average	1611.483	1623.784	47.000	46.942	12.226	7.374
Non-Parent Average	1747.591	1699.773	50.385	47.281	7.100	3.717
95% C.I.	[-373.933, 77.913]	[-470.612, 273.385]	[-9.566, 2.107]	[-9.207, 9.599]	[-7.082, 15.162]	[-1.920, 9.882]
No. of Parents	29	29	33	33	31	31
No. of Non-Parents	22	22	26	26	20	20
Panel C: Age						
Point Estimate (Pre-Post)	-135.197	-101.111	-1.207	-2.540	-12.602***	-6.039*
Pre-72 Average	1603.923	1610.517	47.857	45.722	4.038	3.445
Post-72 Average	1739.120	1711.628	49.065	48.262	16.640	9.484
95% C.I.	[-374.905, 98.690]	[-450.593, 255.590]	[-7.948, 3.323]	[-12.616, 6.430]	[-24.206, -2.695]	[-16.177, 0.366]
No. of Pre-72s	26	26	28	28	26	26
No. of Post-72s	25	25	31	31	25	25
Panel D: Law School Rank						
Point Estimate (T-NT)	-12.380	54.516	2.921	6.070	1.678	-0.156
Top 20 Average	1661.214	1691.137	50.769	52.243	11.400	5.739
Non-Top 20 Average	1673.595	1636.621	47.848	46.173	9.722	5.894
95% C.I.	[-314.307, 271.588]	[-398.417, 487.663]	[-3.923, 9.696]	[-4.590, 15.910]	[-11.085, 16.000]	[-6.194, 8.221]
No. of Top 20s	14	14	13	13	15	15
No. of Non-Top 20s	37	37	46	46	36	36

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: This table presents the P-BOT results by gender (male vs. female), parental status (parents vs. non-parents), age (born pre- vs. post-1972) and school rank (attended vs. did not attend a top-20 law school) from both the simple (difference in means) and non-parametric (local polynomial smoothing) approaches. The three panels display results for the yearly billable hours, total weekly hours, and new matters specifications. Promotion probabilities (propensity scores) are estimated using a probit model, and the samples include those who are marginally promoted (promoted and in the bottom 30th percentile of propensity score). Standard errors are bootstrapped with 500 repetitions. Data is restricted to those who were promoted between waves II and III.

Table 5: Promotions on Perceptions

	Δ Diversity Satisfaction	Δ Any Discrimination	Δ Verbal Discrimination	Δ Assignment Discrimination	Δ Client Discrimination	Δ Colleague Discrimination
Promoted	-0.041 (0.203)	0.042 (0.051)	-0.012 (0.038)	0.061 (0.038)	0.054 (0.040)	-0.006 (0.029)
Constant	4.279*** (0.062)	0.218*** (0.016)	0.101*** (0.012)	0.073*** (0.012)	0.069*** (0.013)	0.054*** (0.009)
Observations	481	536	535	533	534	533
Adjusted R^2	0.423	0.286	0.228	0.195	0.147	0.161

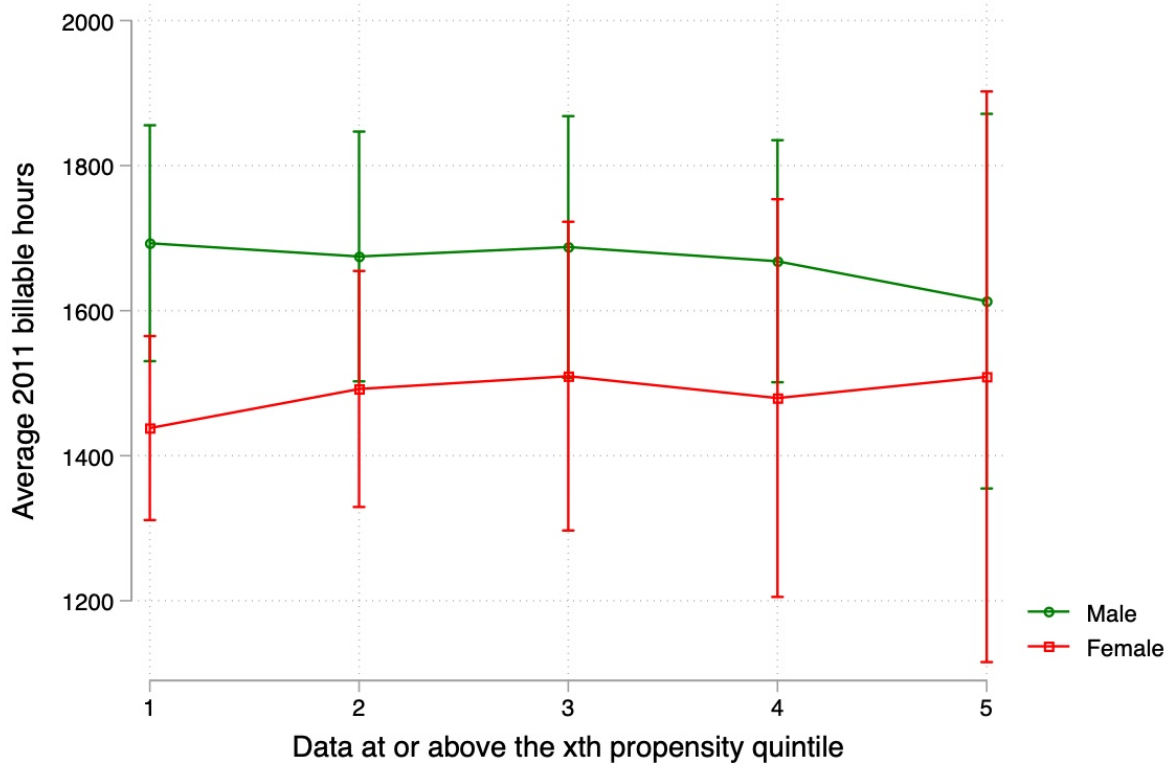
Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: This table presents the results of regressions of various measures of discrimination on promotion. Regressions include individual fixed effects, and standard errors are clustered on the individual level. Estimations use the full sample with no percentile cutoff; however, there is a restriction to those who are private law firm associates in Wave I, as well as those with a 5+ promotion desire. Data from waves II and III.

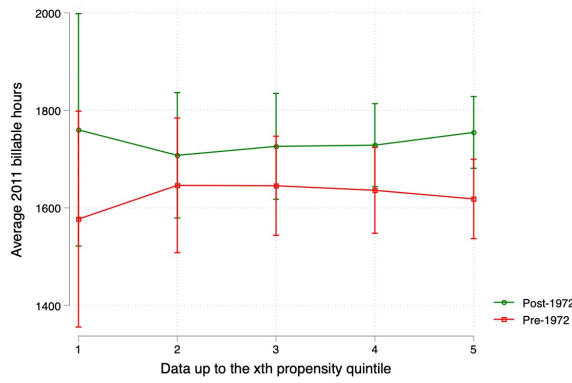
A Appendix: P-BOT Descriptives

Figure A1: Wave III Billable Hours for Different Propensity Score Quintiles by Gender (Nonpromoted)

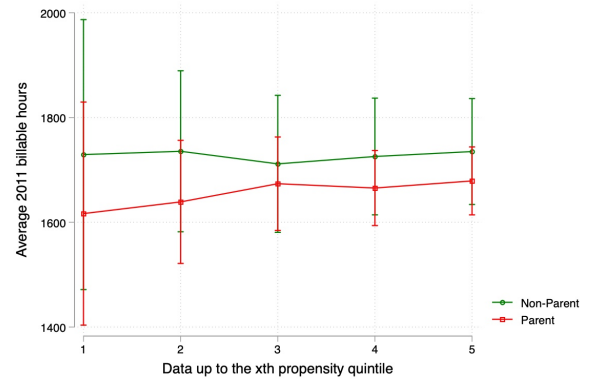


Note: This figure plots average billable hours post-promotion period (2011) among *non-promoted* men and women in different groups of predicted promotion probabilities (1: full sample; 2: quintiles 2-5; 3: quintiles 3-5; 4: quintiles 5-5; 5: quintile 5). Propensity scores estimated using a probit model. 95% confidence intervals are computed around each group's mean.

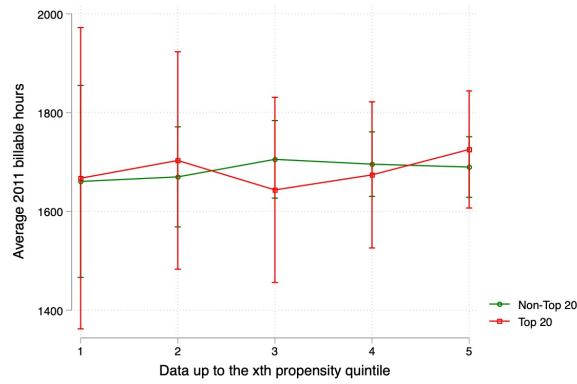
Figure A2: Wave III Billable Hours for Different Propensity Score Quintiles by Lawyer Characteristics



(a) Age



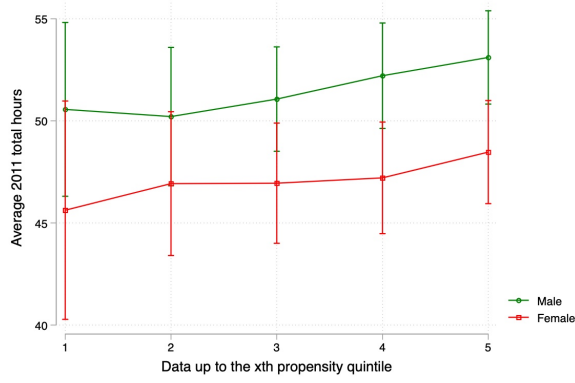
(b) Parental Status



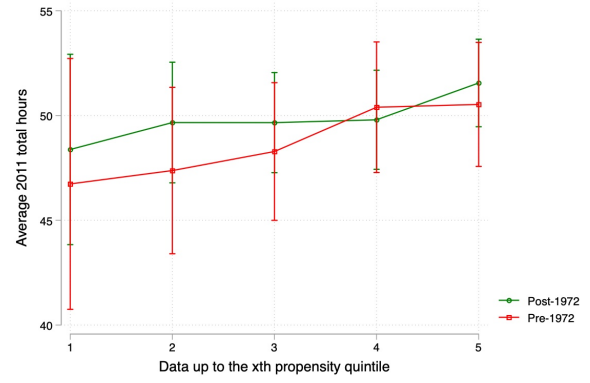
(c) Law School Rank

Note: This figure plots average billable hours post-promotion (Wave III), splitting by three demographics, in different groups of predicted promotion probabilities (1: quintile 1; 2: quintiles 1-2; 3: quintiles 1-3; 4: quintiles 1-4; 5: full sample). Propensity scores estimated using a probit model. 95% confidence intervals are computed around each group's mean.

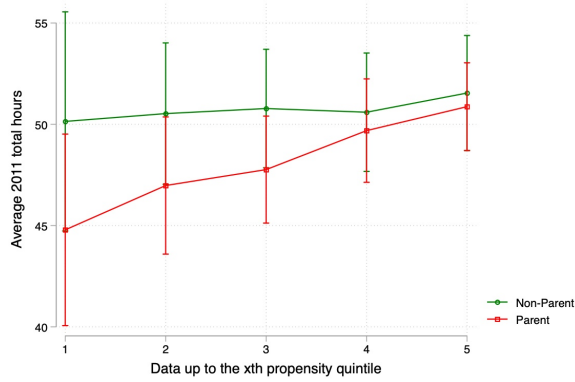
Figure A3: Wave III Total Hours for Different Propensity Score Quintiles by Lawyer Characteristics



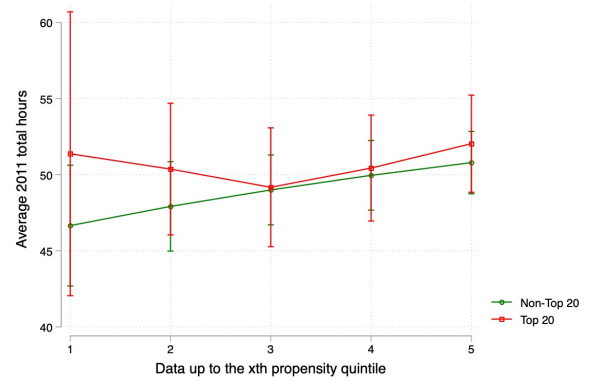
(a) Gender



(b) Age



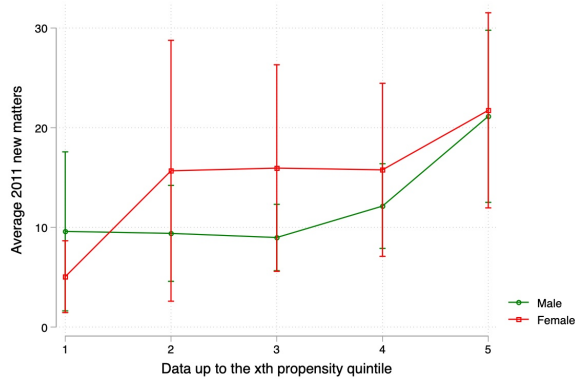
(c) Parental Status



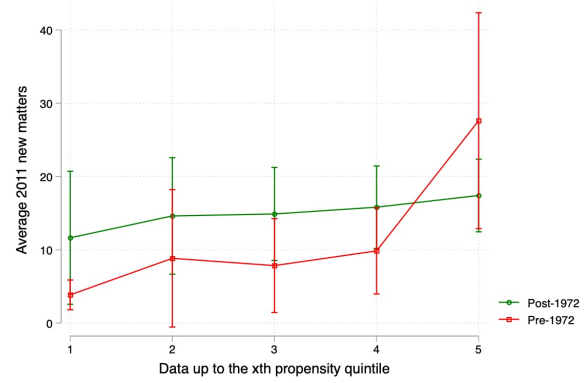
(d) Law School Rank

Note: This figure plots average billable hours post-promotion (Wave III), splitting by three demographics, in different groups of predicted promotion probabilities (1: quintile 1; 2: quintiles 1-2; 3: quintiles 1-3; 4: quintiles 1-4; 5: full sample). Propensity scores estimated using a probit model. 95% confidence intervals are computed around each group's mean.

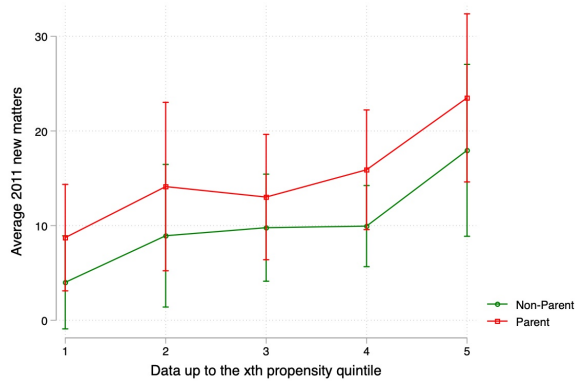
Figure A4: Wave III New Matters for Different Propensity Score Quintiles by Lawyer Characteristics



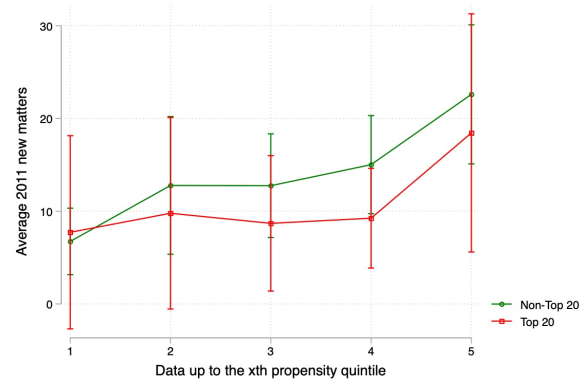
(a) Gender



(b) Age



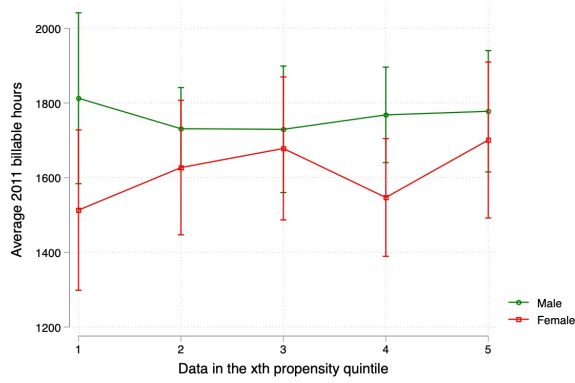
(c) Parental Status



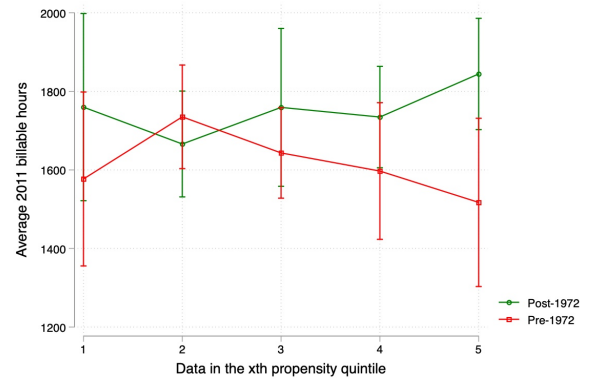
(d) Law School Rank

Note: This figure plots average billable hours post-promotion (Wave III), splitting by three demographics, in different groups of predicted promotion probabilities (1: quintile 1; 2: quintiles 1-2; 3: quintiles 1-3; 4: quintiles 1-4; 5: full sample). Propensity scores estimated using a probit model. 95% confidence intervals are computed around each group's mean.

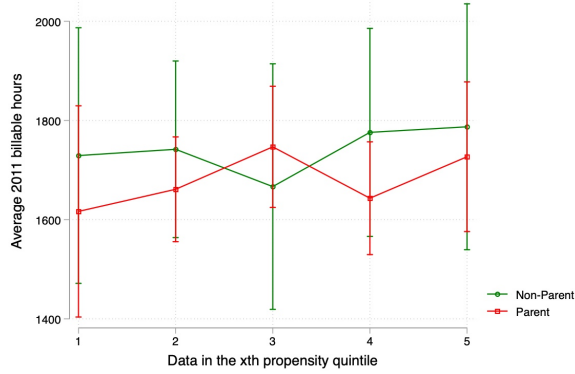
Figure A5: Wave III Billable Hours Within Different Propensity Score Quintiles by Lawyer Characteristics



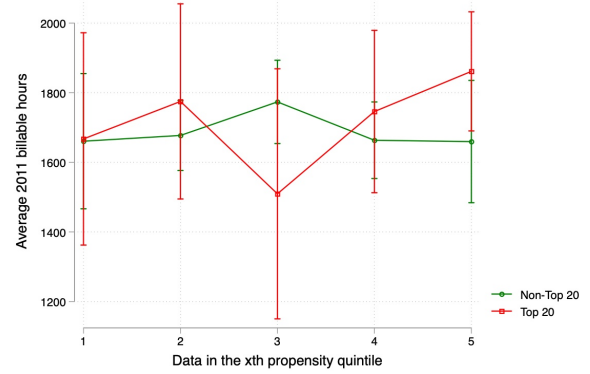
(a) Gender



(b) Age



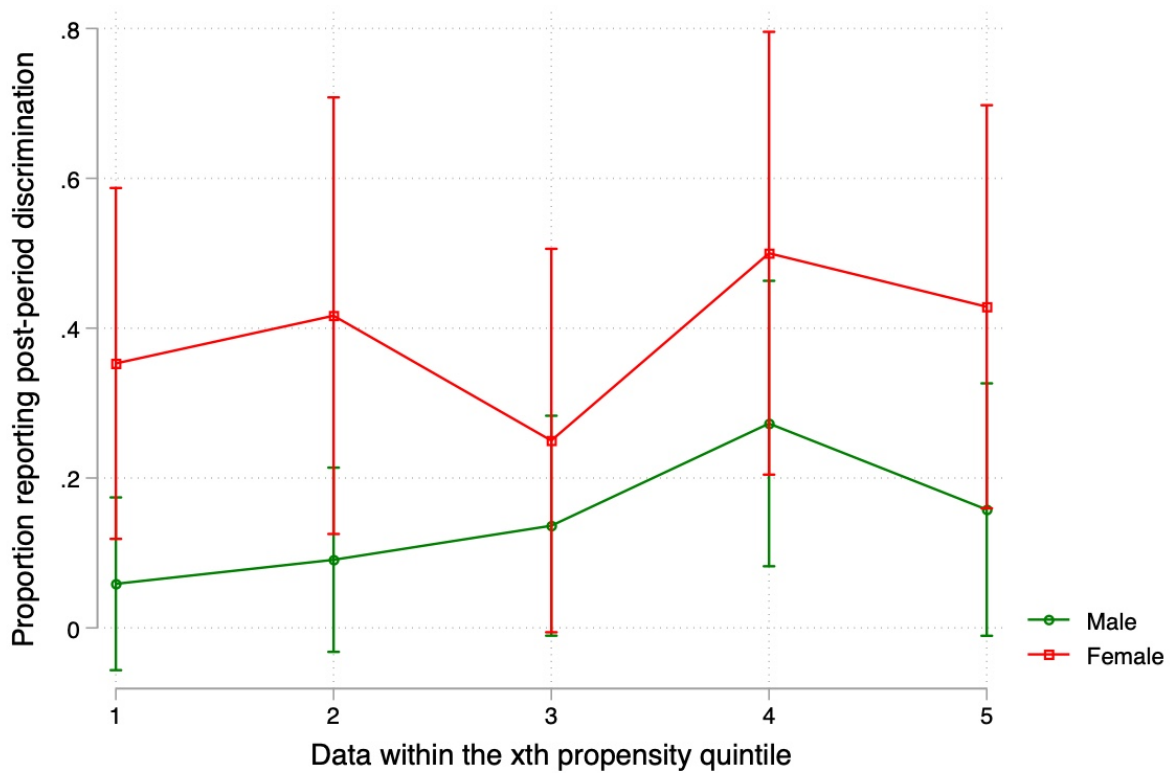
(c) Parental Status



(d) Law School Rank

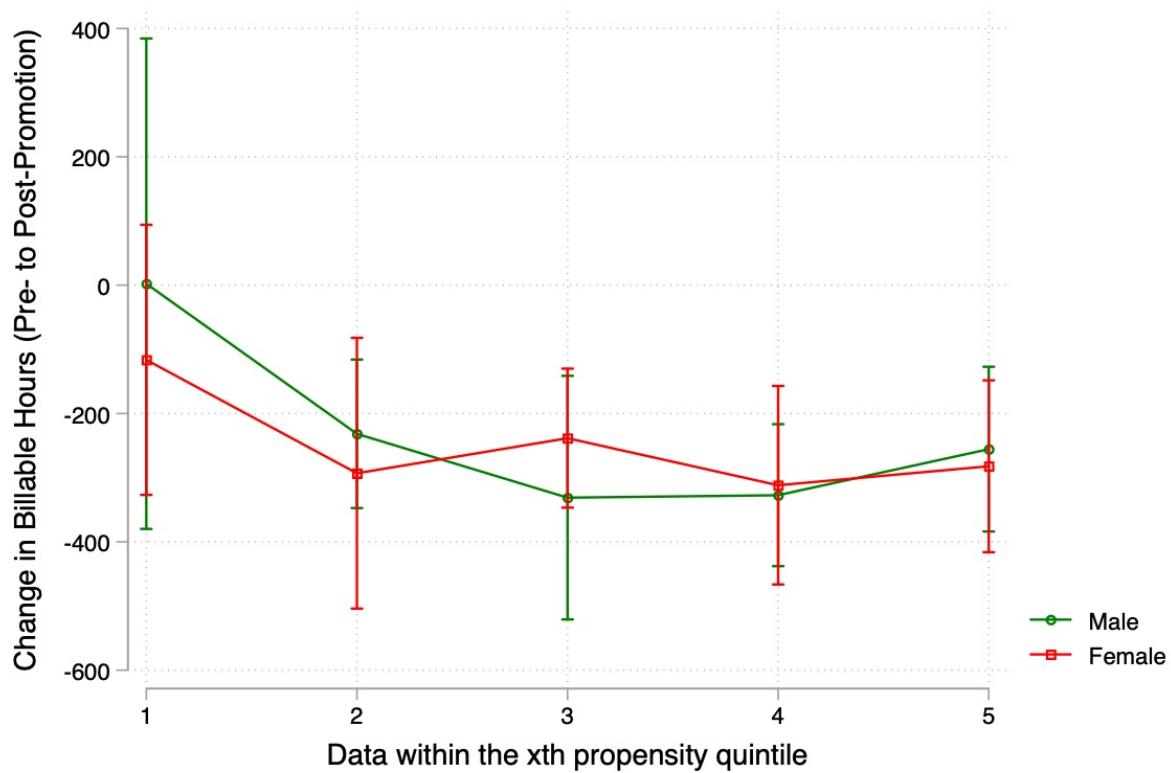
Note: This figure plots average billable hours post-promotion (Wave III), splitting by demographics, within different quintiles of predicted promotion probabilities. Propensity scores estimated using a probit model. 95% confidence intervals are computed around each group's mean.

Figure A6: Reports of Discrimination Within Different Propensity Score Quintiles by Lawyer Characteristics



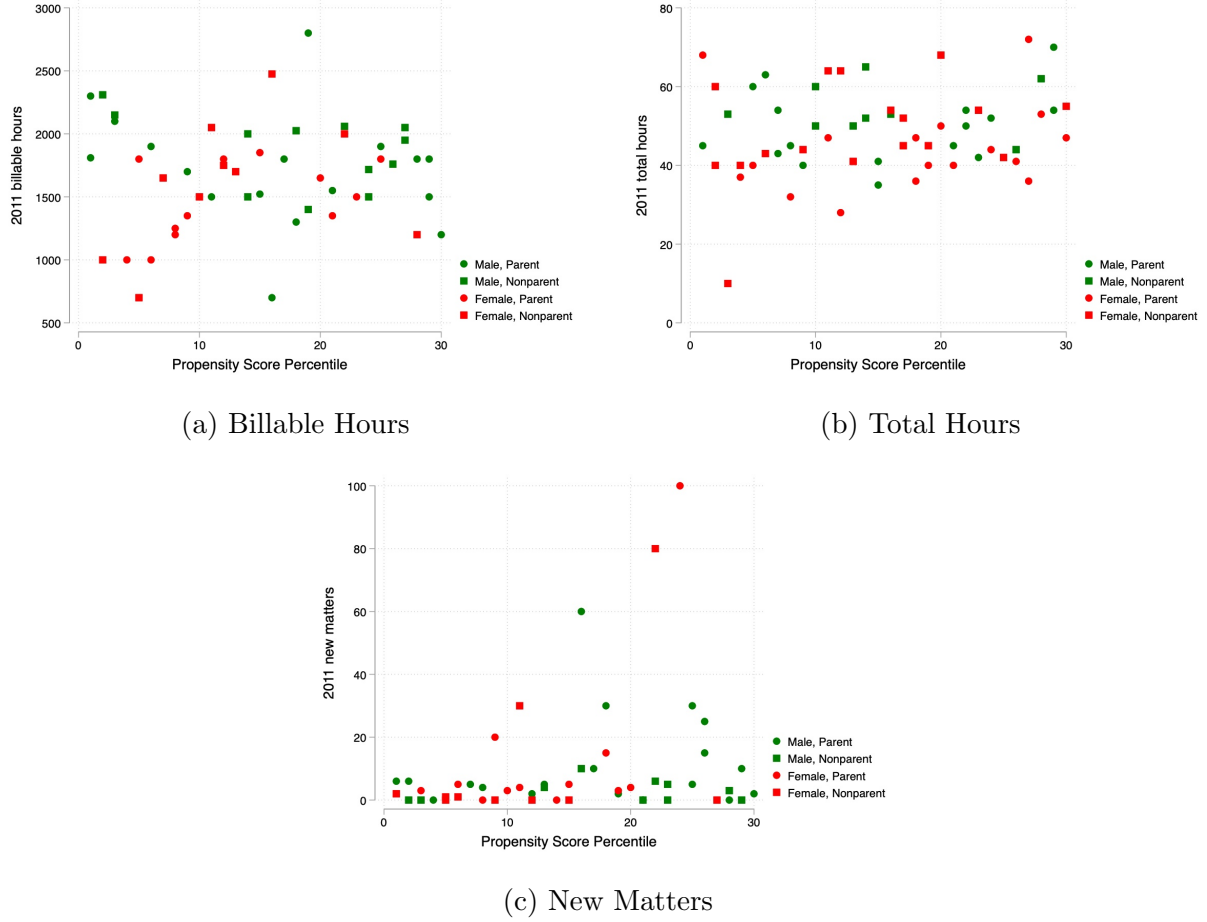
Note: This figure plots average post-promotion reports of discrimination among promoted men and women within the specified quintile of predicted promotion probability. Propensity scores estimated using a probit model. 95% confidence intervals are computed around each group's mean.

Figure A7: Change in Billable Hours (Post- vs. Pre-Promotion) Within Different Propensity Score Quintiles by Lawyer Characteristics



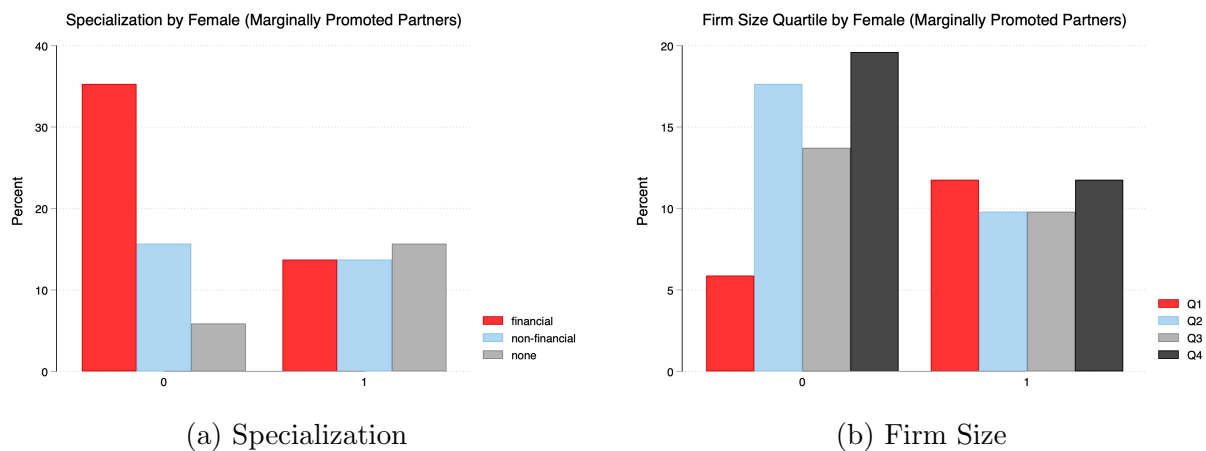
Note: This figure plots the average change in billable hours from pre- to post-promotion among promoted men and women within the specified quintile of predicted promotion probability. Propensity scores estimated using a probit model. 95% confidence intervals are computed around each group's mean.

Figure A8: 2011 Productivity Against Propensity Score by Gender and Parental Status



Note: This figure plots billable hours, total hours, and new matters post-promotion (Wave III) against propensity score percentiles (estimated using a probit model) for those marginally promoted, splitting by gender and parental status. Data is restricted to those who were promoted between waves II and III. Note that, in the new matters scatterplot, the three apparent outliers are not true outliers when considering the full distribution.

Figure A9: Firm Characteristics of Marginal Partners by Gender



Note: This figure plots legal specialization and firm size for marginally promoted partners, splitting by gender. Data is restricted to those who were promoted between waves II and III.

Table A1: Descriptive Statistics (Total Hours)

	<i>Full Sample</i>		<i>Promoted Sample</i>		<i>Marginal Sample</i>	
	Mean	SD	Mean	SD	Mean	SD
Promoted	0.417	0.494	1.000	0.000	1.000	0.000
2006 Total Hours	51.656	11.119	53.865	11.242	48.195	10.218
2011 Total Hours	47.281	12.912	51.143	12.303	48.492	11.215
2006 New Matters	3.911	12.858	5.082	15.614	5.220	9.410
Female	0.474	0.500	0.423	0.495	0.576	0.498
Born in 1972 or Earlier	0.434	0.496	0.403	0.492	0.475	0.504
Top 20 Law School	0.274	0.447	0.281	0.450	0.220	0.418
Parent	0.568	0.496	0.597	0.492	0.559	0.501
Mother	0.257	0.438	0.260	0.440	0.288	0.457
White or Asian	0.804	0.397	0.862	0.346	0.763	0.429
Promotion Desire	7.106	3.136	8.219	2.525	5.797	3.067
Observations	470		196		59	

Note: This table presents descriptive statistics for our total hours sample, as well as the promoted sample that results from selecting only promoted associates and the marginal sample that results from selecting observations who were promoted and had propensity scores in the bottom 30th percentile. Sample is restricted to Wave II associates.

Table A2: Descriptive Statistics (New Matters)

	<i>Full Sample</i>		<i>Promoted Sample</i>		<i>Marginal Sample</i>	
	Mean	SD	Mean	SD	Mean	SD
Promoted	0.590	0.493	1.000	0.000	1.000	0.000
2006 New Matters	3.479	12.739	5.029	16.258	0.765	1.408
2011 New Matters	14.795	34.655	21.376	43.162	10.216	19.763
2006 Billable Hours	1851.014	432.752	1924.571	407.870	1803.412	453.905
Female	0.434	0.496	0.388	0.489	0.431	0.500
Born in 1972 or Earlier	0.427	0.496	0.388	0.489	0.510	0.505
Top 20 Law School	0.292	0.455	0.294	0.457	0.294	0.460
Parent	0.587	0.493	0.618	0.487	0.608	0.493
Mother	0.243	0.430	0.259	0.439	0.235	0.428
White or Asian	0.854	0.354	0.859	0.349	0.902	0.300
Promotion Desire	7.542	2.968	8.388	2.359	6.569	2.934
Observations	288		170		51	

Note: This table presents descriptive statistics for our new matters sample, as well as the promoted sample that results from selecting only promoted associates and the marginal sample that results from selecting observations who were promoted and had propensity scores in the bottom 30th percentile. Sample is restricted to Wave II associates.

Table A3: Determinants of Promotion Propensities Using an OLS Model

	Billable Hours Sample	Total Hours Sample	New Matters Sample
Billable Hours	0.095*** (0.029)		0.096*** (0.028)
Total Hours		0.059** (0.023)	
New Matters	0.067** (0.028)	0.038* (0.022)	0.072*** (0.027)
Promotion Desire	0.142*** (0.030)	0.121*** (0.024)	0.128*** (0.029)
Female	-0.158* (0.087)	-0.088 (0.067)	-0.195** (0.085)
Pre-72	-0.084 (0.057)	-0.046 (0.045)	-0.098* (0.056)
White or Asian	0.014 (0.078)	0.129** (0.057)	-0.014 (0.078)
Top 20 Law School	0.021 (0.060)	0.026 (0.050)	0.026 (0.060)
Parent	-0.023 (0.075)	-0.043 (0.062)	-0.057 (0.074)
Mother	0.223* (0.114)	0.144 (0.089)	0.278** (0.111)
Middle Atlantic	0.003 (0.168)	0.100 (0.119)	-0.007 (0.154)
East North Central	0.068 (0.168)	0.172 (0.116)	0.164 (0.156)
West North Central	0.229 (0.161)	0.267** (0.116)	0.286* (0.150)
South Atlantic	0.078 (0.157)	0.212* (0.111)	0.088 (0.146)
East South Central	0.121 (0.206)	0.174 (0.146)	0.154 (0.208)
West South Central	0.095 (0.172)	0.251** (0.127)	0.169 (0.161)
Mountain	0.046 (0.195)	0.240 (0.153)	0.126 (0.189)
Pacific	0.052 (0.157)	0.177 (0.110)	0.067 (0.145)
Observations	269	470	288
Adjusted / Pseudo R^2	0.176	0.108	0.183
Correctly classified (0.5 prob as threshold)	0.732	0.685	0.726
Correctly classified (prediction: promoted)	0.746	0.635	0.733
Correctly classified (prediction: nonpromoted)	0.694	0.716	0.710

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: This table presents the influences of various characteristics on likelihood of promotion within our billable hours, total hours, and new matters samples using an OLS model to model promotion propensity. Columns report marginal effects, all continuous covariates (billable hours, total hours, new matters, promotion desire) are standardized, and data is restricted to those potentially promoted in Wave III.

Table A4: Determinants of Promotion Propensities Using a Heteroskedastic Probit Model

	Billable Hours Sample	Total Hours Sample	New Matters Sample
Billable Hours	0.220*** (0.048)		0.169*** (0.049)
Total Hours		0.057** (0.023)	
New Matters	0.564*** (0.185)	0.231** (0.110)	0.350*** (0.119)
Promotion Desire	0.103** (0.043)	0.166*** (0.038)	0.139*** (0.042)
Female	-0.097 (0.072)	-0.082 (0.065)	-0.122 (0.076)
Pre-72	-0.087* (0.053)	-0.055 (0.043)	-0.082 (0.052)
White or Asian	0.018 (0.063)	0.126** (0.058)	-0.014 (0.072)
Top 20 Law School	-0.056 (0.058)	0.041 (0.047)	0.016 (0.064)
Parent	-0.034 (0.059)	-0.039 (0.056)	-0.051 (0.062)
Mother	0.213** (0.098)	0.158* (0.084)	0.271*** (0.102)
Middle Atlantic	-0.063 (0.218)	0.122 (0.097)	0.044 (0.154)
East North Central	-0.124 (0.279)	0.208** (0.095)	0.180 (0.163)
West North Central	0.159 (0.241)	0.286*** (0.092)	0.374** (0.147)
South Atlantic	-0.034 (0.261)	0.241*** (0.087)	0.187 (0.149)
East South Central	0.037 (0.232)	0.258** (0.128)	0.207 (0.185)
West South Central	-0.050 (0.250)	0.301*** (0.105)	0.216 (0.159)
Mountain	-0.032 (0.267)	0.249* (0.128)	0.210 (0.173)
Pacific	0.014 (0.260)	0.235*** (0.090)	0.217 (0.155)
Observations	269	470	288
Adjusted / Pseudo R^2	0.028	0.020	0.014
Correctly classified (0.5 prob as threshold)	0.747	0.679	0.753
Correctly classified (prediction: promoted)	0.792	0.624	0.786
Correctly classified (prediction: nonpromoted)	0.667	0.713	0.704

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: This table presents the influences of various characteristics on likelihood of promotion within our billable hours, total hours, and new matters samples using a heteroskedastic probit model to model promotion propensity. To allow the model to converge, only continuous covariates (billable hours, total hours, new matters, promotion desire) are measured with heteroskedastic variances. Columns report marginal effects, all continuous covariates are standardized, and data is restricted to those potentially promoted in Wave III.

Table A5: Determinants of Promotion Propensities Using a Probit Model, No Low Desire

	Billable Hours Sample	Total Hours Sample	New Matters Sample
Billable Hours	0.094*** (0.032)		0.120*** (0.032)
Total Hours		0.069** (0.028)	
New Matters	0.380*** (0.113)	0.027 (0.030)	0.321*** (0.102)
Promotion Desire	0.227*** (0.052)	0.185*** (0.056)	0.232*** (0.055)
Female	-0.095 (0.088)	-0.081 (0.083)	-0.111 (0.089)
Pre-72	-0.069 (0.057)	-0.061 (0.053)	-0.085 (0.057)
White or Asian	-0.024 (0.087)	0.156** (0.072)	-0.053 (0.091)
Top 20 Law School	0.056 (0.064)	0.043 (0.062)	0.089 (0.064)
Parent	-0.053 (0.073)	-0.052 (0.072)	-0.070 (0.074)
Mother	0.291** (0.115)	0.216** (0.106)	0.347*** (0.116)
Middle Atlantic	0.182 (0.186)	0.138 (0.126)	0.086 (0.168)
East North Central	0.230 (0.183)	0.236** (0.120)	0.247 (0.170)
West North Central	0.376** (0.170)	0.292** (0.117)	0.393** (0.157)
South Atlantic	0.272 (0.170)	0.252** (0.113)	0.241 (0.159)
East South Central	0.223 (0.220)	0.292* (0.163)	0.172 (0.223)
West South Central	0.278 (0.182)	0.341** (0.138)	0.294* (0.172)
Mountain	0.227 (0.198)	0.279* (0.160)	0.254 (0.186)
Pacific	0.363** (0.172)	0.274** (0.117)	0.320** (0.160)
Observations	219	344	229
Adjusted / Pseudo R^2	0.201	0.078	0.206
Correctly classified (0.5 prob as threshold)	0.749	0.645	0.738
Correctly classified (prediction: promoted)	0.765	0.638	0.750
Correctly classified (prediction: nonpromoted)	0.667	0.653	0.694

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: This table presents the influences of various characteristics on likelihood of promotion within our billable hours, total hours, and new matters samples using a probit model to model promotion propensity. Individuals with a promotion desire of less than 6 out of 10 are excluded. Columns report marginal effects, all continuous covariates are standardized, and data is restricted to those potentially promoted in Wave III.

Table A6: Determinants of Promotion Propensities Using a Probit Model, No Productivity Indicators

	Billable Hours Sample	Total Hours Sample	New Matters Sample
Promotion Desire	0.147*** (0.025)	0.138*** (0.021)	0.141*** (0.025)
Female	-0.145* (0.084)	-0.106 (0.066)	-0.188** (0.082)
Pre-72	-0.070 (0.056)	-0.051 (0.044)	-0.078 (0.055)
White or Asian	-0.012 (0.078)	0.126** (0.057)	-0.042 (0.079)
Top 20 Law School	0.041 (0.061)	0.024 (0.050)	0.039 (0.060)
Parent	-0.006 (0.075)	-0.047 (0.061)	-0.051 (0.073)
Mother	0.171 (0.109)	0.134 (0.087)	0.243** (0.107)
Middle Atlantic	0.092 (0.164)	0.062 (0.107)	0.054 (0.153)
East North Central	0.158 (0.163)	0.163 (0.105)	0.226 (0.153)
West North Central	0.306** (0.155)	0.240** (0.105)	0.336** (0.146)
South Atlantic	0.152 (0.153)	0.178* (0.099)	0.148 (0.144)
East South Central	0.199 (0.205)	0.160 (0.137)	0.216 (0.209)
West South Central	0.140 (0.171)	0.209* (0.118)	0.197 (0.160)
Mountain	0.112 (0.194)	0.217 (0.143)	0.176 (0.187)
Pacific	0.139 (0.154)	0.163 (0.099)	0.141 (0.144)
Observations	269	470	288
Adjusted / Pseudo R^2	0.143	0.098	0.143

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: This table presents the influences of various non-productivity-related characteristics on likelihood of promotion within our billable hours, total hours, and new matters samples using a probit model to model promotion propensity. Columns report marginal effects, all continuous covariates are standardized, and data is restricted to those potentially promoted in Wave III.

B Appendix: Additional P-BOT Specifications

Table B1: P-BOT Results Using Probit to Estimate Promotion Probability, Marginal 30% of Promoted Partners, Separate Selection Equations

	Billable Hours		Total Hours		New Matters	
	<i>Simple Version</i>	<i>Non-Parametric</i>	<i>Simple Version</i>	<i>Non-Parametric</i>	<i>Simple Version</i>	<i>Non-Parametric</i>
Panel A: Gender						
Point Estimate (F-M)	-248.377*	-657.199	-5.311*	-7.203	2.975	5.903
Female Average	1531.300	1463.360	44.160	45.561	9.850	9.135
Male Average	1779.677	2120.559	49.471	52.764	6.875	3.232
95% C.I.	[-519.258, 28.212]	[-1270.877, 120.935]	[-11.354, 0.625]	[-21.306, 7.422]	[-7.899, 16.694]	[-5.088, 26.765]
No. of Females	20	20	25	25	20	20
No. of Males	31	31	34	34	32	32
Panel B: Parental Status						
Point Estimate (P-NP)	-107.994	13.897	-2.458	8.533	8.419**	0.293
Parent Average	1622.906	1714.044	46.167	50.709	12.219	1.365
Non-Parent Average	1730.900	1700.147	48.625	42.176	3.800	1.071
95% C.I.	[-325.631, 132.799]	[-734.003, 791.030]	[-7.892, 3.626]	[-13.576, 27.249]	[1.876, 17.424]	[-5.566, 5.139]
No. of Parents	32	32	36	36	32	32
No. of Non-Parents	20	20	24	24	20	20
Panel C: Age						
Point Estimate (Pre-Post)	-70.396	321.977	-1.069	7.754	-8.025*	0.849
Pre-72 Average	1663.571	1947.750	45.958	55.317	3.350	3.185
Post-72 Average	1733.968	1625.774	47.028	47.563	11.375	2.336
95% C.I.	[-291.390, 166.029]	[-232.346, 778.343]	[-8.446, 9.428]	[-8.018, 20.704]	[-16.767, 0.339]	[-9.191, 5.901]
No. of Pre-72s	21	21	24	24	20	20
No. of Post-72s	31	31	36	36	32	32
Panel D: Law School Rank						
Point Estimate (T-NT)	91.411	382.793	2.566	11.372	-1.144	35.574
Top 20 Average	1732.133	1749.645	50.706	57.766	9.133	39.874
Non-Top 20 Average	1640.722	1366.852	48.140	46.394	10.278	4.300
95% C.I.	[-227.165, 327.543]	[-119.553, 1088.355]	[-4.302, 9.294]	[-5.358, 24.883]	[-13.644, 10.076]	[-8.069, 53.977]
No. of Top 20s	15	15	17	17	15	15
No. of Non-Top 20s	36	36	43	43	36	36

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: This table presents the P-BOT results by gender (male vs. female), parental status (parents vs. non-parents), age (born pre- vs. post-1972) and school rank (attended vs. did not attend a top-20 law school) from both the simple (difference in means) and non-parametric (local polynomial smoothing) approaches. The three panels display results for the yearly billable hours, total weekly hours, and new matters specifications. Promotion probabilities (propensity scores) are estimated using probit models, with separate selection equations for each group (i.e., men's and women's propensity scores are separately estimated), and the samples include those who are marginally promoted (promoted and in the bottom 30th percentile of propensity score). Standard errors are bootstrapped with 500 repetitions. Data is restricted to those who were promoted between waves II and III.

Table B2: P-BOT Results Using OLS to Estimate Promotion Probability, Marginal 30% of Promoted Partners

	Billable Hours		Total Hours		New Matters	
	<i>Simple Version</i>	<i>Non-Parametric</i>	<i>Simple Version</i>	<i>Non-Parametric</i>	<i>Simple Version</i>	<i>Non-Parametric</i>
Panel A: Gender						
Point Estimate (F-M)	-242.148**	-541.265***	-5.153	-6.407	6.192	-3.093
Female Average	1504.679	1380.592	46.424	44.329	20.192	12.633
Male Average	1746.826	1921.857	51.577	50.736	14.000	15.726
95% C.I.	[-494.116, -0.621]	[-786.956, -231.322]	[-9.960, 1.783]	[-17.468, 2.352]	[-10.255, 24.348]	[-23.303, 13.356]
No. of Females	28	28	33	33	26	26
No. of Males	23	23	26	26	25	25
Panel B: Parental Status						
Point Estimate (P-NP)	-23.549	1.473	-3.021	1.216	7.148	-2.585
Parent Average	1603.724	1616.329	47.364	47.495	20.100	12.760
Non-Parent Average	1627.273	1614.856	50.385	46.278	12.952	15.345
95% C.I.	[-277.236, 215.129]	[-419.284, 346.598]	[-9.259, 2.505]	[-8.940, 11.900]	[-12.755, 21.889]	[-22.916, 14.668]
No. of Parents	29	29	33	33	30	30
No. of Non-Parents	22	22	26	26	21	21
Panel C: Age						
Point Estimate (Pre-Post)	-127.182	55.551	-0.892	-2.902	-14.195	-20.811**
Pre-72 Average	1551.538	1642.008	48.241	45.304	9.920	4.290
Post-72 Average	1678.720	1586.457	49.133	48.206	24.115	25.101
95% C.I.	[-339.593, 134.900]	[-319.873, 387.886]	[-8.041, 4.016]	[-14.062, 6.935]	[-28.918, 7.213]	[-41.559, -4.362]
No. of Pre-72s	26	26	29	29	25	25
No. of Post-72s	25	25	30	30	26	26
Panel D: Law School Rank						
Point Estimate (T-NT)	37.784	108.440	2.998	6.474	-8.156	-9.453
Top 20 Average	1639.812	1683.989	51.083	52.537	11.400	7.685
Non-Top 20 Average	1602.029	1575.549	48.085	46.064	19.556	17.138
95% C.I.	[-229.049, 334.808]	[-345.019, 452.386]	[-4.676, 10.186]	[-5.485, 16.379]	[-24.500, 7.017]	[-24.940, 6.027]
No. of Top 20s	16	16	12	12	15	15
No. of Non-Top 20s	35	35	47	47	36	36

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: This table presents the P-BOT results by gender (male vs. female), parental status (parents vs. non-parents), age (born pre- vs. post-1972) and school rank (attended vs. did not attend a top-20 law school) from both the simple (difference in means) and non-parametric (local polynomial smoothing) approaches. The three panels display results for the yearly billable hours, total weekly hours, and new matters specifications. Promotion probabilities (propensity scores) are estimated using OLS, and the samples include those who are marginally promoted (promoted and in the bottom 30th percentile of propensity score). Standard errors are bootstrapped with 500 repetitions. Data is restricted to those who were promoted between waves II and III.

Table B3: P-BOT Results Using OLS to Estimate Promotion Probability and Lasso to Select Predictors, Marginal 30% of Promoted Partners

	Billable Hours		Total Hours		New Matters	
	Simple Version	Non-Parametric	Simple Version	Non-Parametric	Simple Version	Non-Parametric
Panel A: Gender						
Point Estimate (F-M)	-228.983**	-432.705***	-6.746**	-8.142*	-5.303	-10.170
Female Average	1528.500	1446.859	45.769	44.089	26.318	23.367
Male Average	1757.483	1879.564	52.515	52.232	31.621	33.537
95% C.I.	[-453.147, -48.837]	[-704.969, -168.125]	[-12.688, -0.757]	[-15.164, 0.281]	[-38.960, 29.657]	[-44.074, 25.094]
No. of Females	22	22	26	26	22	22
No. of Males	29	29	33	33	29	29
Panel B: Parental Status						
Point Estimate (P-NP)	-117.056	-47.591	-2.182	-2.809	19.029	7.346
Parent Average	1624.278	1606.605	48.618	46.884	35.676	30.593
Non-Parent Average	1741.333	1654.196	50.800	49.693	16.647	23.247
95% C.I.	[-366.226, 141.545]	[-490.248, 407.746]	[-6.983, 3.956]	[-10.266, 5.123]	[-8.290, 48.325]	[-20.287, 41.355]
No. of Parents	36	36	34	34	34	34
No. of Non-Parents	15	15	25	25	17	17
Panel C: Age						
Point Estimate (Pre-Post)	22.365	161.762	-1.809	-2.435	27.066	14.488
Pre-72 Average	1672.300	1730.388	48.316	46.419	46.316	37.515
Post-72 Average	1649.935	1568.627	50.125	48.854	19.250	23.027
95% C.I.	[-213.962, 218.176]	[-172.776, 455.980]	[-10.112, 4.624]	[-9.669, 7.686]	[-12.589, 76.824]	[-22.332, 66.599]
No. of Pre-72s	20	20	19	19	19	19
No. of Post-72s	31	31	40	40	32	32
Panel D: Law School Rank						
Point Estimate (T-NT)	173.836	120.709	2.376	2.025	4.032	2.418
Top 20 Average	1788.231	1708.533	51.556	49.665	32.417	29.563
Non-Top 20 Average	1614.395	1587.823	49.180	47.640	28.385	27.146
95% C.I.	[-91.564, 440.778]	[-252.174, 531.946]	[-7.488, 10.882]	[-10.479, 13.142]	[-35.244, 63.315]	[-33.600, 68.340]
No. of Top 20s	13	13	9	9	12	12
No. of Non-Top 20s	38	38	50	50	39	39

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: This table presents the P-BOT results by gender (male vs. female), parental status (parents vs. non-parents), age (born pre- vs. post-1972) and school rank (attended vs. did not attend a top-20 law school) from both the simple (difference in means) and non-parametric (local polynomial smoothing) approaches. The three panels display results for the yearly billable hours, total weekly hours, and new matters specifications. Promotion probabilities (propensity scores) are estimated using OLS, with lasso to select predictors; the original set of predictors included all original covariates, their squares, and their interactions. The samples include those who are marginally promoted (promoted and in the bottom 30th percentile of propensity score). Standard errors are bootstrapped with 500 repetitions. Data is restricted to those who were promoted between waves II and III.

Table B4: P-BOT Results Using Heteroskedastic Probit to Estimate Promotion Probability, Marginal 30% of Promoted Partners

	Billable Hours		Total Hours		New Matters	
	Simple Version	Non-Parametric	Simple Version	Non-Parametric	Simple Version	Non-Parametric
Panel A: Gender						
Point Estimate (F-M)	-285.652***	-428.137***	-2.471	-2.476	0.597	-2.758
Female Average	1466.040	1390.488	46.529	47.243	11.708	5.809
Male Average	1751.692	1818.624	49.000	49.720	11.111	8.568
95% C.I.	[-503.224, -97.829]	[-654.186, -215.487]	[-9.802, 3.952]	[-9.074, 4.198]	[-9.976, 14.952]	[-9.014, 4.592]
No. of Females	25	25	34	34	24	24
No. of Males	26	26	25	25	27	27
Panel B: Parental Status						
Point Estimate (T-NT)	-36.735	-65.742	4.591	5.288	-1.647	-1.613
Top 20 Average	1587.176	1551.311	51.000	52.697	10.294	6.160
Non-Top 20 Average	1623.912	1617.053	46.409	47.409	11.941	7.773
95% C.I.	[-245.629, 234.679]	[-367.853, 237.260]	[-0.996, 11.116]	[-2.511, 12.892]	[-14.441, 10.894]	[-9.093, 8.039]
No. of Top 20s	17	17	15	15	17	17
No. of Non-Top 20s	34	34	44	44	34	34
Panel C: Age						
Point Estimate (Pre-Post)	-98.304	-78.989	-0.942	1.605	-15.598***	-8.986**
Pre-72 Average	1557.696	1555.554	47.129	49.181	3.440	3.515
Post-72 Average	1656.000	1634.544	48.071	47.576	19.038	12.502
95% C.I.	[-298.961, 108.889]	[-396.545, 234.953]	[-7.147, 4.716]	[-5.162, 8.476]	[-24.981, -4.892]	[-20.638, -1.469]
No. of Pre-72s	23	23	31	31	25	25
No. of Post-72s	28	28	28	28	26	26
Panel D: Law School Rank						
Point Estimate (T-NT)	-36.735	-65.742	4.591	5.288	-1.647	-1.613
Top 20 Average	1587.176	1551.311	51.000	52.697	10.294	6.160
Non-Top 20 Average	1623.912	1617.053	46.409	47.409	11.941	7.773
95% C.I.	[-245.629, 234.679]	[-367.853, 237.260]	[-0.996, 11.116]	[-2.511, 12.892]	[-14.441, 10.894]	[-9.093, 8.039]
No. of Top 20s	17	17	15	15	17	17
No. of Non-Top 20s	34	34	44	44	34	34

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: This table presents the P-BOT results by gender (male vs. female), parental status (parents vs. non-parents), age (born pre- vs. post-1972) and school rank (attended vs. did not attend a top-20 law school) from both the simple (difference in means) and non-parametric (local polynomial smoothing) approaches. The three panels display results for the yearly billable hours, total weekly hours, and new matters specifications. Promotion probabilities (propensity scores) are estimated using a heteroskedastic probit model that allows for non-constant variance in indicators of productivity (billable or total hours and new matters). The samples include those who are marginally promoted (promoted and in the bottom 30th percentile of propensity score). Standard errors are bootstrapped with 500 repetitions. Data is restricted to those who were promoted between waves II and III.

Table B5: P-BOT Results Using Probit to Estimate Promotion Probability, Marginal 30% of Promoted Partners Reporting No Discrimination in Wave II

	Billable Hours		Total Hours		New Matters	
	Simple Version	Non-Parametric	Simple Version	Non-Parametric	Simple Version	Non-Parametric
Panel A: Gender						
Point Estimate (F-M)	-194.446	-490.575***	-0.777	-2.981	1.087	-2.324
Female Average	1588.400	1452.978	48.348	46.990	9.625	3.896
Male Average	1782.846	1943.553	49.125	49.971	8.538	6.220
95% C.I.	[-472.153, 71.801]	[-772.380, -153.347]	[-8.324, 5.833]	[-15.159, 6.619]	[-10.609, 13.833]	[-8.630, 3.531]
No. of Females	15	15	23	23	16	16
No. of Males	26	26	24	24	26	26
Panel B: Parental Status						
Point Estimate (P-NP)	-80.189	77.437	-2.531	1.959	11.837**	7.266***
Parent Average	1678.458	1756.047	47.560	49.635	13.462	8.425
Non-Parent Average	1758.647	1678.611	50.091	47.676	1.625	1.159
95% C.I.	[-304.489, 210.000]	[-338.431, 451.829]	[-9.274, 3.880]	[-7.085, 13.095]	[1.944, 22.514]	[3.189, 13.087]
No. of Parents	24	24	25	25	26	26
No. of Non-Parents	17	17	22	22	16	16
Panel C: Age						
Point Estimate (Pre-Post)	-194.638	-133.623	-4.672	-6.595	-11.259**	-4.686
Pre-72 Average	1616.762	1664.879	46.458	45.175	3.591	3.487
Post-72 Average	1811.400	1798.502	51.130	51.770	14.850	8.173
95% C.I.	[-466.321, 78.140]	[-515.961, 272.697]	[-11.322, 1.720]	[-18.244, 3.197]	[-22.300, -1.381]	[-16.980, 1.388]
No. of Pre-72s	21	21	24	24	22	22
No. of Post-72s	20	20	23	23	20	20
Panel D: Law School Rank						
Point Estimate (T-NT)	22.142	74.944	5.151	8.625*	-3.083	-0.963
Top 20 Average	1727.909	1765.851	52.800	55.576	6.750	4.562
Non-Top 20 Average	1705.767	1690.907	47.649	46.951	9.833	5.525
95% C.I.	[-310.000, 328.313]	[-347.020, 456.033]	[-2.170, 12.011]	[-1.641, 17.098]	[-14.862, 10.181]	[-8.255, 6.664]
No. of Top 20s	11	11	10	10	12	12
No. of Non-Top 20s	30	30	37	37	30	30

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: This table presents the P-BOT results by gender (male vs. female), parental status (parents vs. non-parents), age (born pre- vs. post-1972) and school rank (attended vs. did not attend a top-20 law school) from both the simple (difference in means) and non-parametric (local polynomial smoothing) approaches. The three panels display results for the yearly billable hours, total weekly hours, and new matters specifications. Promotion probabilities (propensity scores) are estimated using a probit model, and the samples include those who are marginally promoted (promoted and in the bottom 30th percentile of propensity score) and do not report any experiences of discrimination in Wave II. Standard errors are bootstrapped with 500 repetitions. Data is restricted to those who were promoted between waves II and III.

Table B6: P-BOT Results Using Probit to Estimate Promotion Probability, Marginal 30% of Promoted Partners Reporting No Discrimination in Wave III

	Billable Hours		Total Hours		New Matters	
	Simple Version	Non-Parametric	Simple Version	Non-Parametric	Simple Version	Non-Parametric
Panel A: Gender						
Point Estimate (F-M)	-326.350**	-585.801***	-4.023	-5.594	4.745	-2.081
Female Average	1480.067	1346.704	47.120	44.767	13.875	4.039
Male Average	1806.417	1932.505	51.143	50.361	9.130	6.120
95% C.I.	[-553.854, -52.786]	[-890.155, -265.682]	[-10.326, 4.615]	[-16.338, 3.659]	[-9.077, 23.099]	[-8.884, 4.297]
No. of Females	15	15	25	25	16	16
No. of Males	24	24	21	21	23	23
Panel B: Parental Status						
Point Estimate (P-NP)	-34.723	63.197	-1.758	-0.031	5.469	4.339*
Parent Average	1666.652	1692.576	48.192	47.280	13.040	6.839
Non-Parent Average	1701.375	1629.379	49.950	47.310	7.571	2.500
95% C.I.	[-308.008, 231.147]	[-407.946, 495.313]	[-10.413, 4.254]	[-9.931, 11.580]	[-14.586, 18.462]	[-0.863, 10.130]
No. of Parents	23	23	26	26	25	25
No. of Non-Parents	16	16	20	20	14	14
Panel C: Age						
Point Estimate (Pre-Post)	-260.782*	-149.871	-3.576	-4.999	-15.442***	-5.788*
Pre-72 Average	1553.850	1600.049	47.091	44.737	3.158	2.856
Post-72 Average	1814.632	1749.920	50.667	49.736	18.600	8.643
95% C.I.	[-514.208, 18.308]	[-600.079, 292.412]	[-11.432, 2.957]	[-16.766, 5.963]	[-30.476, -2.905]	[-14.778, 0.412]
No. of Pre-72s	20	20	22	22	19	19
No. of Post-72s	19	19	24	24	20	20
Panel D: Law School Rank						
Point Estimate (T-NT)	-37.248	103.773	3.761	9.947**	3.380	1.349
Top 20 Average	1653.200	1731.267	51.818	55.629	13.417	5.966
Non-Top 20 Average	1690.448	1627.494	48.057	45.682	10.037	4.617
95% C.I.	[-351.514, 321.818]	[-387.076, 512.639]	[-1.639, 13.208]	[0.643, 19.018]	[-11.471, 19.429]	[-4.282, 11.470]
No. of Top 20s	10	10	11	11	12	12
No. of Non-Top 20s	29	29	35	35	27	27

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: This table presents the P-BOT results by gender (male vs. female), parental status (parents vs. non-parents), age (born pre- vs. post-1972) and school rank (attended vs. did not attend a top-20 law school) from both the simple (difference in means) and non-parametric (local polynomial smoothing) approaches. The three panels display results for the yearly billable hours, total weekly hours, and new matters specifications. Promotion probabilities (propensity scores) are estimated using a probit model, and the samples include those who are marginally promoted (promoted and in the bottom 30th percentile of propensity score) and do not report any experiences of discrimination in Wave III. Standard errors are bootstrapped with 500 repetitions. Data is restricted to those who were promoted between waves II and III.

Table B7: P-BOT Results Using Probit to Estimate Promotion Probability, No Percentile Cutoff

	Billable Hours		Total Hours		New Matters	
	Simple Version	Non-Parametric	Simple Version	Non-Parametric	Simple Version	Non-Parametric
Panel A: Gender						
Point Estimate (F-M)	-152.665***	-244.411***	-4.636**	-4.975***	0.598	3.221
Female Average	1608.463	1546.855	48.470	47.316	21.742	14.060
Male Average	1761.127	1791.267	53.106	52.292	21.144	10.840
95% C.I.	[-262.727, -44.255]	[-399.237, -84.604]	[-8.296, -1.250]	[-8.896, -0.790]	[-11.021, 13.349]	[-5.200, 12.715]
No. of Females	67	67	83	83	66	66
No. of Males	102	102	113	113	104	104
Panel B: Parental Status						
Point Estimate (P-NP)	-56.094	-69.416	-0.673	-0.866	5.541	4.484
Parent Average	1679.029	1658.211	50.872	49.675	23.495	13.976
Non-Parent Average	1735.123	1727.627	51.544	50.541	17.954	9.492
95% C.I.	[-167.971, 60.569]	[-237.176, 98.158]	[-4.313, 2.551]	[-4.870, 2.882]	[-7.902, 18.096]	[-3.024, 12.093]
No. of Parents	104	104	117	117	105	105
No. of Non-Parents	65	65	79	79	65	65
Panel C: Age						
Point Estimate (Pre-Post)	-136.393***	-96.335	-1.024	-1.055	10.208	-5.991
Pre-72 Average	1618.284	1633.811	50.532	49.438	27.621	8.867
Post-72 Average	1754.676	1730.146	51.556	50.493	17.413	14.859
95% C.I.	[-240.490, -30.654]	[-245.427, 68.246]	[-4.347, 2.606]	[-4.708, 3.136]	[-5.166, 26.615]	[-13.004, 2.013]
No. of Pre-72s	67	67	79	79	66	66
No. of Post-72s	102	102	117	117	104	104
Panel D: Law School Rank						
Point Estimate (T-NT)	35.558	8.468	1.242	1.156	-4.160	-3.890
Top 20 Average	1725.431	1692.707	52.036	50.907	18.440	9.393
Non-Top 20 Average	1689.873	1684.239	50.794	49.750	22.600	13.283
95% C.I.	[-112.723, 160.261]	[-201.829, 216.441]	[-2.633, 4.826]	[-2.699, 5.494]	[-18.245, 10.309]	[-11.792, 4.448]
No. of Top 20s	51	51	55	55	50	50
No. of Non-Top 20s	118	118	141	141	120	120

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: This table presents the P-BOT results by gender (male vs. female), parental status (parents vs. non-parents), age (born pre- vs. post-1972) and school rank (attended vs. did not attend a top-20 law school) from both the simple (difference in means) and non-parametric (local polynomial smoothing) approaches. The three panels display results for the yearly billable hours, total weekly hours, and new matters specifications. Promotion probabilities (propensity scores) are estimated using a probit model, and the samples include those who were promoted (no percentile cutoffs). Standard errors are bootstrapped with 500 repetitions. Sample is restricted to those who were promoted between waves II and III.

Table B8: P-BOT Results Using Probit to Estimate Promotion Probability, Results From Those Promoted in Either Wave II or Wave III, Marginal 30% of Promoted Partners

	Billable Hours		Total Hours		New Matters	
	<i>Simple Version</i>	<i>Non-Parametric</i>	<i>Simple Version</i>	<i>Non-Parametric</i>	<i>Simple Version</i>	<i>Non-Parametric</i>
Panel A: Gender						
Point Estimate (F-M)	-228.529**	-197.525	-2.108	-2.502	-10.567	-15.742
Female Average	1676.675	1701.398	48.927	48.769	11.364	11.904
Male Average	1905.204	1898.924	51.036	51.271	21.931	27.646
95% C.I.	[-420.968, -13.569]	[-462.344, 47.716]	[-6.483, 1.765]	[-7.040, 1.579]	[-30.495, 3.046]	[-41.840, 6.797]
No. of Females	40	40	55	55	44	44
No. of Males	54	54	56	56	58	58
Panel B: Parental Status						
Point Estimate (P-NP)	-30.186	45.462	1.115	0.733	-0.634	4.168
Parent Average	1785.800	1832.581	50.765	50.388	16.938	22.253
Non-Parent Average	1815.986	1787.119	49.649	49.655	17.571	18.085
95% C.I.	[-325.846, 187.800]	[-222.631, 314.211]	[-3.727, 4.610]	[-3.548, 4.535]	[-17.154, 16.860]	[-21.599, 31.350]
No. of Parents	25	25	34	34	32	32
No. of Non-Parents	69	69	77	77	70	70
Panel C: Age						
Point Estimate (Pre-Post)	-186.336*	-114.256	-2.793	-2.148	-6.958	-6.544
Pre-72 Average	1706.860	1734.316	48.582	48.723	13.962	15.826
Post-72 Average	1893.196	1848.572	51.375	50.871	20.920	22.369
95% C.I.	[-429.193, 20.724]	[-399.026, 157.008]	[-7.250, 1.330]	[-5.963, 1.604]	[-26.573, 9.486]	[-31.679, 18.173]
No. of Pre-72s	43	43	55	55	52	52
No. of Post-72s	51	51	56	56	50	50
Panel D: Law School Rank						
Point Estimate (T-NT)	225.205**	255.066**	5.174*	4.049	-17.878***	-21.579***
Top 20 Average	1980.455	1982.251	54.000	52.833	3.000	3.046
Non-Top 20 Average	1755.250	1727.185	48.826	48.784	20.878	24.624
95% C.I.	[7.785, 451.656]	[0.045, 497.822]	[-0.088, 11.738]	[-0.953, 9.832]	[-31.042, -8.469]	[-38.846, -8.256]
No. of Top 20s	22	22	25	25	20	20
No. of Non-Top 20s	72	72	86	86	82	82

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: This table presents the P-BOT results by gender (male vs. female), parental status (parents vs. non-parents), age (born pre- vs. post-1972) and school rank (attended vs. did not attend a top-20 law school) from both the simple (difference in means) and non-parametric (local polynomial smoothing) approaches. The three panels display results for the yearly billable hours, total weekly hours, and new matters specifications. Promotion probabilities (propensity scores) are estimated using a probit model, and the samples include those who are marginally promoted (promoted and in the bottom 30th percentile of propensity score). Standard errors are bootstrapped with 500 repetitions. Sample includes both those promoted between waves I and II and those promoted between waves II and III. Note that we exclude promotion desire and use total hours as a productivity metric across all samples when computing propensity scores due to lack of recorded promotion desire and billable hours in Wave I data.

Table B9: P-BOT Results Using Probit to Estimate Promotion Probability, Results From Those Promoted in Either Wave II or Wave III, No Percentile Cutoff

	Billable Hours		Total Hours		New Matters	
	<i>Simple Version</i>	<i>Non-Parametric</i>	<i>Simple Version</i>	<i>Non-Parametric</i>	<i>Simple Version</i>	<i>Non-Parametric</i>
Panel A: Gender						
Point Estimate (F-M)	-156.596***	-182.771***	-4.571***	-3.961***	-5.437	-8.829
Female Average	1621.446	1628.374	48.326	47.985	18.356	15.963
Male Average	1778.042	1811.145	52.898	51.946	23.793	24.792
95% C.I.	[-259.618, -50.264]	[-335.798, -44.870]	[-6.996, -2.044]	[-6.356, -1.450]	[-15.790, 3.457]	[-19.882, 1.319]
No. of Females	121	121	144	144	132	132
No. of Males	191	191	225	225	208	208
Panel B: Parental Status						
Point Estimate (P-NP)	-101.442*	-87.355	-1.305	-1.207	1.761	2.483
Parent Average	1666.590	1679.472	50.438	49.474	22.552	22.402
Non-Parent Average	1768.032	1766.827	51.743	50.681	20.792	19.919
95% C.I.	[-210.032, 3.566]	[-226.371, 49.494]	[-3.730, 1.375]	[-3.938, 1.184]	[-8.500, 11.203]	[-10.297, 13.336]
No. of Parents	156	156	178	178	172	172
No. of Non-Parents	156	156	191	191	168	168
Panel C: Age						
Point Estimate (Pre-Post)	-177.353***	-161.417**	-1.496	-1.224	6.947	1.186
Pre-72 Average	1613.287	1644.766	50.250	49.536	25.626	21.524
Post-72 Average	1790.639	1806.183	51.746	50.760	18.679	20.339
95% C.I.	[-288.854, -70.218]	[-298.758, -23.156]	[-4.067, 0.988]	[-3.738, 1.222]	[-3.942, 18.551]	[-11.058, 13.059]
No. of Pre-72s	129	129	156	156	147	147
No. of Post-72s	183	183	213	213	193	193
Panel D: Law School Rank						
Point Estimate (T-NT)	83.751	116.968	1.445	2.387	-4.512	-10.000
Top 20 Average	1779.588	1819.650	52.218	52.005	18.205	13.216
Non-Top 20 Average	1695.836	1702.682	50.773	49.619	22.718	23.216
95% C.I.	[-33.686, 201.092]	[-26.895, 260.745]	[-1.279, 4.846]	[-1.013, 5.878]	[-16.481, 10.522]	[-21.156, 2.081]
No. of Top 20s	80	80	87	87	78	78
No. of Non-Top 20s	232	232	282	282	262	262

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: This table presents the P-BOT results by gender (male vs. female), parental status (parents vs. non-parents), age (born pre- vs. post-1972) and school rank (attended vs. did not attend a top-20 law school) from both the simple (difference in means) and non-parametric (local polynomial smoothing) approaches. The three panels display results for the yearly billable hours, total weekly hours, and new matters specifications. Promotion probabilities (propensity scores) are estimated using a probit model, and the samples include those who were promoted (no percentile cutoffs). Standard errors are bootstrapped with 500 repetitions. Sample includes both those promoted between waves I and II and those promoted between waves II and III. Note that we exclude promotion desire and use total hours as a productivity metric across all samples when computing propensity scores due to lack of recorded promotion desire and billable hours in Wave I data.

Table B10: P-BOT Results Using Probit to Estimate Promotion Probability, Marginal 30% of Nonpromoted Associates

	Billable Hours		Total Hours		New Matters	
	Simple Version	Non-Parametric	Simple Version	Non-Parametric	Simple Version	Non-Parametric
Panel A: Gender						
Point Estimate (F-M)	-256.762	-147.253	-8.659**	-11.796**	0.519	2.327
Female Average	1380.571	1397.221	42.100	40.597	7.000	10.762
Male Average	1637.333	1544.473	50.759	52.393	6.481	8.435
95% C.I.	[-665.739, 162.360]	[-629.693, 369.883]	[-16.107, -1.278]	[-21.448, -1.879]	[-8.333, 9.091]	[-10.958, 17.578]
No. of Females	7	7	30	30	9	9
No. of Males	24	24	54	54	27	27
Panel B: Parental Status						
Point Estimate (P-NP)	78.351	35.994	-6.694**	-6.797*	-0.247	6.972
Parent Average	1609.684	1508.213	44.878	45.271	6.522	11.721
Non-Parent Average	1531.333	1472.219	51.571	52.068	6.769	4.748
95% C.I.	[-337.840, 490.500]	[-492.515, 701.849]	[-12.896, -0.646]	[-15.655, 0.912]	[-5.037, 8.375]	[-3.385, 15.864]
No. of Parents	19	19	49	49	23	23
No. of Non-Parents	12	12	35	35	13	13
Panel C: Age						
Point Estimate (Pre-Post)	-64.754	66.415	1.348	2.121	0.700	-0.966
Pre-72 Average	1539.667	1542.579	48.485	49.736	7.000	8.510
Post-72 Average	1604.421	1476.165	47.137	47.615	6.300	9.476
95% C.I.	[-377.975, 244.983]	[-300.700, 418.372]	[-4.801, 7.000]	[-4.465, 10.253]	[-6.463, 7.896]	[-10.875, 7.858]
No. of Pre-72s	12	12	33	33	16	16
No. of Post-72s	19	19	51	51	20	20
Panel D: Law School Rank						
Point Estimate (T-NT)	-39.641	108.087	3.522	1.827	-3.754	-0.910
Top 20 Average	1551.222	1571.676	50.350	49.667	3.900	8.554
Non-Top 20 Average	1590.864	1463.588	46.828	47.840	7.654	9.464
95% C.I.	[-386.048, 409.076]	[-314.309, 464.977]	[-3.480, 9.540]	[-5.360, 9.725]	[-9.462, 7.444]	[-14.414, 13.252]
No. of Top 20s	9	9	20	20	10	10
No. of Non-Top 20s	22	22	64	64	26	26

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: This table presents the P-BOT results by gender (male vs. female), parental status (parents vs. non-parents), age (born pre- vs. post-1972) and school rank (attended vs. did not attend a top-20 law school) from both the simple (difference in means) and non-parametric (local polynomial smoothing) approaches. The three panels display results for the yearly billable hours, total weekly hours, and new matters specifications. Promotion probabilities (propensity scores) are estimated using a probit model, and the samples include those who are marginally nonpromoted (not promoted and in the top 70th percentile of propensity score). Standard errors are bootstrapped with 500 repetitions. Data is restricted to those who were not promoted between waves II and III.

Table B11: P-BOT Results Using Probit to Estimate Promotion Probability, Marginal 30% of Promoted Partners

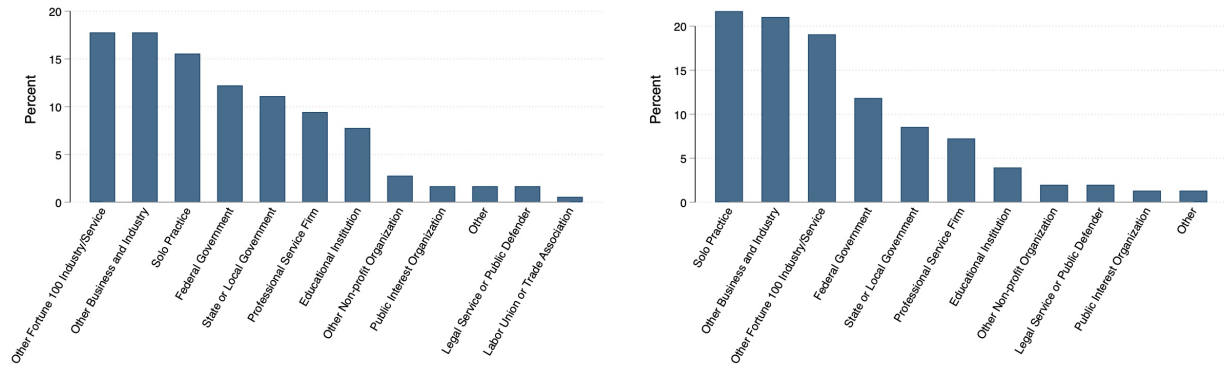
	Pro-Bono On-Job Hours		Pro-Bono Off-Job Hours		All Pro-Bono Hours	
	<i>Simple Version</i>	<i>Non-Parametric</i>	<i>Simple Version</i>	<i>Non-Parametric</i>	<i>Simple Version</i>	<i>Non-Parametric</i>
Panel A: Gender						
Point Estimate (F-M)	8.194	8.882	-6.481	-4.182	3.066	3.932
Female Average	23.810	14.582	10.227	8.987	32.955	22.106
Male Average	15.615	5.700	16.708	13.168	29.889	18.173
95% C.I.	[-20.135, 33.200]	[-5.169, 25.093]	[-28.875, 14.286]	[-32.117, 11.917]	[-30.782, 32.360]	[-25.843, 24.870]
No. of Females	22	22	22	22	22	22
No. of Males	29	29	29	29	29	29
Panel B: Parental Status						
Point Estimate (P-NP)	-22.670	-11.843	-1.577	-4.973	-23.286	-14.916
Parent Average	9.630	5.658	12.923	8.811	21.286	14.074
Non-Parent Average	32.300	17.501	14.500	13.784	44.571	28.991
95% C.I.	[-48.634, 4.695]	[-33.227, 5.133]	[-24.832, 14.870]	[-32.099, 12.511]	[-59.717, 7.945]	[-43.116, 7.711]
No. of Parents	29	29	29	29	29	29
No. of Non-Parents	22	22	22	22	22	22
Panel C: Age						
Point Estimate (Pre-Post)	-1.417	10.613	-12.686	-16.558**	-12.628	-5.527
Pre-72 Average	18.583	14.838	7.542	3.337	25.080	17.736
Post-72 Average	20.000	4.225	20.227	19.895	37.708	23.263
95% C.I.	[-25.284, 25.962]	[-2.037, 25.337]	[-33.456, 6.833]	[-39.718, -0.116]	[-42.143, 23.400]	[-32.527, 16.087]
No. of Pre-72s	26	26	26	26	26	26
No. of Post-72s	25	25	25	25	25	25
Panel D: Law School Rank						
Point Estimate (T-NT)	29.006	13.943	2.000	-0.181	32.729	14.593
Top 20 Average	39.643	18.787	15.000	10.791	54.643	29.578
Non-Top 20 Average	10.636	4.844	13.000	10.972	21.914	14.985
95% C.I.	[-5.473, 70.667]	[-2.546, 37.371]	[-17.216, 43.678]	[-16.700, 33.260]	[-11.824, 82.936]	[-8.780, 49.728]
No. of Top 20s	14	14	14	14	14	14
No. of Non-Top 20s	37	37	37	37	37	37

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: This table presents the P-BOT results by gender (male vs. female), parental status (parents vs. non-parents), age (born pre- vs. post-1972) and school rank (attended vs. did not attend a top-20 law school) from both the simple (difference in means) and non-parametric (local polynomial smoothing) approaches. The three panels display results for the pro-bono hours on the job, not on the job, and overall. Promotion probabilities (propensity scores) are estimated using a probit model (identical to the main specification), and the samples include those who are marginally promoted (promoted and in the bottom 30th percentile of propensity score). Standard errors are bootstrapped with 500 repetitions. Data is restricted to those who were promoted between waves II and III.

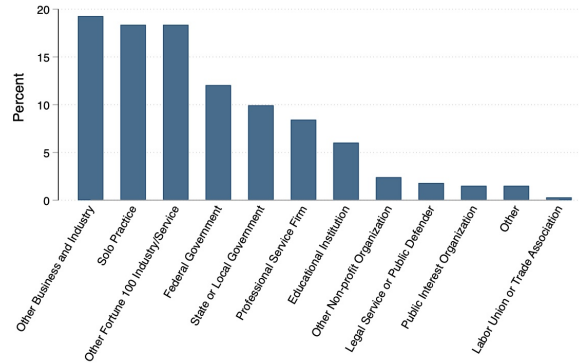
C Appendix: Additional Analyses

Figure C1: Wave II Sectors of Work (Wave I Leavers)



(a) Female

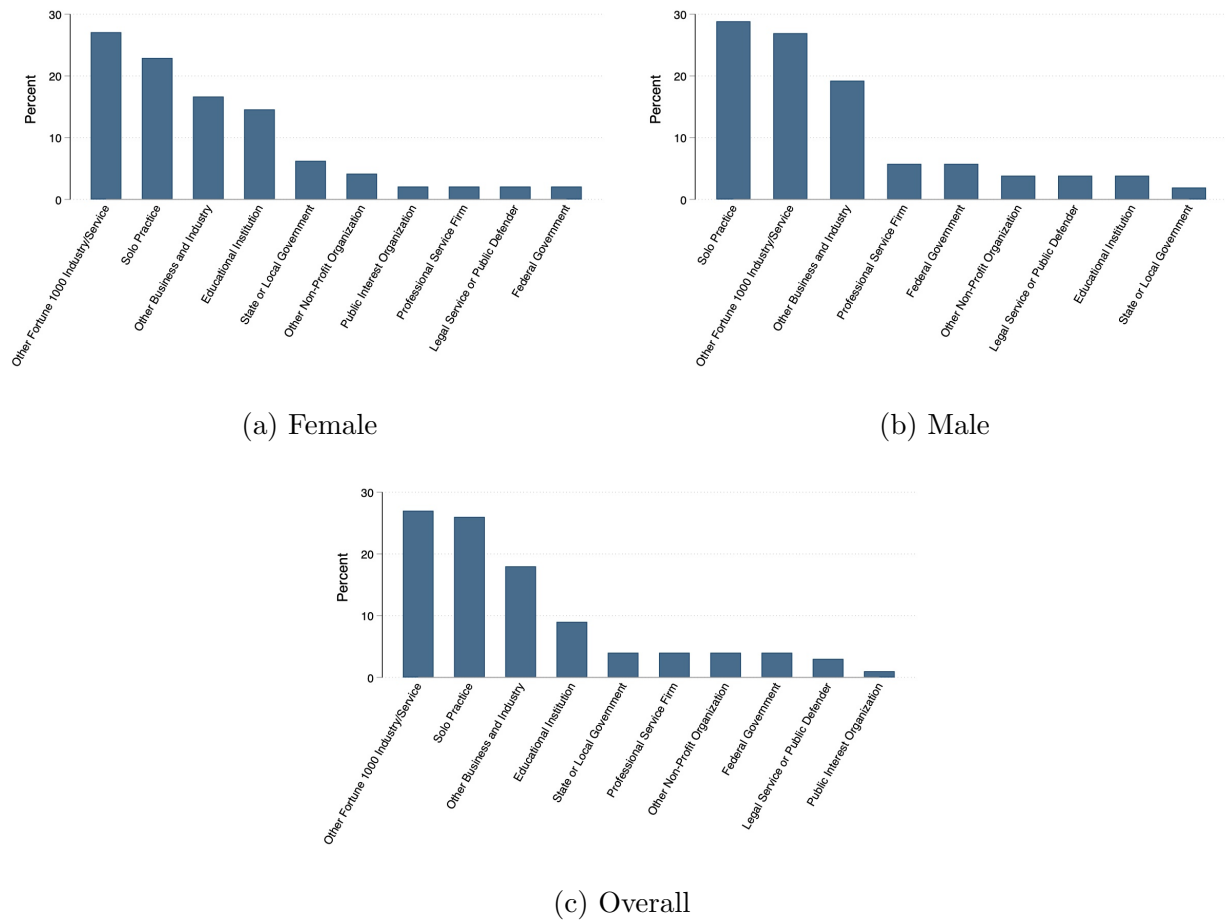
(b) Male



(c) Overall

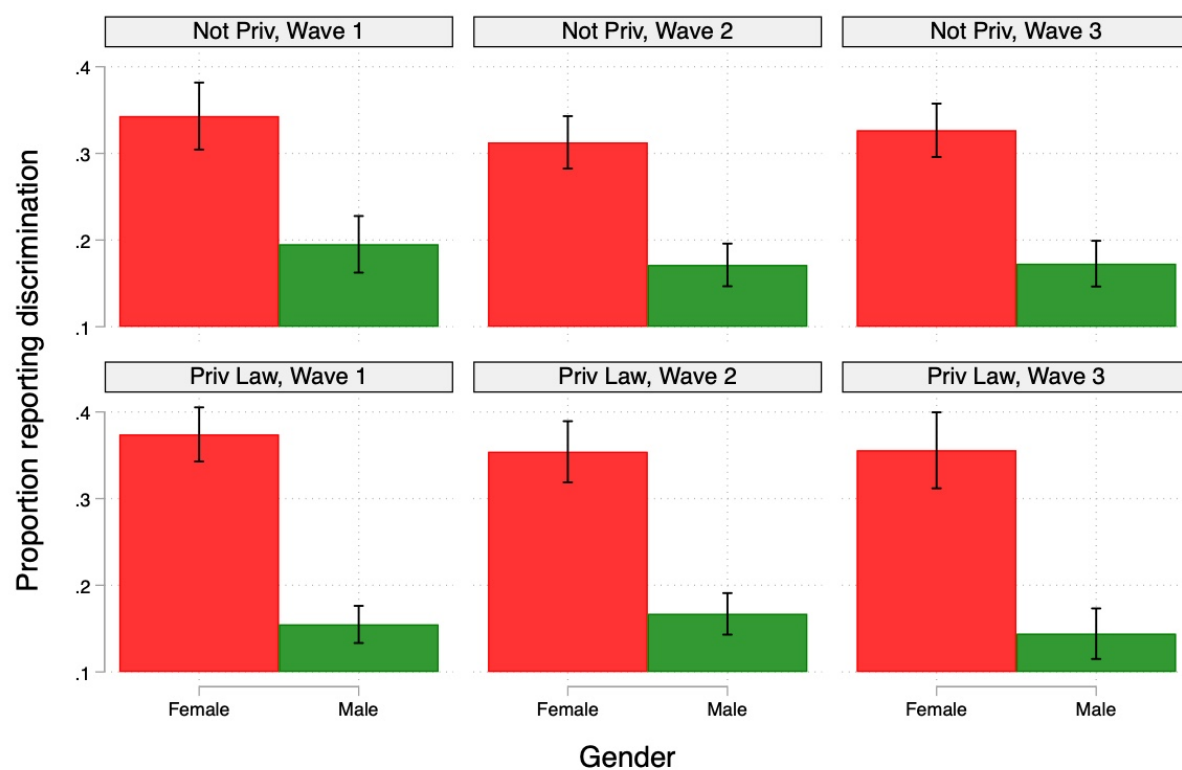
Note: Sample of all Wave I private law firm associates who leave private law after Wave I.

Figure C2: Wave III Sectors of Work (Wave II Leavers)



Note: Sample of all Wave I private law firm associates who leave private law after Wave II.

Figure C3: Experiences of Discrimination by Gender, Private Law Firm, and Wave



Note: Uses the full sample to display the proportion of observations who report any form of discrimination, split by gender, private law firm (those in a private law firm vs. those not; could include those not in law firms), and wave.

Table C1: Descriptive Statistics of Private Law Firm Leavers vs. Stayers

	Wave I Leavers			Wave II Leavers			Stayers		
	Mean	SD	N	Mean	SD	N	Mean	SD	N
Female	0.542	0.499	332	0.480	0.502	100	0.446	0.498	377
Born 1972 or Earlier	0.509	0.501	332	0.410	0.494	100	0.451	0.498	377
White or Asian	0.814	0.390	323	0.780	0.416	100	0.859	0.349	375
Attended Top 20 Law School	0.337	0.474	332	0.300	0.461	100	0.253	0.436	367
Parent in Wave I	0.229	0.421	332	0.370	0.485	100	0.260	0.439	377
Wave I Discrimination	0.278	0.449	331	0.320	0.469	100	0.271	0.445	376
Promoted After Wave I				0.110	0.314	100	0.271	0.445	376
Promotion Desire (Start of Wave II)				6.299	3.333	77	7.567	2.878	275
Parent in Wave II				0.639	0.483	97	0.640	0.481	375
Wave II Discrimination				0.286	0.454	98	0.248	0.432	375
Promoted After Wave II							0.427	0.495	377
Parent in Wave III							0.796	0.404	372
Wave III Discrimination							0.211	0.409	374

Note: Initial sample includes all Wave I private law firm associates. Wave I leavers are those who leave private law after Wave I, while Wave II leavers leave private law after Wave II and stayers remain in private law in all three waves.

Table C2: Likelihood of Private Law Attrition

	Wave I Leavers		Wave II Leavers	
	Linear	Probit	Linear	Probit
Female	0.052 (0.038)	0.050 (0.037)	-0.064 (0.077)	-0.053 (0.069)
Born 1972 or Earlier	0.088** (0.034)	0.086*** (0.033)	-0.090* (0.051)	-0.105** (0.049)
White or Asian	0.002 (0.043)	0.003 (0.042)	-0.109 (0.072)	-0.088 (0.054)
Attended Top 20 Law School	0.085** (0.037)	0.082** (0.035)	-0.001 (0.053)	0.010 (0.049)
Parent in Wave I	-0.095* (0.050)	-0.097* (0.052)	0.151* (0.080)	0.159** (0.076)
Female x Wave I Parent	0.064 (0.073)	0.068 (0.073)	0.066 (0.118)	0.084 (0.103)
Wave I Discrimination	-0.006 (0.037)	-0.007 (0.036)	0.014 (0.063)	0.011 (0.052)
Promoted After Wave I			-0.173*** (0.054)	-0.266** (0.113)
Promotion Desire (Start of Wave II)			-0.019** (0.009)	-0.018** (0.007)
Parent in Wave II			-0.035 (0.073)	-0.033 (0.071)
Female x Wave II Parent			-0.002 (0.100)	-0.009 (0.096)
Wave II Discrimination			0.042 (0.062)	0.033 (0.054)
Constant	0.407*** (0.064)		0.412*** (0.136)	
Observations	911	911	330	330
Adjusted / Pseudo R^2	0.029	0.035	0.055	0.147

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Probit columns report marginal effects. Initial sample includes all Wave I private law firm associates, though the observed sample is smaller due to missingness in variables. Wave I leavers are those who leave private law after Wave I, while Wave II leavers leave private law after Wave II and stayers remain in private law in all three waves. The first two columns compare Wave I leavers to Wave II leavers and stayers, while the second two columns compare Wave II leavers to stayers. Note that firm region is also included in the regressions, though not depicted.

Table C3: Descriptive Statistics of Sample Leavers vs. Stayers

	Wave I Leavers			Wave II Leavers			Stayers		
	Mean	SD	N	Mean	SD	N	Mean	SD	N
Female	0.573	0.496	178	0.494	0.501	249	0.479	0.500	743
Born 1972 or Earlier	0.244	0.430	390	0.442	0.498	249	0.472	0.500	743
White or Asian	0.791	0.408	172	0.816	0.389	201	0.829	0.377	741
Attended Top 20 Law School	0.260	0.439	377	0.246	0.432	248	0.285	0.452	733
Parent in Wave I	0.272	0.445	390	0.229	0.421	249	0.271	0.445	743
Private Law Firm in Wave I	0.972	0.166	390	0.984	0.126	249	0.977	0.150	743
Wave I Discrimination	0.310	0.463	384	0.218	0.414	248	0.275	0.447	741
Promoted After Wave I				0.120	0.326	249	0.157	0.364	743
Promotion Desire (Start of Wave II)				6.250	3.432	104	7.204	3.079	363
Parent in Wave II				0.571	0.496	245	0.587	0.493	733
Private Law Firm in Wave II				0.643	0.480	249	0.658	0.475	742
Wave II Discrimination				0.224	0.418	245	0.251	0.434	734
Promoted After Wave II							0.233	0.423	743
Parent in Wave III							0.739	0.440	731
Private Law Firm in Wave III							0.542	0.499	732
Wave III Discrimination							0.213	0.410	737

Note: Initial sample includes all Wave I associates. Wave I leavers are those who leave the sample or lack position data after Wave I, while Wave II leavers leave or lack position data after Wave II and stayers are present and have position data in all three waves.

Table C4: Likelihood of Sample Attrition

	Wave I Leavers		Wave II Leavers	
	Linear	Probit	Linear	Probit
Female	-0.019 (0.024)	-0.017 (0.023)	0.006 (0.063)	0.004 (0.060)
Born 1972 or Earlier	0.047** (0.021)	0.046** (0.020)	-0.019 (0.044)	-0.014 (0.041)
White or Asian	-0.011 (0.027)	-0.010 (0.026)	0.030 (0.053)	0.026 (0.052)
Attended Top 20 Law School	0.023 (0.023)	0.021 (0.022)	-0.061 (0.044)	-0.070 (0.046)
Parent in Wave I	-0.054* (0.030)	-0.056* (0.034)	-0.014 (0.067)	-0.007 (0.060)
Private Law Firm in Wave I	-0.090 (0.080)	-0.078 (0.057)	0.218*** (0.055)	0.000 (.)
Female x Wave I Parent	0.068 (0.046)	0.071 (0.046)	-0.001 (0.096)	-0.004 (0.091)
Wave I Discrimination	0.010 (0.023)	0.009 (0.022)	-0.057 (0.046)	-0.054 (0.048)
Promoted After Wave I			0.039 (0.067)	0.038 (0.056)
Promotion Desire (Start of Wave II)			-0.006 (0.006)	-0.005 (0.006)
Parent in Wave II			0.036 (0.061)	0.031 (0.057)
Private Law Firm in Wave II			-0.368* (0.218)	-0.261* (0.143)
Female x Wave II Parent			-0.019 (0.086)	-0.019 (0.081)
Wave II Discrimination			0.030 (0.049)	0.022 (0.045)
Constant	0.217** (0.085)		0.267 (0.249)	
Observations	1093	1093	438	434
Adjusted / Pseudo R^2	0.003	0.026	0.021	0.095

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Probit columns report marginal effects. Initial sample includes all Wave I associates, though the observed sample is smaller due to missingness in variables. Wave I leavers are those who leave the sample or lack position data after Wave I, while Wave II leavers leave or lack position data after Wave II and stayers are present and have position data in all three waves. The first two columns compare Wave I leavers to Wave II leavers and stayers, while the second two columns compare Wave II leavers to stayers. Note that firm region is also included in the regressions, though not depicted.

Table C5: Perceptions of Discrimination on Productivity (Promoted Partners)

	Wave III Billable Hours	Wave III Total Hours	Wave III New Matters
Wave II Diversity Satisfaction	-1.106 (38.373)	0.815 (0.723)	0.487 (3.812)
Wave II Any Discrimination	-155.231 (164.684)	-18.468*** (6.745)	-13.151 (12.479)
Wave II Verbal Discrimination	-15.756 (186.460)	14.243 (9.192)	4.959 (12.982)
Wave II Assignment Discrimination	63.966 (191.112)	-17.623** (7.793)	3.213 (14.568)
Wave II Client Discrimination	-86.846 (212.300)	14.763* (8.682)	6.433 (14.883)
Wave II Colleague Discrimination	245.111 (216.115)	24.401*** (8.605)	-5.373 (16.406)
Female	59.403 (241.083)	-7.173 (6.044)	-15.635 (22.089)
Born in 1972 or Earlier	-178.372 (137.266)	-1.595 (3.090)	-3.908 (11.683)
White or Asian	345.087 (232.617)	4.914 (5.172)	-13.019 (19.484)
Top 20 Law School	173.890 (136.480)	8.837** (3.633)	6.801 (11.105)
Parent	67.356 (162.334)	-4.407 (4.223)	-3.278 (13.479)
Female x Parent	-166.969 (250.509)	6.618 (7.197)	8.853 (21.992)
Constant	1651.764*** (303.063)	44.035*** (8.926)	19.978 (23.141)
Observations	64	87	75
Adjusted R ²	0.214	0.069	-0.180

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Data from all private law firm associates promoted between waves II and III. Note that firm region is also included in the regressions, though not depicted.

Table C6: Pre-Period Perceptions of Discrimination on Promotion

	Whether Promoted					
Diversity Satisfaction	0.007 (0.026)					
Female	-0.373** (0.180)	-0.280*** (0.100)	-0.252*** (0.094)	-0.245*** (0.094)	-0.274*** (0.096)	-0.250*** (0.095)
Diversity Satisfaction x Female	0.027 (0.038)					
Born in 1972 or Earlier	-0.069 (0.061)	-0.079 (0.061)	-0.087 (0.061)	-0.091 (0.060)	-0.069 (0.061)	-0.088 (0.061)
White or Asian	0.054 (0.085)	0.037 (0.082)	0.045 (0.082)	0.055 (0.082)	0.048 (0.082)	0.049 (0.082)
Top 20 Law School	0.037 (0.066)	0.022 (0.065)	0.026 (0.065)	0.006 (0.066)	0.014 (0.065)	0.022 (0.066)
Parent	-0.051 (0.081)	-0.032 (0.080)	-0.021 (0.079)	-0.032 (0.080)	-0.041 (0.079)	-0.014 (0.081)
Female x Parent	0.263** (0.122)	0.234* (0.121)	0.222* (0.121)	0.255** (0.121)	0.251** (0.122)	0.228* (0.123)
Any Discrimination		-0.136 (0.118)				
Any Discrimination x Female		0.167 (0.154)				
Verbal Discrimination			-0.026 (0.182)			
Verbal Discrimination x Female			0.068 (0.224)			
Assignment Discrimination				-0.280 (0.220)		
Assignment Discrimination x Female				0.024 (0.266)		
Client Discrimination					-0.300 (0.187)	
Client Discrimination x Female					0.250 (0.236)	
Colleague Discrimination						0.145 (0.186)
Colleague Discrimination x Female						-0.003 (0.258)
Constant	0.667*** (0.144)	0.716*** (0.099)	0.686*** (0.096)	0.696*** (0.099)	0.706*** (0.094)	0.670*** (0.096)
Observations	257	267	267	264	266	264
Adjusted R ²	0.022	0.020	0.014	0.032	0.027	0.017

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Data from all private law firm associates in Wave II.

Table C7: Promotion Desire and Demographics on Perceptions of Discrimination

	Simple	Demographics	Female	Male	Pre-1972	Post-1972	White/Asian	Non-White/Asian
Promotion Desire	-0.011 (0.009)	-0.005 (0.010)	0.002 (0.014)	-0.014 (0.012)	-0.008 (0.013)	-0.013 (0.012)	-0.017 (0.011)	0.013 (0.018)
Female		0.122 (0.087)						
Born in 1972 or Earlier		0.014 (0.051)						
White or Asian		-0.034 (0.076)						
Top 20 Law School		-0.113** (0.052)						
Parent		-0.074 (0.062)						
Female x Parent		0.066 (0.108)						
Constant	0.308*** (0.076)	0.287** (0.113)	0.297*** (0.102)	0.261** (0.112)	0.281** (0.111)	0.324*** (0.102)	0.352*** (0.091)	0.157 (0.127)
Observations	267	267	118	149	114	153	226	41
Adjusted R ²	0.002	0.034	-0.008	0.002	-0.006	0.003	0.009	-0.016

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Data from all private law firm associates in Wave II. Columns 1 and 2 use the full sample, whereas the remaining columns examine demographic subsets.

Table C8: Promotions on Perceptions (All Waves)

	$\Delta DiversitySatisfaction$	$\Delta AnyDiscrimination$	$\Delta VerbalDiscrimination$	$\Delta AssignmentDiscrimination$	$\Delta ClientDiscrimination$	$\Delta ColleagueDiscrimination$
Promoted	0.489*** (0.145)	-0.036 (0.034)	-0.021 (0.025)	-0.022 (0.022)	-0.004 (0.025)	-0.019 (0.034)
Constant	4.251*** (0.019)	0.257*** (0.005)	0.146*** (0.004)	0.098*** (0.003)	0.078*** (0.003)	0.053*** (0.009)
Observations	2741	3036	3027	3013	3012	1673
Adjusted R ²	0.249	0.228	0.246	0.229	0.110	0.094

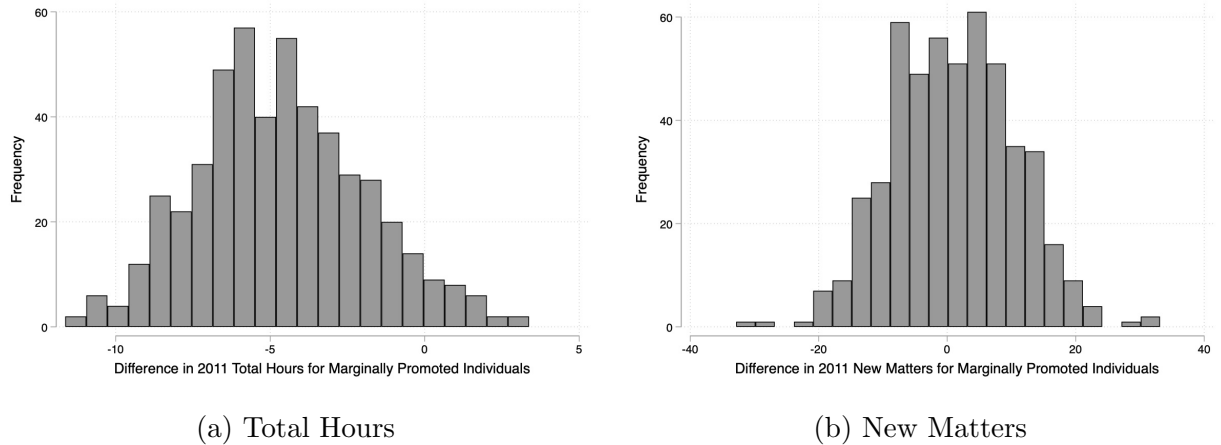
Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: This table presents the results of regressions of various measures of discrimination on promotion. Regressions include individual fixed effects, and standard errors are clustered on the individual level. Estimations use the full sample with no percentile cutoff or promotion desire restriction; however, there is a restriction to those who are private law firm associates in Wave I. Data from all three waves.

D Appendix: Assumption Testing

Figure D1: Perturbation Tests



Note: This figure displays the results of the perturbation test; that is, it shows the distribution of differences in means between the 2011 total hours, as well as new matters, of marginally promoted women and marginally promoted men. To produce these results, we first estimate each individual's propensity score using a probit model. Then, for each promoted individual, we simulate 500 realizations from a standard normal distribution that are subsequently used to simulate propensity scores. We use these scores to redefine samples of marginal individuals and estimate the outcome test within each one.

Table D1: Estimation Sample Stability (Billable Hours Sample)

Excluded Predictor	Share (%) of Marginal Sample in Updated Marginal Sample	Rank correlation Spearman's- ρ
Billable Hours	82.72	0.928
New Matters	90.12	0.853
Promotion Desire	74.07	0.877
Female	90.12	0.980
Born in 1972 or Earlier	93.83	0.984
White or Asian	98.77	1.000
Attended Top 20 Law School	95.06	0.996
Parent	98.77	0.999
Mother	90.12	0.968
Region	93.83	0.954

Note: This table presents the estimation sample stability exercise. We exclude covariates from the probit model for the propensity score and compute (i) the share of the marginally promoted sample also considered marginal under the restricted probit estimation and (ii) the Spearman's- ρ rank correlation statistic for the predicted propensity score across promoted individuals. Data from the billable hours sample.

Table D2: Estimation Sample Stability (Total Hours Sample)

Excluded Predictor	Share (%) of Marginal Sample in Updated Marginal Sample	Rank correlation Spearman's- ρ
Total Hours	90.78	0.949
New Matters	95.04	0.984
Promotion Desire	63.83	0.777
Female	95.74	0.986
Born in 1972 or Earlier	98.58	0.990
White or Asian	90.78	0.954
Attended Top 20 Law School	98.58	0.999
Parent	97.87	0.997
Mother	93.62	0.979
Region	89.36	0.924

Note: This table presents the estimation sample stability exercise. We exclude covariates from the probit model for the propensity score and compute (i) the share of the marginally promoted sample also considered marginal under the restricted probit estimation and (ii) the Spearman's- ρ rank correlation statistic for the predicted propensity score across promoted individuals. Data from the total hours sample.

Table D3: Estimation Sample Stability (New Matters Sample)

Excluded Predictor	Share (%) of Marginal Sample in Updated Marginal Sample	Rank correlation Spearman's- ρ
Billable Hours	82.76	0.903
New Matters	83.91	0.872
Promotion Desire	75.86	0.890
Female	89.66	0.973
Born in 1972 or Earlier	90.80	0.984
White or Asian	98.85	1.000
Attended Top 20 Law School	94.25	0.994
Parent	96.55	0.995
Mother	85.06	0.954
Region	90.80	0.928

Note: This table presents the estimation sample stability exercise. We exclude covariates from the probit model for the propensity score and compute (i) the share of the marginally promoted sample also considered marginal under the restricted probit estimation and (ii) the Spearman's- ρ rank correlation statistic for the predicted propensity score across promoted individuals. Data from the new matters sample.

Table D4: Testing for Monotonicity in Observables (Dep. Variable: Promoted)

	All	Female	Male	Pre-72	Post-72	Top 20	Non-Top 20	Parent	Non-Parent
Panel A: Billable Hours Sample									
2006 Billable Hours	0.102*** (0.028)	0.083* (0.049)	0.102*** (0.034)	0.069 (0.050)	0.121*** (0.032)	0.005 (0.049)	0.147*** (0.029)	0.093** (0.036)	0.126*** (0.045)
2006 New Matters	0.069** (0.030)	0.058** (0.025)	0.107*** (0.028)	0.061** (0.024)	0.392*** (0.090)	0.109*** (0.028)	0.061** (0.029)	0.049** (0.021)	0.175*** (0.051)
Promotion Desire	0.178*** (0.027)	0.237*** (0.031)	0.086* (0.048)	0.180*** (0.045)	0.180*** (0.034)	0.232*** (0.045)	0.154*** (0.034)	0.160*** (0.037)	0.195*** (0.041)
Observations	269	119	150	115	154	80	189	157	112
Panel B: Total Hours Sample									
2006 Total Hours	0.083*** (0.022)	0.045 (0.031)	0.117*** (0.034)	0.054 (0.033)	0.103*** (0.029)	0.147*** (0.033)	0.054* (0.027)	0.075** (0.030)	0.101*** (0.033)
2006 New Matters	0.038* (0.021)	0.044** (0.021)	0.030 (0.045)	0.077*** (0.025)	-0.022 (0.026)	0.155*** (0.030)	0.026 (0.022)	0.042** (0.021)	0.027 (0.045)
Promotion Desire	0.148*** (0.020)	0.167*** (0.027)	0.121*** (0.033)	0.136*** (0.032)	0.160*** (0.026)	0.214*** (0.035)	0.124*** (0.024)	0.140*** (0.027)	0.156*** (0.030)
Observations	470	223	247	204	266	129	341	267	203
Panel C: New Matters Sample									
2006 Billable Hours	0.101*** (0.028)	0.097** (0.049)	0.093*** (0.035)	0.073 (0.051)	0.117*** (0.033)	0.026 (0.055)	0.131*** (0.031)	0.110*** (0.034)	0.099* (0.050)
2006 New Matters	0.072** (0.030)	0.063** (0.028)	0.103*** (0.027)	0.060*** (0.022)	0.348*** (0.075)	0.107*** (0.023)	0.066** (0.031)	0.053** (0.022)	0.168*** (0.042)
Promotion Desire	0.169*** (0.026)	0.240*** (0.030)	0.073 (0.045)	0.166*** (0.043)	0.177*** (0.033)	0.232*** (0.042)	0.142*** (0.033)	0.153*** (0.036)	0.183*** (0.039)
Observations	288	125	163	123	165	84	204	169	119

Note: This table presents the results of testing for monotonicity in observables, which aims to show directional consistency of coefficients across subsamples. Each reported value comes from a separate regression of promotion on the specified row's variable, estimated using a different sample in each column. Data from the sample specified in each panel.

Table D5: Testing for Monotonicity in Observables (Dep. Variable: 2011 Productivity)

	All	Female	Male	Pre-72	Post-72	Top 20	Non-Top 20	Parent	Non-Parent
Panel A: Billable Hours Sample									
2006 Billable Hours	146.704*** (43.215)	249.627*** (42.716)	59.434 (52.499)	121.483** (57.664)	161.465*** (61.031)	102.939 (76.906)	167.721*** (53.011)	168.058*** (58.928)	114.597* (63.021)
2006 New Matters	-54.632*** (19.815)	-32.898*** (11.839)	-110.870** (48.340)	-38.782*** (13.920)	-323.785*** (80.139)	-53.408 (60.143)	-54.329*** (20.689)	-43.072*** (13.387)	-122.832*** (44.915)
Promotion Desire	36.972 (27.764)	60.753* (35.878)	-57.198 (35.531)	-48.299 (33.047)	93.288** (39.410)	45.141 (47.389)	34.275 (34.553)	39.357 (39.713)	35.443 (39.335)
Observations	269	119	150	115	154	80	189	157	112
Panel B: Total Hours Sample									
2006 Total Hours	5.063*** (0.657)	4.381*** (0.924)	4.773*** (0.929)	4.110*** (0.918)	5.801*** (0.914)	5.286*** (1.112)	4.883*** (0.833)	5.243*** (0.813)	4.911*** (1.118)
2006 New Matters	-0.318 (0.754)	0.207 (0.985)	-1.125 (0.728)	0.229 (0.556)	-1.282 (1.249)	-1.520* (0.835)	-0.070 (0.793)	0.013 (0.943)	-1.062 (0.949)
Promotion Desire	3.067*** (0.591)	2.651*** (0.835)	2.280*** (0.801)	3.100*** (0.875)	3.028*** (0.798)	2.626** (1.195)	3.326*** (0.664)	3.229*** (0.751)	3.008*** (0.924)
Observations	470	223	247	204	266	129	341	267	203
Panel C: New Matters Sample									
2006 Billable Hours	1.023 (2.003)	3.682 (3.918)	-0.639 (2.358)	3.130 (4.166)	-0.199 (2.090)	3.637** (1.702)	0.317 (2.677)	3.133 (2.286)	-1.445 (3.553)
2006 New Matters	12.939 (8.514)	4.883* (2.826)	35.803*** (11.965)	11.670 (8.431)	29.147*** (4.218)	34.119* (19.061)	8.771 (5.888)	11.185 (8.866)	22.257** (10.255)
Promotion Desire	1.458 (2.095)	3.531** (1.423)	-1.582 (4.924)	-0.790 (4.983)	2.582 (1.655)	-1.215 (5.175)	2.552 (1.821)	-1.289 (3.608)	4.176* (2.364)
Observations	288	125	163	123	165	84	204	169	119

Note: This table presents the results of testing for monotonicity in observables, which aims to show directional consistency of coefficients across subsamples. Each reported value comes from a separate regression of post-promotion productivity (2011 billable hours, total hours, or new matters, respectively) on the specified row's variable, estimated using a different sample in each column.

E Appendix: Theoretical Background

E.1 Prejudice and the Outcome Test

We adapt and simplify the framework from [Persico \(2009\)](#) to establish a theoretical model of how discrimination may affect promotion decisions within law firms. We refer to the decision-maker who potentially discriminates as the treator and the individual who is potentially discriminated against as the treated. In our setting, the treator is a law firm, while the treated are associates who may or may not be promoted to partner.

Following [Persico \(2009\)](#), we model associates as a mass of atomistic agents indexed by characteristics $g \in \mathcal{G}$. The firm may observe a rich set of characteristics for each associate, including race, gender, professional credentials, and measures of productivity. Accordingly, each g represents a particular combination of characteristics observed by the treator, and $\mathcal{G}$ denotes the set of all such possible combinations.

The econometrician is not required to observe the full set of characteristics observed by the treator. Rather, we assume they observe whether a treated agent's characteristics belong to a partition (with generic element G) of the set $\mathcal{G}$. To illustrate, suppose associates vary only by gender, F or M , and underlying productivity, H or L . The full characteristic space observed by firm management would then be

$$G = \{(F, H), (F, L), (M, H), (M, L)\}.$$

If the researcher observes gender but not underlying productivity, the observable partition consists instead of two elements:

$$G_1 = \{(F, H), (F, L)\} \quad \text{and} \quad G_2 = \{(M, H), (M, L)\}.$$

Thus, even though every promoted associate can be assigned to G_1 or G_2 , the researcher cannot observe their full set of characteristics. This formulation allows for omitted information available to the firm but unavailable to the econometrician.

At a minimum, we assume that the characteristic along which discrimination is being studied – e.g. gender – is observed by the researcher. Hence, in an analysis of gender discrimination, the econometrician’s partition is at least as fine $\{F, M\}$.

For each group g , the firm chooses an action $a_g \geq 0$, with the collection of actions satisfying $\{a_g\} \in \mathcal{A}$. Here, a_g represents the treatment intensity applied to group g among all possible actions $\mathcal{A}$ available to the treator. In our case, a_g represents the count of members of group g who are promoted, and the set $\mathcal{A}$ would capture the constraint that the total number of those promoted cannot exceed a fixed number of available partner positions.²⁴

Let $\pi(a_g, g)$ represent the firm’s expected monetary profit from group g when a_g members of the group are promoted. We assume that the firm’s choice has negligible effects on the environment faced by any particular associate and abstract from strategic interactions across firms. Under this simplification, $\pi(a_g, g) = \pi(g)a_g$ where $\pi(g)$ is expected profit from a person of group g .

We also assume the firm has the following feasible action set:²⁵

$$\mathcal{A} = \left\{ \{a_g\} \text{ s.t. } \int a_g dg \leq C \right\} \quad (4)$$

While we treat monetary profit as the firm’s legitimate objective, the firm’s actual objective may place different effective weights on the profits generated by different groups. We represent this potential distortion through a group-specific multiplicative term $\beta(g)$. In the discrimination literature, such a term is commonly interpreted as reflecting a ‘taste for discrimination.’

Our object of interest is therefore the function $\beta(g)$. Identification becomes more difficult with a coarse partition, as the groups observable to the econometrician may aggregate together types for which $\beta(g)$ differs. We follow the standard simplification that discrimination varies only with the characteristic under study. Specifically, we assume $\beta(g) = \beta(R) \forall g \subset R$; that is, conditional on the protected characteristic of interest, bias does not vary with the

²⁴Note that the set $\mathcal{A}$ can embed more than a binary decision; here, it embeds a constraint.

²⁵Unless otherwise noted, integrals without an explicitly stated domain are taken over $\mathcal{G}$.

remaining observed or unobserved characteristics.²⁶

Again, we observe partitions G of the set $\mathcal{G}$ but, due to missing variables, we do not observe each g . Importantly, we do not require those unobserved characteristics to be independent of the observed group characteristic, and our empirical approach does not require a cardinal or ordinal structure on the underlying g s or the observable groups G .

Although we have some information about the g s or G s for which $\{a_g\} = 0$, this information is not easily exploitable. Associates denied partnership may leave their firms, switch career paths, or otherwise alter their behavior following the denial. Thus, post-decision productivity may not provide a clean counterfactual measure of their value as partners. Our analysis therefore focuses primarily on associates who are promoted.

The law firm chooses the optimal promotion vector $\{a_g^*\}$ that solves

$$\begin{aligned} \max_{a(g)} \quad & \int \beta(g)\pi(g)a_g dg \\ \text{s.t.} \quad & \int a_g dg \leq C, \\ & 0 \leq a_g \leq I(g) \quad \forall g \end{aligned} \tag{5}$$

where $I(g)$ denotes the mass of available associates of type g . The second constraint therefore prevents the firm from promoting more associates of a given type than actually exist.

The central difficulty for conventional outcome tests is inframarginality. When there are groups which the treator would treat with maximum or minimum intensity, this lack of variation prevents us from identifying the bias. In our context, inframarginal groups are those in which the workers are sufficiently productive (or unproductive) to the extent that the firm finds it optimal to promote them all (or none). By contrast, a marginal group contains associates for whom the promotion constraint binds internally: some are promoted and some are not.

Formally, a marginal group is one for which the optimal treatment satisfies $0 < a_g^* < I(g)$.

²⁶Equivalently, the decision-maker's bias depends only on the characteristic being studied and is invariant to other dimensions of heterogeneity. This corresponds to the extended Roy model discussed by [Canay et al. \(2024\)](#).

An inframarginal group is one for which $a_g^* = I(g)$, or $a_g^* = 0$.

Within the partitions $G_1, \dots, G_n$, along characteristics that we think may cause bias, such as gender, consider the marginal subgroups $g_m^{G_1}, \dots, g_m^{G_n}$. The firm's optimal choice of actions $\{a_g^*\}$ is found by maximizing the Lagrangian

$$\begin{aligned} & \mathcal{L}(\{a_g\}, \lambda_0, \lambda_1(g)) \\ &= \int \beta(g)\pi(g)a_g dg - \lambda_0 \left[\int a_g dg - C \right] - \int \lambda_1(g)[a_g - I(g)]dg \end{aligned} \quad (6)$$

subject to $a_g > 0 \forall g$.

We differentiate with respect to each a_g to maximize, and because for marginal groups a_g^* is strictly positive and finite, we have $\beta(g)\pi(g) - \lambda_0 - \lambda_1(g) = 0$. Note also that $a_g^* < I(g)$ implies $\lambda_1(g) = 0$. Hence, we have

$$\beta(g_m)\pi(g_m) = \lambda_0 \quad (7)$$

for all g_m . Then,

$$\frac{\pi(g_m^{G_2})}{\pi(g_m^{G_1})} = \frac{\beta(g_m^{G_1})}{\beta(g_m^{G_2})} \quad (8)$$

Under our maintained assumption that bias depends only on the relevant group characteristic, $\beta(g) = \beta(G_i) \forall g \in G_i$. The preceding equation can therefore be written as

$$\frac{\pi(g_m^{G_2})}{\pi(g_m^{G_1})} = \frac{\beta(G_1)}{\beta(G_2)} \quad (9)$$

which allows us to read off the ratio in the bias between the two groups directly from the ratio in profitabilities of the marginal groups.²⁷ Put differently, if one group must generate

²⁷Up until now we have been referring to profitability $\pi(g)$ as a function of only g . This however may be unrealistic to assume, as the treated groups may adjust their behavior in response to the aggregate action $\mathbf{a}_g$ of a mass of firms. For example, if associates with some characteristic g_1 expect virtually certain promotion, their incentives to maintain high profitability could weaken. Therefore, their profitability $\pi(g_1)$ may decrease as a result of the aggregate action. Conversely, members of a group g_2 who rarely get promoted might attempt to increase their level of profitability. Profitability may thus be better represented as $\pi(g, \mathbf{a}_g)$, where $\mathbf{a}_g$ denotes aggregate treatment of group g .

In the general equilibrium model then, $\pi(g, \mathbf{a}_g)$ is determined as part of an equilibrium. The sharpness of the response of group g to variations in $\pi(g, \mathbf{a}_g)$ determines which groups are treated in equilibrium. The

systematically greater post-promotion value at the promotion margin, this indicates that the group faced a more demanding effective promotion threshold.

Computing the ratio described above requires identifying which promoted associates were actually close to the firm’s promotion margin. The empirical challenge is therefore to construct a sample of promoted associates whose promotion decisions most closely approximate marginal decisions.

E.2 Identifying Marginal Partners

The outcome test does not require the researcher to observe the magnitude of the underlying bias directly. It does, however, require information about individuals near the relevant decision margin.

We do not directly observe the profitability for the marginal groups because marginal status itself is latent. Rather, within an observed partition G , we observe an aggregate of profitabilities $\int_G R_G(g) a_g dg$ over a broader set of groups which may include the marginal group g_m . The confounding groups $g \neq g_m$ are inframarginal.

The firm’s latent promotion threshold, which would identify the associates who only narrowly qualified for partnership, is generally unavailable to the econometrician. This is important because the misspecification of marginal individuals may induce bias in the outcome test: If the underlying productivity distributions differ across groups, average productivity among associates well above the cutoff may differ even when the relevant promotion thresholds do not. On the other hand, differences at the margin can be obscured by the composition of inframarginal promotees.²⁸

To address this problem, we implement [Grau and Vergara \(2025\)](#)’s observational approach, which seeks to limit the scope of inframarginality bias. The central idea is to use

logic of the framework nevertheless continues to apply: comparisons at the margin still reveal differences in effective thresholds. Allowing productivity to respond to aggregate treatment mainly provides a mechanism through which positive masses of several different groups may be promoted in equilibrium.

²⁸[Ayres \(2002\)](#) notes that inframarginality concerns may be less severe when the decision or the outcome is non-dichotomous, as it may become easier for the econometrician to identify whether the decision-making threshold is discriminatory. Our outcomes of interest are indeed non-dichotomous, and we make use of this feature by analyzing the distribution of billable hours (and number of new matters brought to the firm, and total weekly hours worked) and not just the mean.

observable information to identify, among those who were promoted, the associates most likely to have been close to the firm’s promotion cutoff. Outcomes calculated within this subset should more closely approximate outcomes at the true margin and therefore contain less inframarginality contamination than outcomes averaged across all promoted partners. This approach does not require randomized assignment or the use of instruments, though it comes at the cost of certain assumptions, which we further detail in subsequent sections.

Our application assumes that associates’ observable pre-promotion characteristics contain information about the ranking that senior partners use when assessing candidates for promotion. The maintained requirement is that the decision-maker can be represented as ranking individuals according to perceived suitability for treatment, as in [Hull \(2021\)](#).²⁹

For an individual associate, the selection rule into promotion can be formalized by:

$$R_i = \mathbb{1}\{\hat{\pi}(G_i, Z_i) \geq h(G_i, Z_i)\}, \quad (10)$$

where i indexes associates. The indicator variable R_i takes the value 1 if associate i is promoted to partner. The variable G_i is a binary group indicator (e.g., gender), and Z_i is a vector of characteristics of associate i observed by the senior partners making promotion decisions (e.g., productivity as an associate). The function $\hat{\pi}(G_i, Z_i) = \mathbb{E}[\pi_i | G_i, Z_i]$ represents the expected profit contribution if promoted. Here, π_i denotes the actual profit generated by the newly promoted partner. The function $h(G_i, Z_i)$ represents the *effective threshold*, which may vary with G_i and Z_i due to firm policies, estimated future performance, or biases.

Let the latent promotion status be defined as $R_i^* = \hat{\pi}(G_i, Z_i) - h(G_i, Z_i)$, so that the observed promotion decision follows $R_i = \mathbb{1}\{R_i^* \geq 0\}$. We say that a promoted associate (to partner) is *marginal* if $R_i^* = 0$.

[Hull \(2021\)](#) establishes that the observed average behavior of marginal individuals of a given group corresponds to that group’s average effective threshold. Conditioning on $R_i^* = 0$ is therefore valuable for two reasons. First, it is the object that can be identified by the

²⁹Our concept of prejudice combines preferences and stereotypes. The framework does not separately identify these two channels.

outcome test without further assumptions. Second, discriminatory distortions have direct treatment consequences precisely near this margin, because small changes in the decision rule can change whether these associates are promoted.

Under a disparate impact notion of prejudice (Arnold et al., 2022), our approach remains relevant because disparities in promotion rates may emerge for reasons distinct from partners’ subsequent productivity.³⁰ The approach introduced in the next section allows us to test for patterns of prejudice that vary jointly with G_i and Z_i . This flexibility is useful both for assessing the exclusion restrictions emphasized by Canay et al. (2024) and for examining whether senior partners’ decisions respond to predicted productivity in the manner required by the framework.

E.3 The Prediction-Based Outcome Test

In this section, we adapt the framework from Persico (2009) to describe and formalize Grau and Vergara (2025)’s Prediction-Based Outcome Test (P-BOT) in the context of promotion in law firms. We discuss identification and estimation, as well as the scope and limits of the method.

We classify all variables that affect the promotion decision into variables that are observed by the econometrician, X_i , and variables that are not, V_i . With this notation we can write $R_i^* = f(X_i, V_i)$ for some function f so $R_i = \mathbb{1}\{f(X_i, V_i) \geq 0\}$. The only variable we require to belong to X_i is G_i , since the identification of marginal individuals is, ultimately, an input for testing for prejudice against G_i .

E.3.1 Intuition

An outcome test ideally compares associates who were just at the promotion cutoff. Because true marginality is latent, however, the researcher cannot simply select these associates from the data. A naive alternative is to compare the average post-promotion productivity of all promoted associates across groups. However, if productivity distributions vary by group,

³⁰Under disparate treatment notions of prejudice, however, this distinction may compromise the validity of the outcome test (Canay et al., 2024).

differences in averages may be uninformative of the productivity arising at the margin.

The basic strategy of the P-BOT is therefore to identify subsets of promoted associates whose observable characteristics make them especially likely to have been near the cutoff. Restricting attention to such observations should make the resulting outcome comparison more representative of the true marginal comparison and less vulnerable to inframarginality bias.

Formally, for each $G_i \in \{0, 1\}$, the objects of interest to compute the outcome test are:

$$\mathbb{E}[\pi_i | G_i = g, R_i^* = 0] \quad (11)$$

Since R_i^* is unobserved, the econometrician must instead infer which promoted associates are likely to satisfy $R_i^* \approx 0$, or, equivalently, which promoted associates are likely to have post-promotion outcomes similar to those of the truly marginal individuals.

The empirical problem can therefore be stated as follows: using observational data, how can we approximate the productivity of partners for whom $R_i^* = 0$? We begin with the following continuity assumption.

ASSUMPTION 0 (A0): The profit expectation given G_i and R_i^* is continuous in R_i^* at $R_i^* = 0$:

$$\lim_{\varepsilon \rightarrow 0} \mathbb{E}[\pi_i | G_i = g, R_i^* \in [0, \varepsilon]] = \mathbb{E}[\pi_i | G_i = g, R_i^* = 0]$$

A0 requires outcome, π_i , to vary smoothly as associates approach the latent promotion threshold, R_i^* , from above. This forms the intuitive basis for our identification argument: Outcomes for associates at the margin can be approximated by outcomes for associates close to the margin. In simple terms, when ε is small, then

$$\mathbb{E}[\pi_i | G_i = g, R_i^* \in [0, \varepsilon]]$$

constitutes a good approximation of (11). That is, the productivity of marginally promoted associates can be approximated by the productivity of associates with $R_i^* \in [0, \varepsilon]$.

The intuition above is not directly applicable when the econometrician does not observe R_i^* because the promoted associates with $R_i^* \in [0, \varepsilon]$ cannot be directly identified in the data. Instead, we use the following observational object:

$$\mathbb{E}[\pi_i \mid G_i = g, R_i = 1, q(p(X_i)) \leq \tau] \quad (12)$$

to estimate $\mathbb{E}[\pi_i \mid G_i = g, R_i^* \in [0, \varepsilon]] = \mathbb{E}[\pi_i \mid G_i = g, R_i = 1, R_i^* \leq \varepsilon]$, where $p(X_i) = \mathbb{E}[R_i \mid X_i]$ is the propensity score, q is the quantile function (conditional on promotion), and τ is a sufficiently small quantile. Equation (12) is therefore the mean outcome among promoted associates with $G_i = g$ and low propensity scores conditional on promotion.

The intuition for using (12) to estimate (11) is the following: If a promoted associate's propensity score is large, it seems unlikely that the associate was close to the promotion threshold (and therefore, to being marginal). In contrast, if the propensity score of a promoted associate is small, it is more likely that the associate was closer to being denied promotion. This intuition suggests that promoted associates with lower propensity scores, say, $p(X_i) \leq a$, are more likely to have $R_i^* \leq \varepsilon$, for some $a(\varepsilon)$. However, implicit in this intuition are restrictions on the conditional distribution of V_i . Intuitively, if conditional on X_i , the variance of V_i is large, then individuals with large $p(X_i)$ may have draws of V_i such that the true R_i^* is small, and vice versa.

Following this intuition, the analysis below formalizes restrictions on the model and the data that can make (12) work as a reasonable approximation for (11). We call this estimator a Prediction-Based Outcome Test (P-BOT), after [Grau and Vergara \(2025\)](#), because the sample selection procedure relies on the predicted promotion status (i.e., the propensity score). An attractive feature of this approach is that it reframes the central empirical challenge of the outcome test as one of estimating $p(X_i)$.

E.3.2 Assumptions

Following [Grau and Vergara \(2025\)](#), we first make a separability assumption.

ASSUMPTION 1 (A1). There exist functions d and g such that:

$$\mathbb{1}\{f(X_i, V_i) \geq 0\} = \mathbb{1}\{d(X_i) - g(V_i) \geq 0\} \equiv \mathbb{1}\{d(X_i) - W_i \geq 0\}. \quad (13)$$

A1 requires the selection rule to admit an additively separable representation. As discussed by [Vytlačil \(2002\)](#), this imposes an additional monotonicity restriction. One suggestive assessment of A1 is to regress R_i on X_i across subsamples of associates thought to differ in V_i and examine whether the signs of the estimated relationships remain stable. We display results from this diagnostic in [Appendix D](#).³¹

Our second assumption restricts the conditional distribution of W_i given X_i .

ASSUMPTION 2 (A2). The structure of W_i given X_i is given by:

$$W_i = \mu(X_i) + \sigma\zeta_i, \quad (14)$$

where $\sigma > 0$ and ζ_i is a scalar random variable independent from X_i with $\mathbb{E}[\zeta_i] = \bar{\zeta} < \infty$ and strictly increasing CDF.

A2 permits arbitrary correlation between X_i and W_i through the unrestricted $\mu(X_i)$. Its substantive restriction is instead on the conditional variance of W_i , which is assumed to be homoskedastic. While restrictive, this provides a clear version of the identification argument.³²

³¹As [Grau and Vergara \(2025\)](#) note, some structure in the promotion decision leads to $f(X_i, V_i) = h(X_i, V_i) - p(X_i, V_i)$, giving (i) $h(X_i, V_i) = h_X(X_i) + h_V(V_i)$ and (ii) $p(X_i, V_i) = p_X(X_i) + p_V(V_i)$ as sufficient conditions for A1. The former is not directly testable, but the latter implies monotonicity in observables in the expected productivity equation. A related diagnostic is therefore to regress post-promotion productivity on various observables X_i in samples of promoted associates thought to differ V_i and inspect stability of signs. We also detail this diagnostic in [Appendix D](#). Notably, post-promotion productivity may be contaminated by selection in unobservables.

³²It is important to note that A1 and A2 do not imply the absence of inframarginality bias. To the extent that the distribution of (X_i, V_i) varies with G_i , it is still the case that both groups may have different productivity distributions.

E.3.3 Identification

The following proposition summarizes the identification argument. Refer to [Grau and Vergara \(2025\)](#) for a formal proof.

PROPOSITION I. *Assume A1 and A2 hold, and let $\varepsilon > 0$ be fixed. Then:*

$$\mathbb{E}[\pi_i \mid G_i = g, R_i = 1, p(X_i) \leq a(\varepsilon)] = \mathbb{E}[\pi_i \mid G_i = g, R_i = 1, R_i^* \leq \varepsilon + u_i], \quad (15)$$

where $p(X_i) = \mathbb{E}[R_i \mid X_i]$, $a(\varepsilon)$ is constant given ε , with $a'(\varepsilon) > 0$, and

$$u_i = -\sigma(\zeta_i - \zeta).$$

The proposition states that selecting promoted associates with sufficiently low propensity scores amounts to selecting individuals who are close to the latent promotion cutoff, subject to a mean-zero classification error. Since $a'(\varepsilon) > 0$, tightening the propensity-score threshold corresponds to narrowing the implied neighborhood around the true promotion margin. Combined with the continuity condition in A0, this result provides the basis for using outcomes among low-propensity promotees to approximate outcomes among genuinely marginal partners, with the quality of the approximation depending on the behavior of u_i .

In our setting, we therefore isolate promoted lawyers who appear least likely to have received promotion based on their observable pre-promotion characteristics. Under the null of equal promotion thresholds across groups, these marginal partners should exhibit comparable post-promotion productivity across the protected characteristic being studied. If, for example, women face a more demanding promotion threshold, then women observed near the promotion margin should on average exhibit stronger subsequent performance than similarly marginal men. Several features of the result are worth emphasizing.

Prediction. Identification depends on the quality of the predicted promotion ranking rather than on any particular functional form used to estimate it. This gives the procedure an unusual form of robustness to omitted-variable bias in the prediction equation. While unobserved V_i may bias the estimated coefficients of the prediction model, the same bias may

improve propensity score fit. The researcher need not correctly identify why an associate was near the promotion cutoff so long as the observable prediction rule contain information about the associate's relative proximity to that cutoff.

Inframarginality bias. Proposition I does not imply that the proposed estimator is entirely free from inframarginality bias. While $\mathbb{E}[u_i] = 0$ and, therefore, the prediction error is not systematic, nonlinearity in the relationship between outcomes and R_i^* means that mean-zero classification error can still affect the estimand.

As an illustration, consider two associates with realizations $u_1 > 0$ and $u_2 < 0$ such that $u_1 + u_2 = 0$. Suppose $R_1^* > \varepsilon$ but $R_1^* \leq \varepsilon + u_1$. Likewise, assume that $R_2^* \leq \varepsilon$ but $R_2^* > \varepsilon + u_2$. While the first associate does not fall within ε of the margin, they are included in the marginal sample, whereas the second associate does fall within ε but is excluded. If the post-promotion outcome distributions of these locally misclassified associates differ, these rank reversals may induce inframarginality bias.

As such, the validity of (15) as an approximation for (12) depends on the magnitude of the misclassification bias, which is determined by the conditional variance of W_i given X_i ; under the homoskedasticity restriction, this equals $\mathbb{V}(u_i) = \sigma^2 \mathbb{V}(z_i)$. When this variance is small, relatively few associates are misclassified, and they tend to lie close to one another in latent promotion status and pre-promotion decision productivity, limiting the scope for inframarginality bias. However, when this variance is large, associates may be substantially reordered by different realizations of the unobserved component; thus, the scope for inframarginality bias may be substantial.³³

This implies bias is decreasing in the fit of the propensity score. When X_i has poor predictive power, estimated promotion propensities become relatively flat, and selecting associates with the lowest propensities becomes increasingly similar to drawing a random sample of all promotees. Thus, the predictive performance of the promotion model provides an indirect diagnostic for the scope of inframarginality bias.

The role of homoskedasticity. Proposition I is derived under the assumption of ho-

³³Grau and Vergara (2025) present Monte Carlo simulations that show that as the conditional variance increases, that test converges to the average outcome test.

masked conditional variance. If we generalize A2 such that the scalar term becomes an unrestricted function $\sigma(X_i)$, the classification error can remain mean-zero but vary with X_i . As shown by [Grau and Vergara \(2022\)](#), when $\sigma(X_i)$ is unrestricted, $p(X_i) \leq a(\epsilon)$ implies $R_i \leq \epsilon + \Gamma(X_i, u_i)$, with $E[\Gamma(X_i, u_i)] = 0$. The precision with which the propensity score ranks observations around the margin then depends on the particular X_i associated with each observation: some regions of the covariate space may have relatively precise latent rankings, while others may contain substantially more residual variation. The basic interpretation nevertheless remains unchanged. Inframarginality arises from local mistakes in classifying who is close to the promotion threshold, and the fit of the promotion model remains informative about the likely extent of these errors.

Empirical test for inframarginality bias: the perturbation test. The discussion above suggests that, even with heteroskedastic $\sigma(X_i)$, the scope for inframarginality bias remains proportional to uncertainty in the predicted ranking, allowing us to deliberately perturb that ranking and examine whether the substantive conclusion of the outcome test is sensitive to such perturbations. As [Grau and Vergara \(2025\)](#) demonstrate, this can be formalized via a perturbation test. Under A1 and A2, the selection rule is $R_i = \mathbb{1} \left\{ \frac{d(X_i) - \mu(X_i)}{\sigma(X_i)} \geq \zeta_i \right\}$, where the left-hand side and corresponding variance of ζ_i can be estimated. We can then use the estimated variance of ζ_i to simulate perturbations that alter the estimated propensity score and, therefore, the group of marginal associates. Re-estimating the outcome test after these perturbations provides a direct way to assess whether plausible local rank reversals materially change our estimates.

E.3.4 Estimation and implementation

Equation (15) suggests a direct empirical procedure. First, estimate each associate’s propensity for promotion. Second, use those predicted probabilities to rank associates who actually become partners. Third, define a subset near the promotion margin and estimate the group difference in post-promotion outcomes within that subset. Let $\hat{p}(X_i)$ denote the estimated propensity score.

Simple approach. This implementation defines a sample of marginal partners based on the percentile of their predicted promotion propensity. A promoted associate enters the estimation sample when $q(\hat{p}(X_i)) \leq \tau$, where τ is *small* and $a = q^{-1}(\tau)$. Within this truncated sample, the outcome test can be implemented by regressing π_i on G_i . When $G_i = 1$ denotes the potentially disadvantaged group and larger values of π_i represent better partner performance, the sign of the estimated group coefficient indicates how post-promotion productivity differs between the groups near the estimated promotion margin. Notably, the choice of τ involves a conventional bias-variance tradeoff: a large value increases sample size and thus precision, but it also involves a larger share of inframarginal individuals, which may induce bias. At the extreme, when $\tau = 100$, the P-BOT becomes the average outcome test. The econometrician can therefore assess the scope for inframarginality by evaluating the extent to which the choice of τ impacts P-BOT estimates.

Non-parametric approach. We additionally perform non-parametric local regressions to better approximate the limit outlined in A0. Formally, the econometrician can estimate $\mathbb{E}[\pi_i \mid G_i = 0, q(\hat{p}(X_i)) = 1]$ and $\mathbb{E}[\pi_i \mid G_i = 1, q(\hat{p}(X_i)) = 1]$, and assess the extent of prejudice by taking the difference. This approach applies a triangular kernel and thus has the advantage of more heavily weighting observations closer to the margin.

Choice of X_i . The usefulness of the P-BOT depends critically on how much variation in the promotion decision can be predicted by the available covariates. The researcher would therefore like X_i to incorporate rich pre-promotion information whenever doing so improves predictive performance. Flexible machine-learning methods can be useful in this setting because they can accommodate a large predictor set while regularizing variables that contribute noise without improving out-of-sample prediction. Richer X_i also tends to produce greater dispersion in $p(X_i)$, making the ranking of promoted associates by predicted promotion probability more informative about their relative proximity to the cutoff.

Inference. Standard errors must account for the fact that the estimation sample is itself constructed using an estimated object, $\hat{p}(X_i)$. We therefore use bootstrap procedures to construct confidence intervals for the outcome-test estimates.

Perturbation test. The *perturbation test*, developed by [Grau and Vergara \(2025\)](#), begins with estimating a probit model for promotion status. For each promoted associate, we simulate M realizations from a standard normal distributions, and for each of the M simulated realizations, we construct a perturbed latent R_i^* for promoted associates, redefine the subset classified as being close to the promotion margin, and re-estimate the group-specific post-promotion outcome comparison. The distribution of the resulting P-BOT estimates shows how sensitive the estimated discrimination pattern is to plausible errors in the ranking used to identify marginal partners.

E.3.5 Discussion

The prediction-based outcome test has several useful features for our setting. First, because the method requires an informative ranking rather than a structural interpretation of the promotion model’s coefficients, conventional omitted-variable bias in the initial prediction equation is not by itself fatal to the approach. Second, the method can be implemented without an instrument or quasi-random assignment mechanism and can accommodate discrimination that varies non-monotonically across observable characteristics. Third, the potential importance of residual inframarginality can be investigated empirically through prediction fit, sensitivity to the choice of the marginal sample, and perturbation exercises. Finally, implementation is relatively straightforward once a credible promotion-propensity model has been estimated.

The approach also has important limitations. Its identification argument depends on restrictions such as A1 and A2, whose plausibility may vary across institutional settings. Moreover, the P-BOT can only sharply isolate the promotion margin when the available pre-promotion variables predict promotion reasonably well. If promotion decisions depend primarily on information that is unavailable to the econometrician, estimated propensity scores will contain little information about relative distance to the true cutoff, and the P-BOT will increasingly resemble the conventional average outcome test.