Where Gender Gaps Arise: Firm Complaint Handling, Arbitration, and Customer Recovery

Ishitha Kumar* Ivilina Popova†

July 2026

Abstract

Customer complaints against advisors can be resolved by firms or in FINRA arbitration. Using 99,674 complaints with reported damages, we show that customers receive less compensation when their complaint names a female advisor in firm-handled complaints, rather than formal arbitration. There, claimants are 2.4 percentage points less likely to receive payment and, when paid, receive 10.3 percent less than customers with otherwise similar male-advisor complaints. Together, these margins imply 15 percent lower expected compensation. Complaints involving female advisors enter formal arbitration at similar rates. Yet the shortfall is concentrated in firm-handled complaints and narrows by more than half at firms with high complaint rates. In a selectively matched sample of published awards, panel features do not account for the pattern. These patterns point to the greater discretion firms have when handling complaints, compared with formal arbitration, as the key difference between the two resolution processes that can explain the compensation gap.

Keywords: financial advisors; FINRA arbitration; customer complaints; dispute resolution; gender; claimant recovery.
JEL codes: G24; G28; J16; J71; K41.
Acknowledgments and Disclosures: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The author has no relevant or material financial relationships or other potential conflicts of interest related to this research. The author has read the NBER disclosure policy and attests that this statement discloses all sources of research support and all material and relevant financial relationships.

*Department of Finance and Economics, Texas State University. Email: ishitha@txstate.edu.
†Department of Finance and Economics, Texas State University. Email: ip12@txstate.edu.

I Introduction

When a customer alleges that a financial advisor caused them harm, the dispute can be resolved in either of two institutionally distinct ways.[1] The employing firm can investigate and close the complaint itself, deciding whether to compensate the customer and, if so, how much. Or the dispute can enter FINRA arbitration, an externally administered forum with standardized procedures and an adversarial process. The two channels resolve the same kind of dispute, but they differ in how much discretion the resolver has in evaluating and settling the complaint.

In this study, we examine whether customers receive different compensation when their complaint names a female rather than a male advisor in these two channels. If a female–male gap in claimant recovery reflects differences in the underlying disputes, it should appear regardless of where those disputes are resolved. If it reflects discretion in evaluation, the gap should be larger where firms have more discretion and smaller where formal procedures constrain it. The two channels therefore give us different predictions for the same population of complaints.

We find that customers receive less compensation when their complaint names a female advisor, but the difference is concentrated in non-arbitrated complaints and narrows sharply at firms with higher complaint rates. Formal arbitration does not show the same negative pattern. The results are consistent with differences in how firms exercise discretion when handling complaints, while formal arbitration limits that discretion through standardized procedures. The compensation gap therefore appears to arise in how complaints are handled rather than from the formal arbitration process itself. Importantly, this pattern is specific to customer compensation and does not extend to advisors' subsequent careers.

[1] See FINRA, "File a Complaint" (accessed July 17, 2026).

To study these differences, our data combine regulatory disclosure records with advisor registration and firm-affiliation histories. They cover 991,675 advisors and 121,791 customer complaints filed from 2010 to 2023. Each complaint identifies the advisor, firm, state, allegation type, claimed damages when reported, resolution amount, and whether the dispute entered FINRA arbitration. Our main analysis uses 99,674 complaints with reported damages: 74,905 non-arbitrated complaints and 23,814 formal-arbitration complaints. To study what happens inside arbitration, we also hand-match published FINRA award PDFs to disclosure records. This produces 2,793 advisor-level records linked to 1,556 award cases and provides information on panel composition, counsel, claim revisions, and hearing details.

Our empirical approach proceeds in three steps. We first compare customer compensation across non-arbitrated and formal-arbitration complaints, controlling for claim size, allegation type, advisor characteristics, firm-year effects, and state-year effects. We then examine whether the compensation gap varies with the firm environment, particularly complaint rates and female-advisor representation. Finally, we use the matched FINRA awards to study whether observable features of formal arbitration account for the observed results. Because firms and claimants determine whether a dispute enters arbitration, our channel comparisons are conditional on the observed resolution channel. And because many complaints receive no payment, we examine both whether customers receive any compensation and how much they receive when paid.

We start our analysis by estimating customer compensation separately in non-arbitrated and formal-arbitration complaints. Holding claim size, allegation type, advisor characteristics, firms, states, and years fixed, customers whose complaints name female advisors are 2.4 percentage points less likely to receive payment in non-arbitrated complaints. When they are paid, they receive 10.3

percent less. Together, these differences imply about 15 percent lower expected compensation. To illustrate the magnitude, 44.6 percent of non-arbitrated complaints receive a positive payment, and the median payment among those complaints is $41,000. Applying the estimated differences on the two margins gives illustrative expected-compensation benchmarks of roughly $18,300 for complaints naming male advisors and $15,500 for those naming female advisors—a difference of about $2,800 per complaint. We do not find a similar negative pattern in formal arbitration, although complaints involving female advisors enter arbitration at similar rates to those involving male advisors.

We then examine whether the compensation gap varies across firms. If differences in complaint handling matter, the gap should vary with the environment in which firms evaluate complaints rather than with the characteristics of individual advisors. We find that the gap narrows by more than half at firms with higher complaint rates. The same pattern appears at firms employing more women, while the gap does not vary with advisor experience. These results show that the compensation gap tracks the firm environment rather than the individual advisor.

The discretion explanation also makes a prediction within arbitration. Arbitration is not uniformly procedural: judgment has more room to operate in contested cases than in settlements or withdrawals. We therefore examine the matched sample of published FINRA awards, using information on panel composition, claimant counsel, claim revisions, and hearing details. Panel size, arbitrator composition, and FINRA's shift toward all-public panels do not produce detectable differences for complaints involving female advisors. The differences that do appear are on the claimant side: experienced claimant counsel are associated with smaller recovery gains in cases involving female advisors, and claims against female advisors are revised down more before judg-

ment. Because these estimates come from a small, selectively matched sample and do not survive multiple-testing correction, we treat them cautiously.

Finally, we examine whether the compensation gap is accompanied by different subsequent career outcomes for female advisors. Female advisors are less likely to be terminated in both channels and are modestly more likely to re-enter employment after non-arbitrated complaints. In formal arbitration, their employment outcomes are statistically similar to those of male advisors. The compensation gap therefore does not extend to advisors' subsequent careers.

Contribution. This paper contributes to three literatures. The first studies financial-advisor misconduct and the institutions governing the market for advice (Egan et al., 2019; Charoenwong et al., 2019; Gelman et al., 2021; Gurun et al., 2018; Bhattacharya et al., 2012; Stoughton et al., 2011; Guiso et al., 2008; Dimmock and Gerken, 2012). That work explains why misconduct occurs and persists. We show that how complaints are resolved also matters: among observably similar complaints, customer compensation differs depending on whether the dispute is handled by the firm or enters formal arbitration.

The second is the literature on gender and institutional evaluation (Goldin and Rouse, 2000; Goldin, 2014; Blau and Kahn, 2017; Bertrand et al., 2010), and its recent application to advice markets (Egan et al., 2022; Honda, 2025; Clifford et al., 2023; Sarsons, 2019). Egan et al. (2022) document harsher post-misconduct labor-market consequences for female advisors; Honda (2025) shows that those consequences are sensitive to how disclosure events are classified and largely disappear once event types are separated. We contribute to this literature by showing where the difference in customer compensation arises. It appears in firm-handled complaints rather than in formal arbitration and does not extend to advisors' subsequent careers.

The third is the literature on arbitration, repeat players, and discretion in institutional decision-making (Galanter, 1974; Egan et al., 2018; Arnold et al., 2018; Rehavi and Starr, 2014). Much of this literature focuses on whether the arbitration forum itself disadvantages one side. Our results point instead to what happens before formal arbitration, when firms handle complaints internally. This shifts attention toward firm-level complaint handling, where disclosure records report the outcome but not how firms evaluate or settle the complaint.

The rest of the paper proceeds as follows. Section II describes the institutional setting, develops the discretion explanation and its predictions, and presents the data and design. Section III establishes the by-channel pattern. Section IV shows how the internal gap varies with the firm's complaint-processing exposure. Section V opens the arbitration black box using published awards. Section VI reports robustness and scope conditions, and Section VII concludes.

II Data and Empirical Design

A Institutional Setting and Dispute-Resolution Channels

Financial advisors operate inside a disclosure regime that records customer complaints, regulatory actions, terminations, and other reportable events. Throughout, we use *advisor* for the individual and reserve *adviser* for regulatory terms such as "investment adviser" in Form ADV and the Investment Adviser Registration Depository.[2] Customer disputes involving advisors can be resolved through two institutionally distinct channels.

The first channel is handled within the employing firm. A customer alleges that the advisor's

[2]The distinction is purely terminological: common industry usage favors *advisor*, whereas SEC forms and related statutes use *adviser*.

conduct caused financial harm, and the complaint is recorded in the advisor's disclosure history if it satisfies reporting requirements. The firm may deny the claim, close it without payment, or resolve it monetarily. We refer to these cases as non-arbitrated complaints. The firm possesses information about the advisor and the underlying transaction and has substantial latitude in how it investigates and settles the complaint.

The second channel is formal FINRA arbitration, the dominant forum for customer disputes involving FINRA member firms. Arbitration is more formal and more adversarial than within-firm resolution. Parties submit claims and responses, exchange information, and, when the case is not settled or withdrawn, present evidence before one or more arbitrators. FINRA Rule 12403 governs the composition of three-arbitrator panels and, after the 2011 reform, permits customer claimants in eligible cases to elect an all-public panel.[3] The forum is therefore externally administered even though it remains industry-specific.

B Discretion and the Location of Disparities

The two channels differ along many dimensions, but the key difference for this paper is how much discretion the resolver has in evaluating and settling the complaint. Non-arbitrated resolution is bilateral and unstandardized. Disclosure records report the outcome, but not how the firm reached it. The firm holds private information about the advisor and the transaction and chooses whether to deny, close, or pay. Formal arbitration limits some of that discretion through procedure. Claims and responses are filed, discovery is structured, and panels are drawn under rules that constrain composition. When cases are contested, panels issue written awards. Arbitration does not eliminate

[3] See FINRA Rule 12403 (accessed July 17, 2026).

judgment, but formal procedures place greater constraints on how it is exercised.

This difference gives the two explanations distinct predictions. Suppose first that any female–male difference in claimant recovery reflects the underlying merits of the disputes—that complaints against female advisors are, conditional on observables, genuinely weaker. Then the difference should appear in both channels at similar magnitudes. It should also be visible at the entry stage if parties assess those differences when deciding whether a complaint enters arbitration. A merits explanation also predicts milder downstream consequences for female advisors if their complaints involve less serious conduct. The discretion explanation can produce the same career pattern if firms that pay less also impose milder employment consequences. Downstream employment outcomes therefore cannot distinguish the two explanations. We report them in Section C to establish the scope of the compensation result and to connect the analysis to the literature on post-misconduct labor markets.

Suppose instead that the difference arises from discretion in evaluation. Then the gap should be larger in non-arbitrated complaints and smaller where formal procedures constrain judgment. Arbitration entry should not differ systematically by advisor gender once claim size and case characteristics are held fixed. The gap should also vary with the firm environment rather than with the advisor's own attributes, narrowing where firms have greater complaint-processing exposure and remaining unrelated to advisor experience. Within arbitration, any remaining difference should appear in contested cases and claimant-side litigation rather than in observable panel features such as panel size or arbitrator composition.

Figure 1 maps the complaint-to-resolution sequence and shows where each test enters the analysis. The two explanations are not exhaustive or mutually exclusive, but they make different

predictions that the data can distinguish. Section III examines arbitration entry, the channel comparison, and downstream careers. Section IV asks whether the gap varies with the firm environment or advisor attributes. Section V examines variation within formal arbitration. Throughout, we use *discretion* to mean latitude over how a dispute is evaluated and settled, not a claim about any party's intentions.

C Data

C.1 Complaint Disclosures and Advisor Records

Our primary data come from regulatory disclosure records for registered financial advisors between 2010 and 2023. We obtain broker-side disclosures from FINRA BrokerCheck, which draws from Forms U4 and U5 filed through the Central Registration Depository, and supplement them with adviser-side records from the Investment Adviser Registration Depository and related Form ADV filings. These sources provide advisor registration histories, firm affiliations, customer disputes, regulatory actions, terminations, and subsequent employment paths.

The combined advisor universe contains 991,675 unique individuals. From it we construct a customer-dispute sample of 121,791 complaints filed from 2010 through 2023. Each observation contains the complaint year, advisor experience, registration status, firm affiliation, state, allegation category, claimed damages when reported, and the resolution amount recorded in the disclosure.

Each observation is one reported customer complaint as it appears on an advisor's disclosure record. If a complaint later proceeds to arbitration, it still remains a single observation in our file; we classify that row by its recorded resolution channel rather than duplicating it as separate firm-resolution and arbitration observations. We therefore do not reconstruct a second firm-level

claimant-recovery outcome for a complaint that later appears in arbitration. Instead, we work with the reported complaint universe and partition it by resolution channel. Specifically, the pooled complaint sample contains one row per dispute with reported damages. We then split that sample into complaints that never appear as formal arbitration disclosures and complaints that do. We therefore study the same complaint universe in three samples: all complaints, non-arbitrated complaints, and formal-arbitration complaints.

We classify a dispute as *formal arbitration* whenever the disclosure record identifies it as filed in or resolved through FINRA arbitration. Under this definition, arbitration includes disputes that settle, are withdrawn, or are decided by a panel. This coding explains why the arbitration share is larger than a reader might expect if one had only published awards in mind: the formal-arbitration channel is broader than the subset of disputes that culminate in a written award.

The recovery analysis requires claimed damages, so the estimation samples are smaller than the raw complaint counts. The final pooled complaint sample contains 99,674 observations with reported damages. The channel-specific claimant-recovery samples contain 74,905 non-arbitrated complaints and 23,814 formal-arbitration complaints after applying the same controls and fixed effects used in the baseline regressions. Female advisors account for 9.3% of the pooled estimation sample and 8.0% of the formal-arbitration estimation sample. Because the excluded raw-file rows coincide with both missing claimed damages and missing channel coding, the relevant restriction is entry into the fully coded claimant-recovery sample rather than clean selection on damages reporting alone; Appendix Table Q1 audits that sample-entry stage directly.

We classify advisor gender from first names using the GenderChecker database (GenderChecker, 2020). First-name-based gender coding is standard in this literature; see, for example, Egan et al.

(2022) and Honda (2025). GenderChecker returns male, female, and unisex labels, and no advisor in our data is left unmatched. Of the 991,675 unique advisors, 60.5% are classified as male, 21.5% as female, and 18.0% as unisex. Our main female indicator equals one only for names classified as female; male and unisex names form the omitted category.[4]

C.2 Published FINRA Awards

The disclosure data identify whether a dispute enters formal arbitration and record its resolution amount, but they reveal relatively little about how the arbitration was conducted. To observe features of the forum itself, we use FINRA's Arbitration Awards Online database, which provides written awards for disputes that reach a panel decision. FINRA posts each award as an individual PDF file. These PDFs identify arbitrators, counsel, allegations, products, expungement and counterclaim requests, relevant dates, and in some cases amended claim amounts.

The award data cover a narrower part of the dispute-resolution process than the disclosure data. FINRA publishes an award only when the dispute reaches a panel decision; arbitrations that settle, are withdrawn, or otherwise close without an award do not appear. The award archive contains 8,253 cases during our sample window. We hand-match the award PDFs back to disclosure records and parse their contents to construct measures of panel composition, claimant representation, product type, expungement, counterclaims, claim revisions, and hearing details. The matching protocol begins with exact case-number links whenever the disclosure row reports a FINRA arbitration number that also appears on the award PDF. When that identifier is missing or unusable, we search candidate awards using advisor and claimant names, firm names, filing or award dates, hearing site, and claimed damages, and we retain a match only when those fields line up uniquely on manual

[4]Appendix Table H1 reports the robustness of the results to dropping unisex names entirely.

review. Parsed fields are then accepted only when the extracted PDF text and the hand check agree on the underlying identifiers and roster structure; ambiguous links and unreadable parses are dropped rather than forced.

The match yields 2,793 advisor-level arbitration rows linked to 1,556 unique award cases. Because the institutional file is organized at the advisor-row level rather than at the unique-case level, later row counts can exceed the number of distinct matched cases whenever one award maps to more than one advisor disclosure row. The match rate is well below 100% because most successful links rely on an exact FINRA case number reported in the disclosure record, and awards without that shared identifier rarely contain enough overlapping person-level information to be matched uniquely back to a CRD-identified advisor. Of the matched records, 2,205 have readable panel rosters, 1,771 contain usable claimant-representation text, and the more specialized institutional tests operate on smaller subsets still. The matched-award sample allows us to examine features of the arbitration process that are not observed in the disclosure data. Because the estimating file remains at the advisor-row level, firm-level clustering does not by itself absorb all within-award dependence; Appendix Table P1 reports two-way firm-and-case clustering for the leading matched-award interactions.

D Measuring Claimant Recovery

Our primary outcome is claimant recovery relative to the amount claimed:

$$R_{i,c,t} = \ln(1 + \text{Resolution}_{i,c,t}) - \ln\left(1 + \text{Damages}_{i,c,t}\right), \tag{1}$$

where i indexes advisors, c indexes complaints, and t indexes calendar years; $\text{Resolution}_{i,c,t}$ is the dollars received by the claimant and $\text{Damages}_{i,c,t}$ is the claimed damages amount. A lower value of $R_{i,c,t}$ means that the claimant recovered a smaller share of the amount claimed.

The resolution amount is recorded on the disclosure row in the channel being analyzed. In non-arbitrated complaints, it is the firm-level resolution amount. In formal arbitration, it is the resolution amount recorded on the arbitration disclosure. Many disputes receive no payment. Those zero-recovery cases remain in the estimation samples, and when $\text{Resolution}_{i,c,t} = 0$, the outcome becomes $R_{i,c,t} = -\ln(1 + \text{Damages}_{i,c,t})$. Zero-payment disputes are common: 55.4% of non-arbitrated complaints and 46.4% of formal-arbitration complaints in the channel-specific baseline estimation samples have zero recovery.[5] The negative sample means therefore reflect both partial recoveries and the large share of denied or unpaid claims.

Because the outcome is log-like and many disputes remain at zero, it combines whether the claimant receives any positive payment with the amount paid conditional on payment. Following Chen and Roth (2024), we interpret $R_{i,c,t}$ as a claimant-recovery index. Appendix Table O1 reports an indicator for any positive recovery, conditional log resolution among paid cases, and a PPML specification in dollar levels.

We winsorize claimed damages at the 0.5th and 99.5th percentiles to limit the influence of extreme reported claim amounts. The resolution amount is left in reported dollars in the baseline specification so that the extensive margin and the full dispersion of positive payouts remain visible in the core claimant-recovery index. Appendix Table O1 shows that symmetric winsorization of the resolution amount leaves the split-sample baseline estimates essentially unchanged. In the main

[5] Among the 74,905 non-arbitrated baseline observations, 33,415 (44.6%) have positive recovery. The conditional-log-resolution regression retains 32,808 observations; the difference reflects additional singleton fixed-effect drops after restricting the sample to paid cases.

arbitration analysis, the denominator uses the filed claim amount recorded in the disclosure data. The latest amended claim amount parsed from the award PDF is used only in the claim-revision analysis and in the corresponding denominator robustness check on the matched-award subset.

E Sample Characteristics

Table 1 reports summary statistics by gender within the two channel-specific samples with reported damages, prior to the fixed-effect absorption used in the regressions. In the non-arbitrated sample, female advisors average 18.1 years of experience versus 20.2 for men and have IA-registration rates of 34.1% versus 32.9%. The broad covariate profiles are otherwise similar across gender. Claimed damages are somewhat larger in the formal-arbitration channel than in the non-arbitrated channel, which is consistent with the idea that larger disputes are more likely to leave the firm and enter the external forum.

[Insert Table 1 about here]

In raw means, claimants recover slightly less against female advisors in both channels. The regression analysis asks where that conditional gap appears once complaint type, claim size, advisor characteristics, firm-state-year complaint environments, and state-year conditions are held constant.

Table 2 describes the published-award subsample used for the institutional tests. Two-thirds of matched cases that can be parsed cleanly involve three-arbitrator panels, and one-third involve single-arbitrator panels. Female arbitrator representation is limited: only 4% of one-arbitrator panels are female, 41% of three-arbitrator panels include at least one woman, and essentially no parsed panels are all female. The limited variation in arbitrator gender also constrains how precisely we can estimate panel-gender interactions.

[Insert Table 2 about here]

F Observation Counts Across Specifications

Sample sizes differ across tables only because each specification requires different data elements; the design is the same throughout. The two channel-specific estimation counts do not sum exactly to the pooled sample because 955 complaints survive pooled estimation but become singleton fixed-effect cells once the sample is split. On the award side, later tables impose additional requirements such as readable panel rosters, claimant-text extraction, or three-arbitrator panels. Appendix Table J1 lists every step and the effective N for each specialized regression.

G Empirical Strategy

G.1 Selection into Formal Arbitration

We first test whether complaints involving female advisors are more likely to enter the formal-arbitration channel:

$$Arbitration_{i,c,t} = \delta Female_i + \lambda \widetilde{D}_{i,c,t} + X'_{i,t}\psi + \mu_{k(c)} + \alpha_{f\times t} + \alpha_{s\times t} + \varepsilon_{i,c,t}, \tag{2}$$

where $Arbitration_{i,c,t}$ equals one when the complaint appears as a formal arbitration disclosure, $\widetilde{D}_{i,c,t} = \ln(1 + \text{Damages}_{i,c,t})$, $X_{i,t}$ contains advisor experience and registration status, $\mu_{k(c)}$ are complaint-type fixed effects, and $\alpha_{f\times t}$ and $\alpha_{s\times t}$ are firm-by-year and state-by-year fixed effects. We estimate this equation as a linear probability model. The coefficient δ tells us whether observably similar complaints involving female advisors are more likely to move from the pooled complaint

universe into formal arbitration.

G.2 Channel-Specific Claimant Recovery

We then estimate the baseline claimant-recovery specification:

$$R_{i,c,t} = \beta Female_i + \gamma \widetilde{D}_{i,c,t} + X'_{i,t}\psi + \mu_{k(c)} + \alpha_{f\times t} + \alpha_{s\times t} + \varepsilon_{i,c,t}. \quad (3)$$

We run equation (3) three ways: on the pooled complaint sample, on non-arbitrated complaints only, and on formal-arbitration complaints only. The comparison shows how the conditional female coefficient differs across the observed resolution channels. As a further discipline on observable selection, Appendix Table R1 later re-estimates the non-arbitrated specification within common-support and high predicted-arbitration-risk subsets constructed from the baseline observables and fixed effects.

G.3 Firm-Environment Heterogeneity

To study the complaint-handling environment, we augment the baseline specification with one firm characteristic at a time:

$$R_{i,c,t} = \beta_1 Female_i + \beta_2 (Female_i \times M_{f,t}) + \gamma \widetilde{D}_{i,c,t} + X'_{i,t}\psi + \mu_{k(c)} + \alpha_{f\times t} + \alpha_{s\times t} + \varepsilon_{i,c,t}, \quad (4)$$

where M denotes one moderator at a time: the complaint rate or female-advisor share in the advisor's firm-state-year cell, the advisor's own experience, or log discretionary RAUM. The complaint-rate, female-share, and log-RAUM measures are mean-centered before interacting with Female, so

the Female main effect in those columns is evaluated at the sample mean of the corresponding moderator rather than at zero. Only the RAUM measure is constant within firm-year cells and therefore absorbed in levels by the firm-by-year fixed effects. The complaint-rate and female-share measures vary at the firm-state-year level, and advisor experience varies at the advisor level, so those main effects remain identified in the reported tables. The coefficient of interest is β_2, which measures how the female–male gap varies with that moderator.

G.4 Published-Award and Career Analyses

We use the same basic control structure in the published-award analyses, replacing $M_{f,t}$ with forum-side variables such as judgment status, panel size, public-arbitrator share, female arbitrator share, claimant-counsel repeat-player status, expungement, and counterclaims. These specifications retain complaint-type, firm-by-year, and state-by-year fixed effects, with standard errors clustered at the firm level. Appendix Table P1 supplements the baseline award tables with two-way firm-and-case clustering and multiple-testing adjustments for the leading interaction family.

For employment outcomes, we estimate

$$Y_{i,c,t} = \theta_1 Female_i + \theta_2 R_{i,c,t} + \theta_3 (Female_i \times R_{i,c,t}) + X'_{i,t}\kappa + \mu_{k(c)} + \alpha_{f\times t} + \alpha_{s\times t} + \eta_{i,c,t}, \qquad (5)$$

separately within the non-arbitrated and formal-arbitration channels, matching the paper's main claimant-recovery split. Here $Y_{i,c,t}$ is one post-resolution employment outcome and $R_{i,c,t}$ is the claimant-recovery measure defined in equation (1). Termination records whether the advisor separates after the complaint-resolution event. Re-employed next year equals one if the advisor is observed anywhere in the registration records by the following year, whether at the same firm or

a different one. Switch firm records a change in employer, while exit industry is evaluated on the separated-advisor subsample and is mechanically right-censored near the end of the sample period. Late re-entry and time to re-entry are also measured only for separated advisors and distinguish delayed returns from immediate post-event employment outcomes. Tables 5, 6, and 11 use this specification. We use the employment analysis to establish the scope of the compensation result rather than as a second headline result.

III Main Results by Dispute-Resolution Channel

This section tests the predictions from Section B. We first examine whether advisor gender predicts entry into arbitration, then compare claimant recovery across the two resolution channels, and finally examine advisors' subsequent careers.

A Selection into Formal Arbitration

We begin with arbitration entry. If complaints against female advisors differ systematically in their merits, those differences may affect whether a complaint enters arbitration. If the gap instead arises during evaluation, arbitration entry should depend on claim size and case characteristics rather than advisor gender.

Table 3 reports estimates of equation (2). The Female coefficient is -0.0180 (s.e. 0.0115), a 95% confidence interval of roughly $[-0.041,\ 0.005]$. Complaint size and IA registration predict arbitration entry strongly; advisor gender does not. The interval rules out a female difference larger than roughly four percentage points in the negative direction and half a percentage point in the

positive, against an arbitration share of roughly one quarter.

[Insert Table 3 about here]

Complaints involving female and male advisors therefore enter formal arbitration at similar rates after controlling for observable case and advisor characteristics. We do not find evidence of broad gender-specific sorting into arbitration at the entry stage.

B Claimant Recovery by Channel

Table 4 reports the claimant-recovery regressions. Panel A gives the pooled and split-sample specifications; Panel B gives a pooled interaction model that tests directly whether the female coefficient differs across channels while allowing baseline slopes to vary.

Figure 2 plots the three estimates with their confidence intervals. The non-arbitrated estimate lies below zero, the formal-arbitration estimate is imprecise and spans zero, and the pooled estimate lies between them.

The pooled all-complaints coefficient is -0.152 (s.e. 0.120). Pooling the two channels combines firm-handled and formal-arbitration complaints and masks the difference between them.

In non-arbitrated complaints, the claimant-recovery index is 0.294 log points lower against female advisors, conditional on complaint type, claim size, advisor characteristics, firm-state-year environment, and state-year conditions. The gap operates on both parts of the outcome. Appendix Table O1 shows that claimants are 2.4 percentage points less likely to receive any positive payment (-0.024, s.e. 0.009) against a base rate of 44.6 percent, and, conditional on payment, they recover 10.3 percent less (-0.109, s.e. 0.039). Combining the two margins, expected recovery against

female advisors is about 15 percent lower than against observably similar male advisors. To illustrate the magnitude, 44.6 percent of non-arbitrated complaints receive a positive payment, and the median payment among paid complaints is $41,000. Applying the estimated differences on the two margins gives illustrative expected-compensation benchmarks of roughly $18,300 for complaints naming male advisors and $15,500 for those naming female advisors—a difference of about $2,800 per complaint. The result therefore reflects both whether claimants are paid and how much they receive when paid.

In formal arbitration, the point estimate is 0.431 (s.e. 0.555), with a confidence interval running from roughly -0.66 to 1.52. Formal arbitration does not show the negative pattern observed in non-arbitrated complaints.

[Insert Table 4 about here]

Panel B directly compares the female coefficient across channels. In the fully interacted specification, the Female $\times$ Formal arbitration interaction is 0.833 (s.e. 0.466, $p = 0.074$). The estimate is positive and large enough to offset the non-arbitrated gap, although the confidence interval is wide at $[-0.08,\ 1.75]$. Appendix Table M1 reports a simpler common-slopes specification with a smaller, statistically insignificant interaction. We therefore interpret the cross-channel test as supporting the direction of the difference rather than a precise estimate of its magnitude.

Two additional checks address whether the result can be explained by differences in case composition. Appendix Table R1 restricts attention to complaints that look observably arbitration-prone—cases that could plausibly have gone either way—and the non-arbitrated female coefficient remains negative there, at -0.245 (s.e. 0.109) in the common-support band. The gap is therefore not confined to disputes that were never candidates for the external forum. And the distributional

evidence in Section B shows the internal gap is present across the recovery distribution rather than confined to a tail.

The main result is that the conditional compensation gap appears in non-arbitrated complaints, where firms handle the outcome, rather than in formal arbitration. If unobserved case weakness alone drives the gap, a similar negative pattern should appear in both channels. Section IV next examines whether the gap also varies with the firm environment, and Section V examines formal arbitration in more detail.

C Downstream Employment Outcomes by Channel

We next examine whether the compensation gap is accompanied by different subsequent career outcomes for female advisors. Equation (5) follows Egan et al. (2022) on post-misconduct labor-market penalties and Honda (2025) on the sensitivity of those outcomes to how disclosure events are classified.

These outcomes establish the scope of the compensation result but do not distinguish the merits and discretion explanations. Both can predict milder career consequences for female advisors when their complaints are resolved more cheaply. The channel and firm-environment tests in Sections B and IV provide the relevant distinction.

Table 5 reports termination following resolution. Female advisors are less likely to be terminated in both channels—by 0.00739 (s.e. 0.00313) in non-arbitrated complaints and 0.0109 (s.e. 0.00491) in formal arbitration—and more claimant-favorable resolutions raise termination probability in both columns. This pattern is consistent with either weaker underlying complaints or differences in how complaints are handled. What it establishes is that the compensation gap is not accompanied by a

harsher post-event labor-market penalty for female advisors, the comparison emphasized by Egan et al. (2022) and Honda (2025).

[Insert Table 5 about here]

Table 6 reports the remaining outcomes. In the non-arbitrated channel, female advisors are more likely to be re-employed the next year (0.0232, s.e. 0.0126) and more likely to re-enter after a delay (0.0107, s.e. 0.00507), with positive Female $\times$ Recovery interactions in the delayed columns (0.00142, s.e. 0.000627 for late re-entry; 0.00623, s.e. 0.00255 for months to re-entry). Switch-firm and exit-industry differences are small. In formal arbitration, the female coefficients and interactions are insignificant throughout.

[Insert Table 6 about here]

The compensation gap therefore does not extend to advisors' subsequent careers. Female advisors do not face a broader career disadvantage in these outcomes, and the formal-arbitration employment results are statistically similar across advisor gender.

IV Firm-Level Heterogeneity in the Non-Arbitrated Channel

Section III shows that the compensation gap is concentrated in non-arbitrated complaints. We now ask whether it varies across firms. If differences in complaint handling matter, the gap should vary with the environment in which firms evaluate complaints rather than with the characteristics of individual advisors. In particular, it should narrow where firms have greater complaint-processing exposure.

A Selection into Arbitration Across Firm Environments

We first revisit arbitration entry, allowing the gender coefficient to vary with complaint-environment characteristics. Table 7 reports the corresponding heterogeneity tests.

Most firm and advisor characteristics do not change the selection result. The Female $\times$ Complaint rate interaction is -0.00105 (s.e. 0.00313), the Female $\times$ Experience interaction is -0.000270 (s.e. 0.000946), and the Female $\times$ log Discr. RAUM interaction is -0.00176 (s.e. 0.00155), all economically tiny and statistically insignificant. The one clear interaction is the local female-advisor share: the coefficient on Female $\times$ Firm female share is 0.210 (s.e. 0.093). A 10-percentage-point increase in the local female share is therefore associated with roughly a 2.1-percentage-point increase in the arbitration propensity of female-advisor complaints relative to otherwise similar male-advisor complaints. The local female-advisor share therefore predicts some gender-specific sorting into arbitration. This means the female-share heterogeneity in recovery below may reflect both complaint handling and differences in which cases remain in the non-arbitrated channel. The complaint-rate interaction shows no comparable entry effect and provides a cleaner test. The broader result remains unchanged: advisor gender does not predict arbitration entry on average.

[Insert Table 7 about here]

B The Non-Arbitrated Recovery Gap Across Firms

Table 8 examines how the compensation gap varies across firms within non-arbitrated complaints. In columns (1) and (4), complaint rate and firm female share are mean-centered before interaction,

so the Female coefficients of -0.294 (s.e. 0.085) and -0.357 (s.e. 0.112) are the female–male gaps at the sample means of those moderators rather than at zero. Column (1) then shows a Female $\times$ Complaint rate interaction of 0.0552 (s.e. 0.0117), while column (4) shows a Female $\times$ Firm female share interaction of 2.167 (s.e. 1.036). By contrast, the Female $\times$ Experience interaction is -0.00297 (s.e. 0.00978) and statistically indistinguishable from zero.

The complaint-rate interaction is economically large. One standard deviation of the firm-state-year complaint rate is about 2.95 complaints per advisor, so the interaction implies a 0.16 log-point narrowing of the gap over that range—more than half of the 0.294 baseline. At one standard deviation above the mean complaint rate, more than half of the baseline female–male gap is absent. Advisor experience, by contrast, is unrelated to the gap. The compensation difference therefore varies with the firm's complaint environment rather than with the advisor's own experience. In column (4), the main effect on Firm female share is -1.250 (s.e. 0.561), which is the slope for male-advisor complaints: a 10-percentage-point increase in local female share is associated with about a 0.125 log-point lower claimant-recovery index against male advisors. The Female $\times$ Firm female share interaction of 2.167 offsets more than that amount, so the corresponding slope for female-advisor complaints is positive. The complaint-rate interaction is the cleaner result because complaint intensity does not also predict female-specific sorting into arbitration. The female-share interaction points in the same direction, but high-female-share environments also send somewhat more female-advisor complaints into arbitration. We therefore interpret that estimate as combining differences in the firm environment with some difference in case composition. Together, the results show that the non-arbitrated compensation gap varies systematically with the environment in which firms handle complaints.

[Insert Table 8 about here]

The firm-size interaction is negative in the RAUM-available subsample under the continuous log-RAUM specification: Female $\times$ log Discr. RAUM equals -0.0263 (s.e. 0.0151). That estimate suggests a somewhat larger non-arbitrated gap at larger firms, but Appendix Table G1 reports the same specification with a large-firm indicator instead of log-RAUM, and there the interaction is 0.111 (s.e. 0.222), statistically insignificant and directionally different. We therefore place less weight on firm size than on complaint rate and female-advisor share. Firm size is less directly connected to complaint handling, and the estimates are available only in the smaller RAUM sample.

The firm-level results strengthen the channel comparison. The gap is concentrated where firms handle complaints internally, and it narrows as firms process more complaints while remaining unrelated to advisor experience. This pattern is consistent with differences in how firms exercise discretion when handling complaints. A case-composition explanation would require complaints against female advisors to become systematically stronger precisely where firms process more disputes. The contemporaneous firm-state-year measure cannot show whether the narrower gap reflects learning, standardization, or another feature of complaint handling; that would require prior complaint exposure to predict subsequent outcomes.[6]

V Opening the Arbitration Black Box

Sections III and IV show that the compensation gap is concentrated in non-arbitrated complaints and varies with the firm environment. We now examine what happens within formal arbitration.

[6]By the Frisch–Waugh–Lovell theorem (Frisch and Waugh, 1933; Lovell, 1963), re-estimating the interaction specifications on the residualized non-arbitrated outcome reproduces the Table 8 interaction coefficients exactly, so we omit the residualized version as redundant.

This analysis is exploratory because FINRA publishes awards only for disputes that reach a panel decision, and only a subset can be matched cleanly back to CRD-identified advisors.

The discretion explanation also makes a prediction within arbitration. Arbitration is not uniformly procedural: judgment has more room to operate in contested cases than in settlements or withdrawals. If discretion matters, any female-related variation within arbitration should be more visible in contested litigation than in observable panel features. We use the matched award sample to examine panel rosters, counsel, claim revisions, and hearing details. Because this sample is small and selectively matched, we discuss the inference limits at the end of Section B.

A Observable Features of Arbitration

We first examine observable features of the arbitration process. The appendix tables report the full set of estimates.

Cases involving female advisors do not sort into different arbitration outcomes. Appendix Table S1 gives female coefficients of -0.009 (s.e. 0.017) for settlements, -0.019 (s.e. 0.014) for judgments, -0.031 (s.e. 0.049) for denials, and -0.001 (s.e. 0.003) for dismissals, all within about three percentage points of zero, estimated on the 21,079 formal-arbitration rows with identifiable outcome status. We therefore find little evidence that cases involving female advisors sort into different arbitration outcomes.

FINRA's procedural thresholds also do not produce gender-specific differences. Small claims are processed differently above and below \$50,000 and \$100,000, but Figures 3 and 4 show no gender-specific bunching at either cutoff, and Appendix Table S2 confirms it parametrically: the Female $\times$ \$50k–\$100k interaction is 0.110 (s.e. 0.970) and the Female $\times$ \$100k+ interaction

is 0.648 (s.e. 0.829). Higher-damage bins do recover more—1.037 (s.e. 0.483) and 1.763 (s.e. 0.509)—but that gradient is not gender-specific.

[Insert Figures 3 and 4 about here]

Restricting the analysis to matched awards does not reveal a clearer pattern in panel-adjudicated cases. Appendix Table S3 reports a female main effect of -0.154 (s.e. 0.507) and a Female $\times$ Judgment interaction of -0.177 (s.e. 0.612), which remains -0.165 (s.e. 0.662) once the judgment effect is allowed to vary with female arbitrator share.

We also do not find detectable differences across the observable panel features we can measure. Appendix Table S4 gives a Female $\times$ Three-arbitrator interaction of 0.542 (s.e. 1.663) and a Female $\times$ Female-arbitrator-share interaction of 0.820 (s.e. 2.853); Appendix Table C1 gives -1.705 (s.e. 3.995) for public-arbitrator share and 0.479 (s.e. 0.475) in judgments; Appendix Table C2 gives 2.508 (s.e. 2.410) for female chairs and -0.489 (s.e. 0.652) for repeat-player arbitrators. The sharpest available test is FINRA's move toward all-public panels, a reform that shifted composition materially: Figure 5 shows the uptake and Appendix Table C3 confirms a first stage of 0.339 (s.e. 0.089). Even that shift produces no detectable female interaction, with estimates of -1.316 (s.e. 2.253), 4.459 (s.e. 3.153), and 8.982 (s.e. 6.331) on the all-public, majority-public, and post-rule indicators.

[Insert Figure 5 about here]

Overall, observable panel features do not produce detectable female interactions in the selectively matched award sample. The standard errors are large, so these estimates cannot rule out arbitration-side effects of the size found in non-arbitrated complaints. They do show that no measured feature of

panel composition stands out as explaining the pattern. We therefore turn to claimant-side litigation in contested cases.

B Claimant-Side Litigation Dynamics

Table 9 reports two tests of claimant-side litigation. In column (1), claimant representation is absorbed by the fixed effects and yields no interpretable interaction. In column (2), claimant-attorney repeat-player status does: the Female $\times$ Repeat-player interaction is -3.755 (s.e. 2.045), while the repeat-player main effect is 0.358 (s.e. 0.637). The estimated difference therefore comes from the interaction rather than from a uniform recovery advantage associated with repeat-player counsel.

The sign does not imply that experienced counsel reduce recovery. Rather, the recovery advantage associated with repeat-player claimant counsel is smaller in cases involving female advisors. This pattern can reflect differences in the cases taken by repeat-player counsel or differences in how those cases proceed once they become contested. The estimate therefore points to claimant-side litigation rather than to a uniform disadvantage created by experienced counsel.

[Insert Table 9 about here]

Panel B examines claim revisions and must be read jointly across the main and interaction effects. The dependent variable is the change from initial claimed damages to the latest amount observed in the award PDF, so lower values imply larger downward revisions. The Female main effect of 1.825 (s.e. 0.720) says that outside judgment cases, claims involving female advisors are revised down less. The Female $\times$ Judgment interaction of -2.593 (s.e. 1.352) reverses that in

judgment cases, implying a female effect of about -0.77 log points there. Thus, claims against female advisors are revised down more sharply in the contested judgment subset.

These estimates should be interpreted cautiously. The matched-award sample is small and selectively linked. Appendix Table P1 shows that two-way clustering by firm and award case leaves the nominal p-values almost unchanged, but neither interaction survives Benjamini–Hochberg or Bonferroni adjustment across the award-side family. We therefore treat the claimant-counsel and claim-revision results as exploratory. Within the dimensions we can observe, the results point more toward how contested cases are litigated than toward who sits on the arbitration panel.

VI Robustness and Scope Conditions

This section examines whether the by-channel pattern is robust to alternative outcome definitions, samples, time windows, and coding choices. We also report additional checks for the matched-arbitration analysis.

A Outcome Construction with Many Zero Payments

Appendix Table O1 addresses an important feature of the baseline outcome: roughly half of the disputes in the estimation sample receive zero payment. The claimant-recovery measure therefore combines whether a claimant receives any payment with the amount received when paid. Following Chen and Roth (2024), we therefore do not interpret the baseline log-like coefficient as a literal percentage effect in dollars recovered.

The decomposition shows that the non-arbitrated gap appears on both margins. The female

coefficient is negative for the probability of any positive recovery and for conditional log resolution among paid cases. The PPML specification in dollar levels is imprecise in every channel, so we do not interpret the baseline coefficient as a dollar elasticity. This suggests that the gap is more visible in proportional recovery than in dollar-weighted totals. Appendix Table O1 also shows that symmetric winsorization of the resolution amount leaves the split-sample baseline estimates essentially unchanged, so the non-arbitrated result is not being driven by a few extreme payouts.

B Distributional Evidence Across Channels

Table 10 asks whether the gender difference is concentrated in a small part of the recovery distribution. The pattern differs by channel. In non-arbitrated complaints, the female coefficient in the lowest predicted-recovery quartile is -0.383 (s.e. 0.187). The higher-quartile interactions are 0.131 (s.e. 0.167), 0.170 (s.e. 0.235), and 0.0604 (s.e. 0.188), so the gap does not vary systematically across the upper three quartiles. In formal arbitration, by contrast, the lowest-quartile female coefficient is 0.398 (s.e. 0.687), and the Q2–Q4 interactions of -1.070 (s.e. 0.965), -0.290 (s.e. 0.787), and 1.439 (s.e. 1.309) are all statistically insignificant and do not trace a stable pattern.

[Insert Table 10 about here]

The non-arbitrated gap is therefore not confined to an extreme tail of the distribution, while formal arbitration does not show a comparable pattern across quartiles.

C Observable Arbitration Proneness

Appendix Table R1 asks whether the non-arbitrated result is confined to complaints that never looked likely to reach arbitration in the first place. We construct a predicted-arbitration-risk index from the baseline observables and fixed effects, excluding advisor gender, and then re-estimate the non-arbitrated claimant-recovery regression within progressively more arbitration-prone subsets.

The negative non-arbitrated female coefficient remains in these restricted samples. In the common-support overlap band, it is -0.245 (s.e. 0.109), and in the top half of the predicted-risk distribution it is -0.304 (s.e. 0.137). In the top quartile of predicted arbitration risk, the estimate attenuates to -0.177 (s.e. 0.180) and loses significance, but its confidence interval still includes the full-sample gap. The pooled overlap-band interaction remains positive but statistically insignificant. The overlap exercise therefore leaves the main by-channel result unchanged.

D Time Trends

Appendix Figure K1 and Appendix Table K1 examine whether the non-arbitrated result is driven by one unusual part of the sample period. The non-arbitrated female coefficient becomes more negative after the mid-2010s. In formal arbitration, the early-period estimate is positive and statistically significant, but the coefficient drifts downward and becomes imprecise in the later subperiods. The formal-arbitration estimates therefore do not show a stable time pattern. Across subperiods, the non-arbitrated coefficient is estimated more consistently.

E Name Coding and Award-Match Balance

Appendix Table H1 shows that the main result is not created by the treatment of ambiguous names. GenderChecker leaves no advisor unmatched. When we drop unisex names entirely and compare only clearly male and clearly female advisors, the non-arbitrated female coefficient remains close to baseline, while the formal-arbitration coefficient remains statistically insignificant. Dropping unisex names therefore leaves the main result essentially unchanged.

Appendix Table I1 examines whether the published-award subsample is strongly selected by advisor gender. The female share differs only modestly between matched and unmatched arbitration rows. The larger differences between matched and unmatched observations are economic and procedural, which is expected because written awards represent a narrower subset of the broader arbitration universe that also includes settlements and withdrawals.

Appendix Table Q1 addresses a separate sample-construction issue. In the raw file, the excluded complaint rows are exactly the observations with both missing claimed damages and missing channel coding, so a clean conditional test for damages reporting is not identified. The relevant restriction is therefore entry into the fully coded claimant-recovery sample rather than selection on reported damages alone. Female-advisor complaints are modestly more likely to enter the paper sample descriptively—86.0% versus 83.8% for male-advisor complaints—so the restriction is not selecting only unusually strong female cases. The raw file still does not support a sharper conditional damages-reporting test.

F Other Case Features

Appendix Tables D1 and E1 examine two additional families of award-level covariates. The product-type subsamples are small, so we report those results descriptively. Expungement requests and counterclaims affect the level of award outcomes in expected ways, but they do not generate meaningful interactions with advisor gender.

G Award-Sample Inference Diagnostics

Appendix Table P1 reports the main inference checks for the award-side interactions. Two-way clustering by firm and award case leaves the nominal p-values for the claimant-attorney-repeat and claim-revision interactions almost unchanged. However, neither interaction survives Benjamini–Hochberg or Bonferroni adjustment across the main award-side family. We therefore treat both results as exploratory. Wild-cluster bootstrap p-values are not reported for these absorbed-interaction specifications.

H Threshold and Placebo Diagnostics

Appendix Table F1 reports local threshold analyses around the $50,000 and $100,000 procedural cutoffs, and Appendix Table L1 reports fake-cutoff and rule-date exercises. The first stages around the procedural thresholds do not support a regression-discontinuity design, and the fake-cutoff placebos are not clean nulls. We therefore do not use the thresholds for identification. The main analysis instead relies on the full-sample channel comparison and firm-environment heterogeneity.

Finally, Table 11 replaces the observed claimant-recovery measure in the by-channel employ-

ment regressions with residual claimant recovery. The immediate-re-employment advantage for women in the non-arbitrated channel remains under residualization, the delayed-re-entry result becomes statistically insignificant, and the formal-arbitration interactions remain statistically insignificant throughout. That stability leaves the main channel comparison intact.

VII Conclusion

Customer complaints against financial advisors can be resolved within the employing firm or in formal FINRA arbitration. Using 99,674 complaints with reported damages, we show that customers receive less compensation when their complaint names a female advisor, but the difference is concentrated in non-arbitrated complaints. There, claimants are 2.4 percentage points less likely to receive payment and, when paid, receive 10.3 percent less than customers with otherwise similar male-advisor complaints. Together, these margins imply about 15 percent lower expected compensation. Complaints involving female advisors enter formal arbitration at similar rates, and formal arbitration does not show the same negative pattern.

The firm-level results point to differences in complaint handling. The non-arbitrated compensation gap narrows by more than half when the firm-state-year complaint rate is one standard deviation higher, while the gap does not vary with advisor experience. The result also remains negative when we restrict the sample to complaints that were observably plausible candidates for arbitration. These patterns show that the gap varies with the environment in which firms handle complaints rather than with the individual advisor.

Within formal arbitration, observable panel features do not account for the pattern in the matched-award analysis. Panel size, arbitrator gender, public-arbitrator share, female chairs, repeat-player

arbitrators, and FINRA's 2011 shift toward all-public panels do not produce detectable female interactions. The differences that do appear concern claimant-side litigation: the recovery advantage associated with repeat-player claimant counsel is smaller in cases involving female advisors, and claims against female advisors are revised down more sharply before judgment. Because these estimates come from a small, selectively matched sample and do not survive multiple-testing correction, we treat them cautiously.

The compensation gap also does not extend to advisors' subsequent careers. Female advisors are less likely to be terminated in both channels and are modestly more likely to re-enter employment after non-arbitrated complaints. In formal arbitration, their employment outcomes are statistically similar to those of male advisors. The paper's main result is therefore about differences in customer compensation rather than a broader career disadvantage for female advisors.

These findings shift attention toward firm-level complaint handling. Disclosure records report whether a complaint is paid and how much the customer receives, but they do not show how firms evaluate or settle complaints internally. The evidence is consistent with the greater discretion firms have in handling complaints, compared with the standardized procedures of formal arbitration, as an important difference between the two resolution processes. The fact that the gap narrows sharply at firms with higher complaint rates also identifies complaint-processing exposure as a useful margin for future research and regulatory attention.

The analysis has two main limitations. Firms and claimants determine whether a dispute enters arbitration, so the channel comparison is conditional on the observed resolution channel. We also do not observe the within-firm bargaining process, and the published-award sample includes only disputes that reach a written award. Data on internal negotiations, amended claims, and

counsel strategy would allow a more direct analysis of how firms evaluate and settle non-arbitrated complaints. The present evidence establishes where the compensation gap appears: in firm-handled complaints rather than in formal arbitration.

References

Arnold, David, Will Dobbie, and Crystal S. Yang (2018) "Racial Bias in Bail Decisions," *Quarterly Journal of Economics*, 133 (4), 1885–1932, 10.1093/qje/qjy012.

Bertrand, Marianne, Claudia Goldin, and Lawrence F. Katz (2010) "Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors," *American Economic Journal: Applied Economics*, 2 (3), 228–255, 10.1257/app.2.3.228.

Bhattacharya, Utpal, Andreas Hackethal, Simon Kaesler, Benjamin Loos, and Steffen Meyer (2012) "Is Unbiased Financial Advice to Retail Investors Sufficient? Answers from a Large Field Study," *Review of Financial Studies*, 25 (4), 975–1032, 10.1093/rfs/hhr135.

Blau, Francine D. and Lawrence M. Kahn (2017) "The Gender Wage Gap: Extent, Trends, and Explanations," *Journal of Economic Literature*, 55 (3), 789–865, 10.1257/jel.20160995.

Charoenwong, Ben, Alan Kwan, and Tarik Umar (2019) "Does Regulatory Jurisdiction Affect the Quality of Investment-Adviser Regulation?" *American Economic Review*, 109 (10), 3681–3712, 10.1257/aer.20180412.

Chen, Jiafeng and Jonathan Roth (2024) "Logs with Zeros? Some Problems and Solutions," *Quarterly Journal of Economics*, 139 (2), 891–936, 10.1093/qje/qjad054.

Clifford, Christopher P., William Christopher Gerken, and Tian Qiu (2023) "Racial Concordance in the Market for Financial Advice," 10.2139/ssrn.4323223, SSRN Scholarly Paper No. 4323223.

Dimmock, Stephen G. and William C. Gerken (2012) "Predicting Fraud by Investment Managers," *Journal of Financial Economics*, 105 (1), 153–173, 10.1016/j.jfineco.2012.01.002.

Egan, Mark, Gregor Matvos, and Amit Seru (2018) "Arbitration with Uninformed Consumers," *American Economic Review*, 108 (8), 2331–2366.

——— (2019) "The Market for Financial Adviser Misconduct," *Journal of Political Economy*, 127 (1), 233–295, 10.1086/700735.

——— (2022) "When Harry Fired Sally: The Double Standard in Punishing Misconduct," *Journal of Political Economy*, 130 (5), 1184–1248, 10.1086/718964.

Frisch, Ragnar and Frederick V. Waugh (1933) "Partial Time Regressions as Compared with Individual Trends," *Econometrica*, 1 (4), 387–401, 10.2307/1907330.

Galanter, Marc (1974) "Why the "Haves" Come Out Ahead: Speculations on the Limits of Legal Change," *Law & Society Review*, 9 (1), 95–160, 10.2307/3053023.

Gelman, Michael, Zaheer Khan, Amir Shoham, and Shlomo Y. Tarba (2021) "Does Local Competition and Firm Market Power Affect Investment Adviser Misconduct?" *Journal of Corporate Finance*, 66, 101810, 10.1016/j.jcorpfin.2020.101810.

GenderChecker (2020) "GenderChecker Name Database," Commercial name-gender classification database.

Goldin, Claudia (2014) "A Grand Gender Convergence: Its Last Chapter," *American Economic Review*, 104 (4), 1091–1119, 10.1257/aer.104.4.1091.

Goldin, Claudia and Cecilia Rouse (2000) "Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians," *American Economic Review*, 90 (4), 715–741, 10.1257/aer.90.4.715.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales (2008) "Trusting the Stock Market," *Journal of Finance*, 63 (6), 2557–2600, 10.1111/j.1540-6261.2008.01408.x.

Gurun, Umit G., Noah Stoffman, and Scott E. Yonker (2018) "Trust Busting: The Effect of Fraud on Investor Behavior," *Review of Financial Studies*, 31 (4), 1341–1376, 10.1093/rfs/hhx058.

Honda, Jun (2025) "Financial Adviser Misconduct and Labor Market Penalties: Uncovering Racial Disparities in the Absence of Gender Gaps," https://arxiv.org/abs/2503.12837, arXiv:2503.12837.

Lovell, Michael C. (1963) "Seasonal Adjustment of Economic Time Series and Multiple Regression Analysis," *Journal of the American Statistical Association*, 58 (304), 993–1010, 10.1080/01621459.1963.10480682.

Rehavi, M. Marit and Sonja B. Starr (2014) "Racial Disparity in Federal Criminal Sentences," *Journal of Political Economy*, 122 (6), 1320–1354, 10.1086/677255.

Sarsons, Heather (2019) "Interpreting Signals in the Labor Market: Evidence from Medical Referrals," Working paper, University of British Columbia.

Stoughton, Neal M., Youchang Wu, and Josef Zechner (2011) "Intermediated Investment Management," *Journal of Finance*, 66 (3), 947–980.

Figures

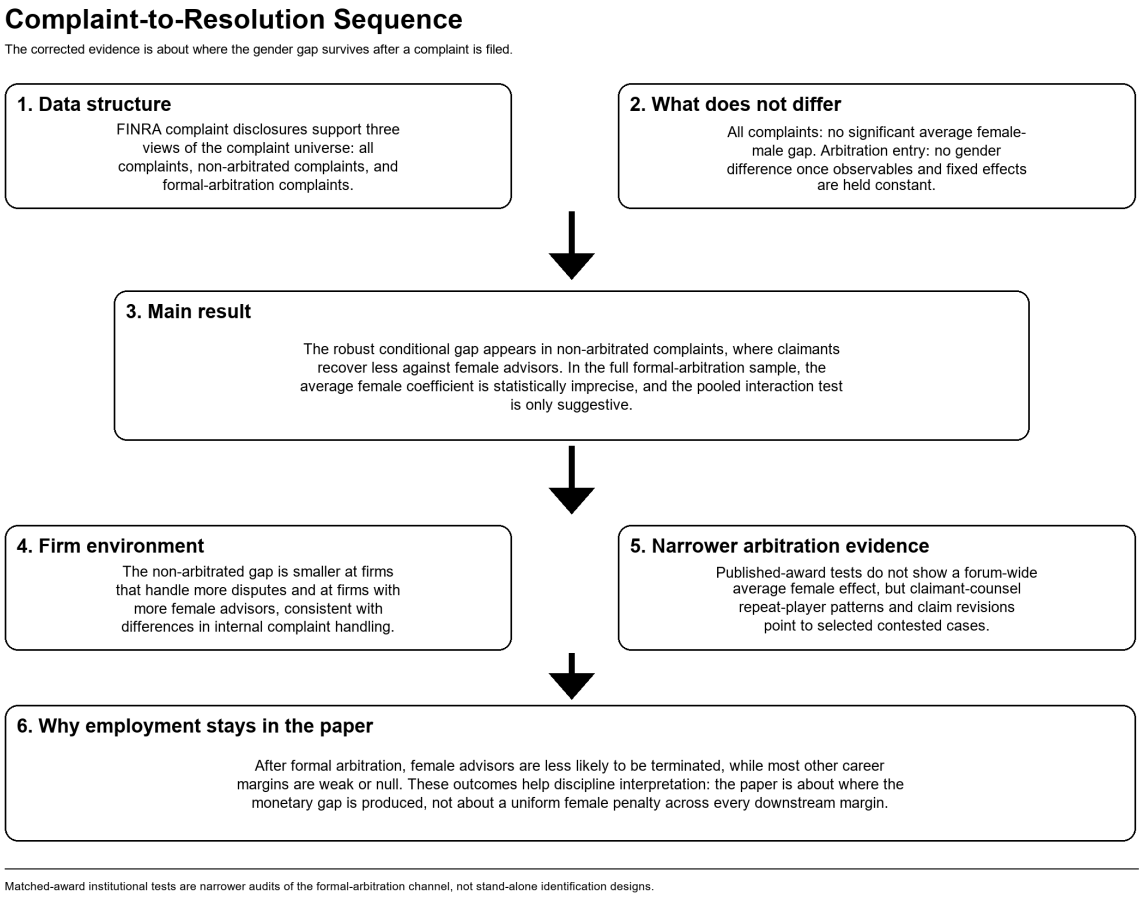
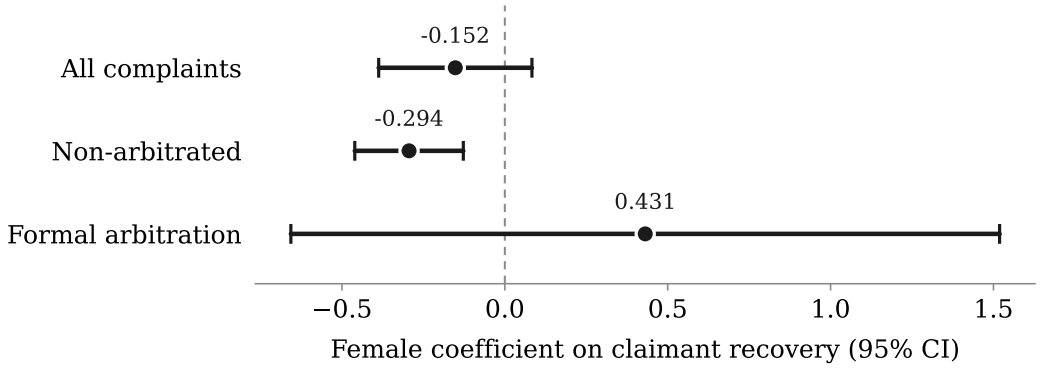


Figure 1: Complaint-to-Resolution Sequence and Main Empirical Margins

Notes: This figure summarizes the empirical logic of the paper. The robust and precisely estimated female–male gap appears in non-arbitrated complaints. In formal arbitration, the point estimate is positive but not precisely estimated, and pooled interaction tests move in the same direction. The matched-award tests open the arbitration forum itself; within that selected sample, the female-related patterns that remain visible line up with claimant-side litigation dynamics rather than with measured panel composition.

Figure 2: The Female Coefficient on Claimant Recovery, by Resolution Channel



Notes: Each row plots the Female coefficient from equation (3) with its 95% confidence interval, estimated on the pooled complaint sample, on non-arbitrated complaints, and on formal-arbitration complaints. All three specifications use the same controls and complaint-type, firm-by-year, and state-by-year fixed effects, with standard errors clustered at the firm level. The corresponding estimates appear in Table 4.

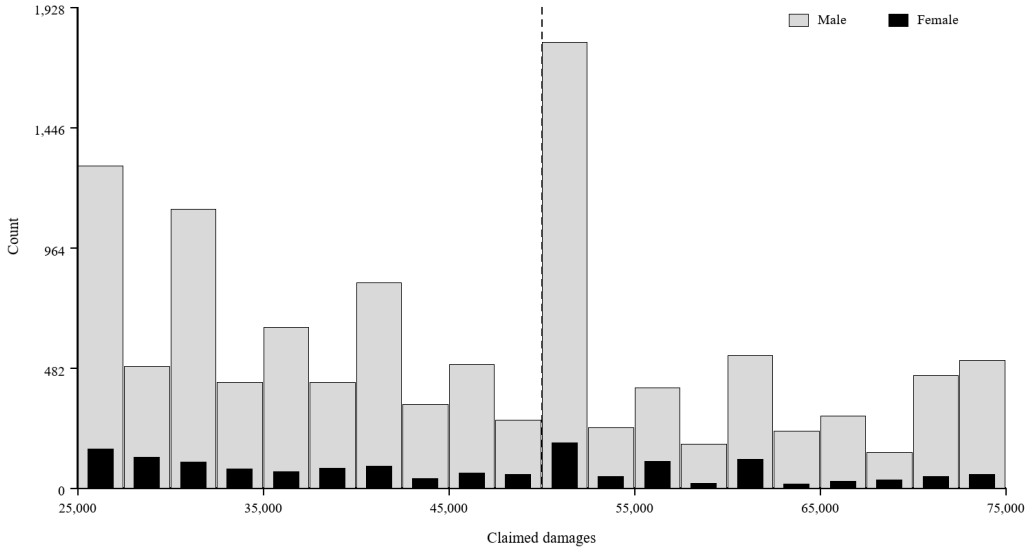


Figure 3: Distribution of Claimed Damages Around the $50,000 Threshold

Notes: Light gray bars show the male-advisor distribution and black bars show the female-advisor distribution in the formal-arbitration sample. The dashed line marks the $50,000 procedural threshold.

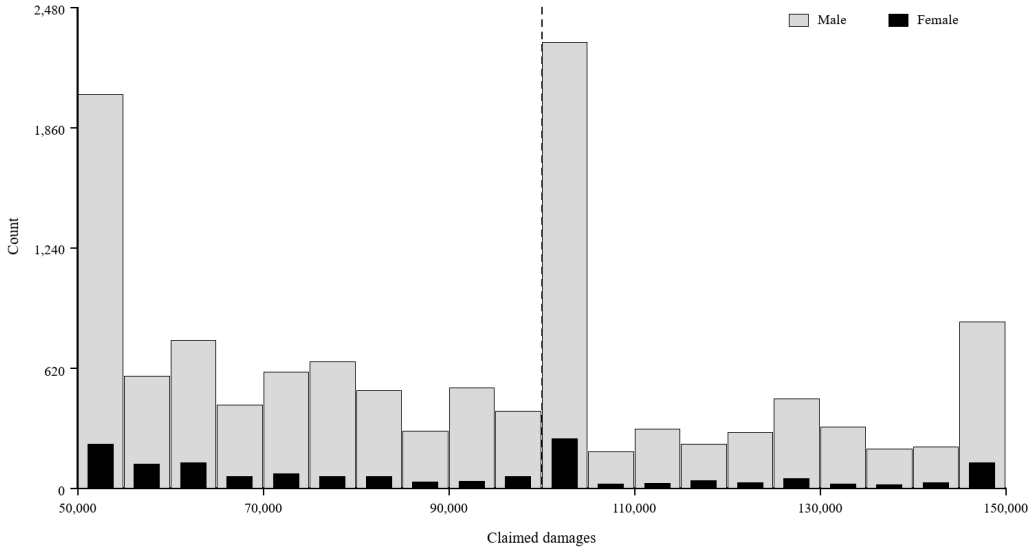


Figure 4: Distribution of Claimed Damages Around the $100,000 Threshold

Notes: Light gray bars show the male-advisor distribution and black bars show the female-advisor distribution in the formal-arbitration sample. The dashed line marks the $100,000 threshold associated with three-arbitrator procedures.

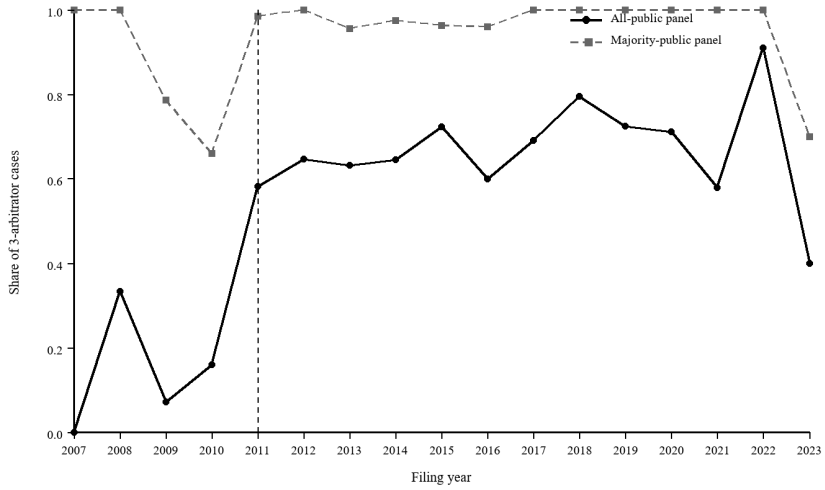


Figure 5: All-Public Panel Uptake After 2011

Notes: The figure reports the share of all-public and majority-public panels by year in the matched three-arbitrator award sample. The dashed vertical line marks the 2011 rule change that made the all-public option permanent.

Tables

Table 1: Summary Statistics by Channel and Advisor Gender

	Male	Female
Panel A. Non-arbitrated complaints		
Claimant recovery (R)	−3.039 (6.673)	−3.282 (6.599)
Firm female share	0.171 (0.085)	0.232 (0.120)
Experience (years)	20.234 (10.246)	18.127 (9.629)
IA registration	0.329 (0.470)	0.341 (0.474)
Log damages	7.827 (5.256)	7.588 (5.215)
Positive damages	0.708 (0.455)	0.697 (0.460)
Observations	69,769	7,451
Panel B. Formal arbitration		
Claimant recovery (R)	−2.644 (6.398)	−3.042 (6.555)
Firm female share	0.160 (0.092)	0.278 (0.191)
Experience (years)	17.142 (8.896)	16.150 (7.521)
IA registration	0.715 (0.452)	0.767 (0.423)
Log damages	8.556 (5.323)	8.920 (4.667)
Positive damages	0.739 (0.439)	0.807 (0.395)
Observations	23,160	1,967

Notes: Means with standard deviations in parentheses. These summary statistics are computed on the full channel-specific samples with reported damages, prior to fixed-effect absorption in the regression specifications. The corresponding claimant-recovery regressions therefore use slightly smaller samples: 74,905 non-arbitrated rows and 23,814 formal-arbitration rows. Claimant recovery (R) is the log resolution-to-damages ratio $\ln(1+\text{Resolution}) - \ln(1+\text{Damages})$. Indicator variables appear as sample shares in this table. Table A1 defines each variable used in the paper.

Table 2: Matched-Award Panel Composition

Panel type	N	Share	Any female arbitrator	All female panel
One-arbitrator panels	476	34%	4%	–
Three-arbitrator panels	939	66%	41%	0%
Parsed one- and three-arbitrator cases	1,415	100%	–	–

Notes: This table uses the case-level FINRA award roster file underlying the matched-award institutional analysis. It is limited to the 1,542 matched award cases with readable case-level text and a parsed panel roster. Rare parsed panels with sizes other than one or three (127 cases) are excluded so that the tabulation matches the one- versus three-arbitrator distinction used in the institutional tests.

Table 3: Selection into the Formal-Arbitration Channel

	Arbitration
Female	-0.0180
	(0.0115)
Experience (years)	-0.00300^{***}
	(0.000603)
IA registration	0.221^{***}
	(0.0169)
Log damages	0.0422^{***}
	(0.00245)
Positive damages	-0.418^{***}
	(0.0263)
Complaint Type FE	Yes
Firm $\times$ Year FE	Yes
State $\times$ Year FE	Yes
Observations	99,674
Firm clusters	1,331
R^2	0.451

Notes: Linear-probability estimates on the complaint sample with reported damages. The dependent variable equals one when the complaint later appears as a formal-arbitration disclosure and zero for non-arbitrated complaint disclosures. Robust standard errors clustered at the firm level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 4: Claimant Recovery Across Dispute-Resolution Channels

Panel A. Baseline by-channel specifications

	All complaints (1)	**Non-arbitrated** (2)	**Formal arbitration** (3)
Female	−0.152	−0.294***	0.431
	(0.120)	(0.0940)	(0.555)
Experience (years)	0.00789***	0.00424	0.0413***
	(0.00288)	(0.00342)	(0.0138)
IA registration	−0.287***	−0.346***	−1.373***
	(0.0600)	(0.0330)	(0.215)
Log damages	−0.00888	−0.115***	−0.137
	(0.0425)	(0.0362)	(0.123)
Positive damages	−9.835***	−8.702***	−8.848***
	(0.490)	(0.445)	(1.284)
Complaint Type FE	Yes	Yes	Yes
Firm × Year FE	Yes	Yes	Yes
State × Year FE	Yes	Yes	Yes
Observations	99,674	74,905	23,814
Firm clusters	1,331	1,003	787
R^2	0.578	0.594	0.737

Panel B. Formal cross-channel interaction test

	Pooled interaction (4)
Female (non-arbitrated effect)	−0.280***
	(0.103)
Female × Formal arbitration	0.833*
	(0.466)
Implied female effect in formal arbitration	0.553
	(0.455)
Baseline controls × Formal arbitration	Yes
Complaint Type FE	Yes
Firm × Year FE	Yes
State × Year FE	Yes
Observations	99,674
Firm clusters	1,331

Notes: OLS estimates of the baseline claimant-recovery specification. The dependent variable is the log resolution-to-damages ratio $R = \ln(1 + \text{Resolution}) - \ln(1 + \text{Damages})$. Panel A reports the pooled and split-sample regressions. Column (1) uses the full complaint sample with reported damages. Column (2) restricts to non-arbitrated complaints, and column (3) restricts to formal-arbitration disclosures. Columns (2) and (3) do not sum exactly to column (1) because 955 complaints survive in the pooled specification but drop as singleton fixed-effect cells when the sample is split by channel and re-estimated separately. Because zero payments are common, Appendix Table O1 decomposes this baseline outcome into payment incidence and conditional payment size and shows that the non-arbitrated coefficient should be read as a claimant-recovery-index difference rather than as a literal percentage difference in dollars recovered. Panel B reports the pooled interaction test closest to the split-sample specification: the Female coefficient is the non-arbitrated effect, the Female × Formal arbitration coefficient is the cross-channel difference, and the implied female effect in formal arbitration is their sum. All baseline controls are interacted with the arbitration indicator in Panel B. Appendix Table M1 reports the simpler common-slopes interaction and the coefficient path across specifications. Robust standard errors clustered at the firm level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 5: Termination Following Complaint Resolution, by Channel

	Non-arbitrated complaints	**Formal arbitration**
	(1)	(2)
Female	-0.00739^{**}	-0.0109^{**}
	(0.00313)	(0.00491)
Claimant recovery (R)	0.00399^{***}	0.00234^{***}
	(0.000448)	(0.000818)
Experience	-0.000348^{*}	-0.000245
	(0.000190)	(0.000346)
IA registration	-0.0255^{***}	-0.00648
	(0.00206)	(0.00494)
Log damages	0.00171^{**}	0.00112
	(0.000841)	(0.00143)
Positive damages	0.0116	-0.00197
	(0.0104)	(0.0171)
All FEs	Yes	Yes
Observations	74,905	23,814
Firm clusters	1,003	787
R^2	0.278	0.614

Notes: OLS estimates of post-resolution termination by channel. Column (1) uses the non-arbitrated complaint regression sample and column (2) uses the formal-arbitration regression sample. Both columns include the baseline claimant-recovery controls, complaint-type fixed effects, firm-by-year fixed effects, and state-by-year fixed effects. Robust standard errors clustered at the firm level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 6: Post-Resolution Employment Outcomes, by Channel

Panel A. Immediate outcomes

	Non-arbitrated complaints			Formal arbitration		
	Re-employed next year	Switch firm	Exit industry	Re-employed next year	Switch firm	Exit industry
Female	0.0232^{*}	0.00635	-0.00639	0.0151	0.000814	0.0553
	(0.0126)	(0.00478)	(0.0154)	(0.0213)	(0.00504)	(0.0527)
Recovery	-0.00214^{***}	-5.52×10^{-5}	0.00289^{***}	0.00782^{***}	0.000586^{*}	-0.00621^{*}
	(0.000497)	(0.000204)	(0.000770)	(0.00114)	(0.000351)	(0.00357)
Female × Recovery	2.90×10^{-5}	1.20×10^{-5}	0.000379	-0.00339	-0.000894	0.00140
	(0.00111)	(0.000485)	(0.00164)	(0.00218)	(0.000903)	(0.00689)
All FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	74,905	74,905	7,436	23,814	23,814	582
R^2	0.330	0.447	0.835	0.339	0.382	0.827

Panel B. Delayed outcomes

	Non-arbitrated complaints		Formal arbitration	
	Late re-entry	Months to re-entry	Late re-entry	Months to re-entry
Female	0.0107^{**}	0.0337	-0.00301	-0.0441
	(0.00507)	(0.0241)	(0.00904)	(0.0750)
Recovery	-0.000510	0.000850	0.000370	0.00519
	(0.000362)	(0.00136)	(0.00111)	(0.00838)
Female × Recovery	0.00142^{**}	0.00623^{**}	-0.000240	-0.00796
	(0.000627)	(0.00255)	(0.00183)	(0.0132)
All FEs	Yes	Yes	Yes	Yes
Observations	7,436	4,148	582	384
R^2	0.375	0.356	0.664	0.700

Notes: OLS estimates of post-resolution employment outcomes by channel. Panel A reports next-year re-employment, firm switching, and industry exit. Panel B reports late re-entry and months to re-entry on the separated-advisor subsample. The non-arbitrated columns use the non-arbitrated complaint regression sample and the formal-arbitration columns use the arbitration regression sample. "Recovery" denotes claimant recovery (R) entering the employment regression as a right-hand-side regressor. All specifications include complaint-type fixed effects, firm-by-year fixed effects, state-by-year fixed effects, and the baseline claimant-recovery controls. Robust standard errors clustered at the firm level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 7: Firm Environment and Selection into Formal Arbitration

	(1) Complaint rate	**(2)** Experience	**(3)** Firm size	**(4)** Firm female share
Female	−0.0180	−0.0132	−0.0252*	−0.0274**
	(0.0114)	(0.0198)	(0.0131)	(0.0115)
Complaint rate	0.0121***			
	(0.00402)			
Female × Complaint rate	−0.00105			
	(0.00313)			
Experience	−0.00301***	−0.00298***	−0.00413***	−0.00301***
	(0.000604)	(0.000596)	(0.000614)	(0.000604)
Female × Experience		−0.000270		
		(0.000946)		
Female × log Discr. RAUM			−0.00176	
			(0.00155)	
Firm female share				0.0108
				(0.0606)
Female × Firm female share				0.210**
				(0.0929)
All FEs	Yes	Yes	Yes	Yes
Observations	99,674	99,674	66,793	99,674
R^2	0.451	0.451	0.411	0.451

Notes: Linear-probability estimates of arbitration-entry heterogeneity. Each column adds one complaint-environment moderator and its interaction with Female. Complaint rate and firm female share are measured at the advisor's firm-state-year level in the estimating file, so their main effects remain identified and are reported even with firm-by-year and state-by-year fixed effects. In the continuous complaint-rate, female-share, and RAUM specifications, the moderator is mean-centered before interaction, so the Female main effect is evaluated at the sample mean of that moderator. Column (3) uses the smaller RAUM-available sample. Robust standard errors clustered at the firm level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 8: Firm Environment and Non-Arbitrated Claimant Recovery

	(1) Complaint rate	(2) Experience	(3) Firm size	(4) Firm female share
Female	-0.294^{***}	-0.240	-0.228^{*}	-0.357^{***}
	(0.0850)	(0.210)	(0.132)	(0.112)
Complaint rate	0.149			
	(0.0966)			
Female × Complaint rate	0.0552^{***}			
	(0.0117)			
Experience	0.00426	0.00450	0.00603	0.00418
	(0.00340)	(0.00343)	(0.00373)	(0.00345)
Female × Experience		-0.00297		
		(0.00978)		
Female × log Discr. RAUM			-0.0263^{*}	
			(0.0151)	
Firm female share				-1.250^{**}
				(0.561)
Female × Firm female share				2.167^{**}
				(1.036)
All FEs	Yes	Yes	Yes	Yes
Observations	74,905	74,905	49,601	74,905
R^2	0.595	0.594	0.573	0.594

Notes: OLS estimates of the non-arbitrated claimant-recovery specification. Each column adds one complaint-environment moderator and its interaction with Female. Complaint rate and firm female share are measured at the advisor's firm-state-year level in the estimating file, so their main effects remain identified and are reported even with firm-by-year and state-by-year fixed effects. In the continuous complaint-rate, female-share, and RAUM specifications, the moderator is mean-centered before interaction, so the Female main effect is evaluated at the sample mean of that moderator. The column (1) Female coefficient coincides with the baseline non-arbitrated estimate in Table 4 because mean-centering evaluates it at the sample mean of the moderator; its standard error differs because the specification also estimates the interaction. Column (3) uses the RAUM-available subsample. Robust standard errors clustered at the firm level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 9: Claimant-Side Litigation Dynamics in Matched Awards

	Panel A. Dependent variable: Claimant recovery (R)		Panel B. Dependent variable: Log claim revision
	(1) Claimant represented	(2) Attorney repeat-player	(3) Female × Judgment
Female advisor	−1.358	−0.296	1.825**
	(0.847)	(1.087)	(0.720)
Claimant represented	Omitted		
Female × claimant represented	Omitted		
Claimant-attorney repeat player		0.358	
		(0.637)	
Female × claimant-attorney repeat player		−3.755*	
		(2.045)	
Judgment award			0.305
			(0.344)
Female × judgment award			−2.593*
			(1.352)
Experience	0.0231	0.0203	−0.0194
	(0.0253)	(0.0250)	(0.0120)
IA registration	0.595	0.555	−0.00363
	(0.535)	(0.520)	(0.336)
Log damages	−1.504***	−1.475***	−0.791***
	(0.161)	(0.168)	(0.169)
Positive damages	5.200**	4.707**	−2.021
	(2.228)	(2.308)	(2.176)
All FEs	Yes	Yes	Yes
Observations	986	986	1,206
R^2	0.744	0.746	0.781

Notes: Panel A reports matched-award claimant-recovery regressions. Columns (1) and (2) begin from 1,771 matched-award rows with usable claimant-representation text and use 986 rows after baseline-covariate screening and fixed-effect absorption; the claimant-represented indicator and its interaction with Female are omitted in column (1) because they are collinear with the fixed effects in that subset. These columns use 141 firm clusters and 548 unique award cases. Panel B reports the claim-revision regression, where the dependent variable is $\ln(1+\text{Latest claimed damages}) - \ln(1+\text{Initial claimed damages})$ using the parsed award PDF. Lower values therefore indicate larger downward revisions in claimed damages. Column (3) begins from 2,069 latest-claim-known matched-award rows and uses 1,206 after fixed-effect absorption, with 169 firm clusters and 692 unique award cases. All specifications include complaint-type fixed effects, firm-by-year fixed effects, and state-by-year fixed effects, with robust standard errors clustered at the firm level. Appendix Table P1 reports the corresponding two-way firm-and-case clustering and multiple-testing diagnostics. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 10: Distributional Gender Gaps by Resolution Channel

	Non-arbitrated	**Formal arbitration**
	(1)	(2)
Female (Q1)	-0.383^{**}	0.398
	(0.187)	(0.687)
Female × Q2	0.131	−1.070
	(0.167)	(0.965)
Female × Q3	0.170	−0.290
	(0.235)	(0.787)
Female × Q4	0.0604	1.439
	(0.188)	(1.309)
Experience	0.00485	0.0493^{***}
	(0.00331)	(0.0156)
All FEs	Yes	Yes
Observations	74,905	23,814
R^2	0.595	0.739

Notes: Main channel-specific specifications augmented with Female-by-quartile interactions based on predicted claimant recovery within each channel. The omitted category is the lowest predicted-recovery quartile. Robust standard errors clustered at the firm level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 11: Employment Outcomes by Channel: Residual Claimant Recovery

Panel A. Immediate outcomes

	Non-arbitrated complaints			Formal arbitration		
	Re-employed next year	Switch firm	Exit industry	Re-employed next year	Switch firm	Exit industry
Female	0.0238**	0.00632	−0.00880	0.0248	0.00269	0.0431
	(0.0103)	(0.00396)	(0.0111)	(0.0225)	(0.00560)	(0.0444)
Residual recovery	−0.00232***	-1.82×10^{-5}	0.00280***	0.00768***	0.000507	−0.00689*
	(0.000462)	(0.000201)	(0.000763)	(0.00113)	(0.000386)	(0.00384)
Female × Residual recovery	0.00196	−0.000388	0.00127	−0.00202	-8.83×10^{-5}	0.0137
	(0.00140)	(0.000568)	(0.00202)	(0.00311)	(0.00153)	(0.0113)
All FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	74,905	74,905	7,436	23,814	23,814	582
R^2	0.330	0.447	0.835	0.339	0.382	0.828

Panel B. Delayed outcomes

	Non-arbitrated complaints		Formal arbitration	
	Late re-entry	Months to re-entry	Late re-entry	Months to re-entry
Female	0.00562	0.0117	−0.00223	−0.0142
	(0.00408)	(0.0188)	(0.00702)	(0.0566)
Residual recovery	−0.000463	0.00121	0.000355	0.00545
	(0.000362)	(0.00150)	(0.00112)	(0.00869)
Female × Residual recovery	0.000980	0.00268	1.01×10^{-5}	−0.0128
	(0.00116)	(0.00495)	(0.00213)	(0.0198)
All FEs	Yes	Yes	Yes	Yes
Observations	7,436	4,148	582	384
R^2	0.374	0.355	0.664	0.700

Notes: Same specifications as Table 6, replacing observed claimant recovery with residual claimant recovery estimated within each channel's baseline claimant-recovery regression. "Residual recovery" denotes residual claimant recovery serving as the right-hand-side regressor in the employment equations. The non-arbitrated columns use the non-arbitrated residual and the arbitration columns use the formal-arbitration residual. Robust standard errors clustered at the firm level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Appendix

This appendix collects the supporting material referenced in the main text. Appendix exhibits appear in place rather than being gathered at the very end so that each table or figure sits next to the discussion that uses it.

Appendix A. Variable Definitions

Table A1 defines the outcome, treatment, firm-environment, panel, and employment variables used in the paper.

Table A1: Variable Definitions

Variable	Definition
Female advisor	Indicator equal to one for financial advisors classified as female by the first-name dictionary.
Arbitration	Indicator equal to one when a complaint appears as a formal FINRA arbitration disclosure.
Claimant recovery (R)	Log resolution-to-damages ratio, $\ln(1 + \text{Resolution}) - \ln(1 + \text{Damages})$. Lower values indicate smaller claimant recovery relative to the amount claimed. Compact table headers abbreviate this as "Recovery (R)" where column widths require.
Residual claimant recovery	Residual from the baseline claimant-recovery regression in the sample relevant to the application. The main text does not tabulate residualized re-estimates of Table 8 because the interaction coefficients are identical by the Frisch–Waugh–Lovell theorem; Table 11 instead uses channel-specific residual recovery as a right-hand-side regressor in the employment specifications.
Judgment award	Indicator equal to one when the arbitration outcome is an award or judgment rather than a settlement or other exit.
Experience	Advisor experience in years at the time of the complaint or arbitration event.
IA registration	Indicator for investment-adviser registration status in the disclosure record.
Log damages	Natural log of one plus claimed damages.
Positive damages (`Mis_Dmgs`)	Indicator equal to one when reported claimed damages exceed zero.
Complaint rate	Number of complaint disclosures in the advisor's firm-state-year cell divided by the number of advisors in that cell.
Firm female share	Share of female advisors in the advisor's firm-state-year cell.

Continued on next page

Variable	Definition
Firm size	Firm-scale measure based on log discretionary RAUM.
Female arbitrator share	Share of arbitrators on the panel coded as women.
Public arbitrator share	Share of arbitrators on the panel coded as public rather than industry or non-public.
All-public panel	Indicator equal to one for three-arbitrator cases with only public arbitrators.
Majority-public panel	Indicator equal to one when a majority of panel members are coded public.
Female chair	Indicator equal to one when the panel chair is coded female.
Arbitrator repeat-player	Panel-level mean count of prior appearances by the arbitrators on that panel in the parsed award corpus.
Claimant-attorney repeat player	Indicator equal to one when the claimant's attorney had at least one prior appearance in the parsed award corpus before that case, and zero otherwise.
Latest claimed damages	Most recent damages amount parsed from the award PDF.
Expungement request	Indicator for cases that include an expungement request.
Counterclaim	Indicator for cases that include a respondent counterclaim.
Termination	Indicator equal to one if the advisor separates after the complaint-resolution event.
Re-employed next year	Indicator equal to one if the advisor is observed anywhere in the registration data by the following year, whether at the same firm or a different one.
Switch firm	Indicator equal to one if the advisor's post-event employer differs from the pre-event employer.
Exit industry	Indicator equal to one on the separated-advisor subsample if the advisor is not observed again in the registration data within the sample window.
Late re-entry	Indicator equal to one on the separated-advisor subsample if the advisor returns after the immediate next-year horizon.
Months to re-entry	Elapsed months from separation to subsequent registration among advisors who later return.

Appendix B. Allegation-Level Gender Coefficients

Because measured gender differences in this literature can depend on how disclosure events are classified, Table B1 estimates the female coefficient separately within the major allegation categories. No single allegation family carries the paper's main by-channel pattern.

Table B1: Allegation-Specific Gender Gaps in Claimant Recovery (Arbitration)

	Unsuitability	**Misrepresentation**	**Fiduciary duty**	**Negligence**	**Supervision**
Female advisor	0.790	-0.971	0.120	-0.327	-0.327
	(2.096)	(0.955)	(0.868)	(0.848)	(1.073)
Observations	421	637	1,017	1,004	789
R^2	0.724	0.776	0.756	0.762	0.761

Notes: Each column reports the female-advisor coefficient from a separate matched-award claimant-recovery regression for the allegation family named in the column heading. All models include experience, IA registration, log damages, the positive-damages control, complaint-type fixed effects, firm-by-year fixed effects, and state-by-year fixed effects. Standard errors are clustered at the firm level.

Appendix C. Arbitrator-Side Institutional Tests

These tables audit whether observable panel-side characteristics explain the narrower matched-award results. Table C1 studies public-arbitrator share, Table C2 studies female chairs and arbitrator repeat-player status, and Table C3 studies the all-public panel rule and the post-rule period. Across all three, the female interactions are statistically insignificant.

Table C1: Public-Arbitrator Share and Claimant Recovery in Arbitration

	All matched awards	**Judgment cases**
Female advisor	0.382	-0.364
	(3.454)	(0.450)
Public arbitrator share	-3.180***	0.537**
	(1.082)	(0.252)
Female × public share	-1.705	0.479
	(3.995)	(0.475)
Experience	-0.0204	-0.00707
	(0.0203)	(0.00506)
IA registration	1.097**	1.147*
	(0.521)	(0.629)
Log damages	-1.308***	-1.070***
	(0.165)	(0.0914)
Observations	1,317	396
R^2	0.739	0.970

Notes: The dependent variable is claimant recovery (R) in the matched-award sample. Column (1) begins from 2,205 rows with panel public-share coding and uses 1,317 after fixed-effect absorption. Column (2) restricts the sample to judgment cases with panel public-share coding, moving from 958 raw rows to 396 estimation rows. All models include baseline controls, complaint-type fixed effects, firm-by-year fixed effects, and state-by-year fixed effects. Standard errors are clustered at the firm level.

Table C2: Female Chairs, Repeat-Player Arbitrators, and Claimant Recovery in Arbitration

	Female chair	**Repeat-player status**
Female advisor	0.0150	0.228
	(1.247)	(0.936)
Female chair	-1.857	
	(1.191)	
Female × female chair	2.508	
	(2.410)	
Arbitrator repeat-player		0.108
		(0.229)
Female × repeat-player		-0.489
		(0.652)
Experience	0.0229	0.000326
	(0.0276)	(0.0179)
IA registration	0.111	0.839**
	(0.486)	(0.407)
Log damages	-1.678***	-1.075***
	(0.311)	(0.140)
Observations	945	1,557
R^2	0.794	0.710

Notes: Column (1) begins from 1,757 three-arbitrator matched-award rows with chair-gender coding and uses 945 after fixed-effect absorption. Column (2) begins from 2,538 matched-award rows with arbitrator repeat-player coding and uses 1,557 after fixed-effect absorption. All models include baseline controls, complaint-type fixed effects, firm-by-year fixed effects, and state-by-year fixed effects. Standard errors are clustered at the firm level.

Table C3: The All-Public Panel Rule and Claimant Recovery in Arbitration

	First stage	Recovery: all-public	Recovery: majority-public	Recovery: post-rule
Post-2011 rule	0.339***			
	(0.0894)			
Female advisor		0.355	-4.404	-0.952
		(1.426)	(2.864)	(3.978)
All-public panel		-1.342		
		(0.874)		
Female × all-public		-1.316		
		(2.253)		
Majority-public panel			-1.672	
			(1.123)	
Female × majority-public			4.459	
			(3.153)	
Post-rule level				1.860
				(2.849)
Female × post-rule				8.982
				(6.331)
Observations	341	889	889	165
R^2	0.387	0.796	0.795	0.800

Notes: Column (1) estimates the first stage from the post-2011 all-public rule within the three-arbitrator rule-window subsample, moving from 385 raw rows to 341 after fixed-effect absorption. Columns (2) and (3) use 1,666 three-arbitrator matched-award rows with public-composition coding and 889 estimation rows after absorption. Column (4) uses 405 three-arbitrator rule-window rows and 165 estimation rows after absorption. Blank cells indicate that the regressor is not included in that specification. Baseline controls and fixed effects follow the matched-award specifications. Standard errors are clustered at the firm level.

Appendix D. Product-Type Splits

Table D1 reports product-level matched-award coefficients for annuities, REIT/LP interests, private placements, options, and structured products. The cells are extremely thin, so the estimates are descriptive only and are not used for inference.

Table D1: Product-Type Splits in Claimant Recovery (Arbitration)

	Annuities	**REIT / LP**	**Private placement**	**Options**	**Structured products**
Female advisor	Omitted	-6.113	Omitted	-2.409	-0.719
		(0.620)		(8.154)	(4.293)
Observations	34	74	22	23	40
R^2	0.932	0.974	–	0.848	0.886

Notes: Each column reports the female-advisor coefficient from a separate matched-award claimant-recovery regression within the product family named in the column heading. All models include baseline controls and the same fixed effects as the main matched-award specifications. In the annuities and private-placement columns, the female-advisor coefficient is omitted because the thin cell is absorbed by the fixed effects or yields a degenerate perfect-fit specification. The product-level cells are extremely thin throughout, so these estimates are reported descriptively and are not used for inference.

Appendix E. Expungement and Counterclaims

Table E1 adds expungement and counterclaim indicators to the matched-award claimant-recovery regressions. These case features affect award levels, but they do not explain the female-related pattern.

Table E1: Expungement, Counterclaims, and Claimant Recovery in Arbitration

	Expungement	**Counterclaim**
Female advisor	0.138	-0.132
	(1.309)	(0.705)
Expungement request	0.484	
	(0.468)	
Female × expungement	-0.443	
	(1.533)	
Counterclaim		1.398**
		(0.630)
Female × counterclaim		-0.562
		(1.722)
Observations	1,557	1,557
R^2	0.710	0.711

Notes: The dependent variable is claimant recovery (R) in the matched-award sample. Both columns begin from 2,538 matched-award rows and use 1,557 after baseline-covariate screening and fixed-effect absorption. All models include baseline controls, complaint-type fixed effects, firm-by-year fixed effects, and state-by-year fixed effects. Standard errors are clustered at the firm level.

Appendix F. Threshold Diagnostics

Table F1 reports local threshold diagnostics around the $50,000 and $100,000 procedural cutoffs. The results are mixed and do not support a persuasive RDD interpretation, which is why the paper relies on the full-sample channel evidence instead.

Table F1: Local Threshold Diagnostics Around $50,000 and $100,000

	$50k first stage	**$50k outcome**	**$100k first stage**	**$100k outcome**
Above threshold	0.0236		−0.0230	
	(0.0180)		(0.0157)	
Female advisor		2.453		−0.421
		(1.715)		(1.326)
Female × Above threshold		−1.231		0.970
		(1.559)		(1.211)
Distance from threshold	-1.45×10^{-6}	−0.00102*	-5.11×10^{-7}	0.000369**
	(1.01×10^{-6})	(0.000584)	(3.62×10^{-7})	(0.000169)
Above threshold × Distance	3.16×10^{-6}***	0.000420	1.39×10^{-6}***	−0.000171**
	(1.21×10^{-6})	(0.000257)	(5.29×10^{-7})	(7.69×10^{-5})
Observations	3,566	3,157	4,192	3,571
R^2	0.002	0.921	0.004	0.850

Notes: Columns (1) and (3) estimate local first-stage relationships between claimed-damages thresholds and advisor gender composition in the formal-arbitration sample. Columns (2) and (4) estimate local claimant-recovery regressions around the same thresholds. The diagnostics are mixed rather than clean nulls, so these specifications are best read as descriptive checks rather than as a source of identification.

Appendix G. Firm-Size Heterogeneity

Table G1 repeats the selection and non-arbitrated recovery specifications with a large-firm indicator and its interaction with advisor gender in the reduced sample for which the RAUM-based size measure is observed.

Table G1: Firm-Size Heterogeneity in Selection and Non-Arbitrated Claimant Recovery

	Selection	Non-arbitrated recovery
Female advisor	-0.000231	-0.373**
	(0.0127)	(0.161)
Female × large firm	-0.0246	0.111
	(0.0177)	(0.222)
Experience	-0.00299***	0.00421
	(0.000604)	(0.00343)
IA registration	0.221***	-0.346***
	(0.0169)	(0.0330)
Log damages	0.0422***	-0.115***
	(0.00244)	(0.0362)
Positive damages	-0.418***	-8.702***
	(0.0263)	(0.445)
Observations	99,674	74,905
R^2	0.451	0.594

Notes: Each column reports the large-firm heterogeneity specification for the outcome named in the column heading, using the paper's channel definitions. The main large-firm indicator is absorbed in the underlying specification; the coefficient of interest is the interaction between female advisor status and the large-firm indicator. Baseline controls and the same fixed effects as the corresponding main-table specifications are included throughout. Standard errors are clustered at the firm level.

Appendix H. Gender Coding and Unisex Names

Table H1 documents the handling of unisex names and shows that dropping them leaves the main non-arbitrated result close to baseline.

Table H1: GenderChecker Validation and Unisex-Name Robustness

	Male	Unisex	Diff.	N (Unisex)
Panel A. Mean claimant recovery by name bucket				
Non-arbitrated	-3.018	-3.145	-0.127	11,512
Formal arbitration	-2.591	-2.944	-0.353	3,513
Panel B. Female coefficient with and without unisex names				
	Non-arbitrated	**Formal arbitration**	**SE**	**N**
Baseline channel specification	-0.294	0.431	(0.094 / 0.555)	74,905 / 23,814
Male/Female-only sample	-0.278	0.328	(0.101 / 0.610)	63,532 / 20,397

Continued on next page

Table H1: GenderChecker Validation and Unisex-Name Robustness (continued)

	Male	Unisex	Diff.	N (Unisex)
Panel C. Unisex share in key environments				
Large firm = 0		15.38%		
Large firm = 1		13.88%		
High female share = 0		14.75%		
High female share = 1		14.00%		
High misconduct = 0		15.01%		
High misconduct = 1		14.46%		
Low experience = 0		13.67%		
Low experience = 1		16.43%		
Panel D. Classification coverage and handling				
Ground-truth validation sample		Unavailable in current data		
Dictionary unmatched share		0.00%		
Unisex share, all unique advisors		17.96%		
Unisex share, complaint unique advisors		14.42%		
Unisex share, arbitration unique advisors		13.84%		
Treatment of unisex names		Pooled into omitted group in baseline dropped in male/female-only robustness		

Notes: Panel A compares mean claimant recovery for unisex and clearly male names in the channel-specific samples. Panel B reports the baseline non-arbitrated and formal-arbitration female coefficients together with the corresponding male/female-only robustness. Panel C reports unisex-name shares across the key firm-environment margins used in the heterogeneity analysis. Panel D documents coverage and coding. GenderChecker returns male, female, and unisex labels, and no names are unmatched in the current data.

Appendix I. Award-Match Balance

Table I1 compares matched and unmatched arbitration rows on pre-resolution observables and on resolution composition. The award-PDF match is only weakly gender-selective.

Table I1: Balance of Matched and Unmatched Award-PDF Rows

Variable	Matched	Unmatched	Diff.	SMD
Panel A. Baseline observables				
Female	0.060	0.078	-0.018	-0.071
Unisex	0.147	0.138	0.010	0.027
Experience	21.417	18.975	2.442	0.246
IA registered	0.149	0.426	-0.277	-0.642
BD active	0.476	0.413	0.063	0.127
Log claimed damages	11.146	9.743	1.403	0.316
Positive claimed damages	0.904	0.817	0.087	0.253
Complaint year	2014.328	2014.546	-0.218	-0.064
Panel B. Resolution composition				
Award / judgment	0.398	0.010	0.388	1.098
Settled	0.530	0.753	-0.223	-0.479

Notes: Matched indicates that the arbitration row was linked to a FINRA award PDF. Panel A reports pre-resolution observables. Panel B reports resolution-status composition to document the award-publication margin. SMDs are matched-minus-unmatched mean differences scaled by the pooled standard deviation. Sample: 2,793 matched rows and 5,532 unmatched rows.

Appendix J. Sample Construction and Observation Counts

Table J1 collects the complaint-side and award-side sample ladders in one place. The point is not that the design changes across tables, but that different institutional variables require different intersections of award matching, parsed text, and fixed-effect support.

Table J1: Sample Construction and Observation Counts

Sample step	**N**	**Restriction or reason**
Full complaint universe	121,791	All customer-complaint disclosures in the 2010–2023 sample window.
Pooled claimant-recovery sample	99,674	Complaints with reported damages used in the pooled claimant-recovery and arbitration-entry regressions.
Non-arbitrated rows before absorption	77,220	Channel-specific subset of the pooled sample before fixed-effect absorption.
Non-arbitrated claimant-recovery regressions	74,905	Non-arbitrated rows after baseline-covariate screening and fixed-effect absorption.
Formal-arbitration rows before absorption	25,127	Channel-specific subset of the pooled sample before fixed-effect absorption.

Continued on next page

Table J1: Sample Construction and Observation Counts (continued)

Sample step	N	Restriction or reason
Formal-arbitration claimant-recovery regressions	23,814	Formal-arbitration rows after baseline-covariate screening and fixed-effect absorption.
Pooled-only residual rows	955	Complaints that survive in the pooled specification but become singleton fixed-effect cells once the sample is split by channel.
Published FINRA awards	8,253	Award cases in the archive during the sample window.
Matched advisor-level award rows	2,793	Arbitration disclosures linked to a specific FINRA award PDF.
Unique matched award cases	1,556	Distinct award case IDs represented by the matched advisor-level rows.
Readable panel-roster rows	2,205	Matched rows with a parseable panel roster and public-share coding.
Usable claimant-representation rows	1,771	Matched rows with claimant-representation text that can be coded.
Post-2010 parsed institutional window	2,538	Narrower matched-award window used by several appendix institutional tables.
Judgment and panel tables	1,750 / 1,657	Main matched-award regressions after covariate screening, plus female-arbitrator-share coding for the second column.
Claimant-dynamics tables	986 / 1,206	Representation and repeat-player columns use 986 rows; the claim-revision column uses 1,206 rows with latest-claim parsing.
Public-share tables	1,317 / 396	All matched awards with public-share coding, then the judgment-only subset.
Female-chair and repeat-player tables	945 / 1,557	Three-arbitrator chair analyses require 945 rows; repeat-player and expungement-style tests use 1,557 rows.
All-public rule tables	341 / 889 / 165	First-stage rule-window sample, three-arbitrator claimant-recovery sample, and post-rule claimant-recovery window.

Notes: Complaint-side counts are listed in the order used by the main text. Award-side counts mix advisor-level rows and unique-case counts because the institutional file is organized at the advisor-row level, so one published award can map to more than one advisor disclosure row. The point of the table is to show why effective sample sizes vary across the institutional extensions: each table requires a different intersection of successful award matching, parsed text, panel coding, and non-singleton fixed-effect support.

Appendix K. Time Trends in Gender Gaps by Channel

Figure K1 and Table K1 track the female coefficient over time in the pooled complaint sample, the non-arbitrated channel, and the formal-arbitration channel. The non-arbitrated channel becomes

more negative over time, whereas the formal-arbitration channel starts positive in the early subperiod and then turns imprecise later on.

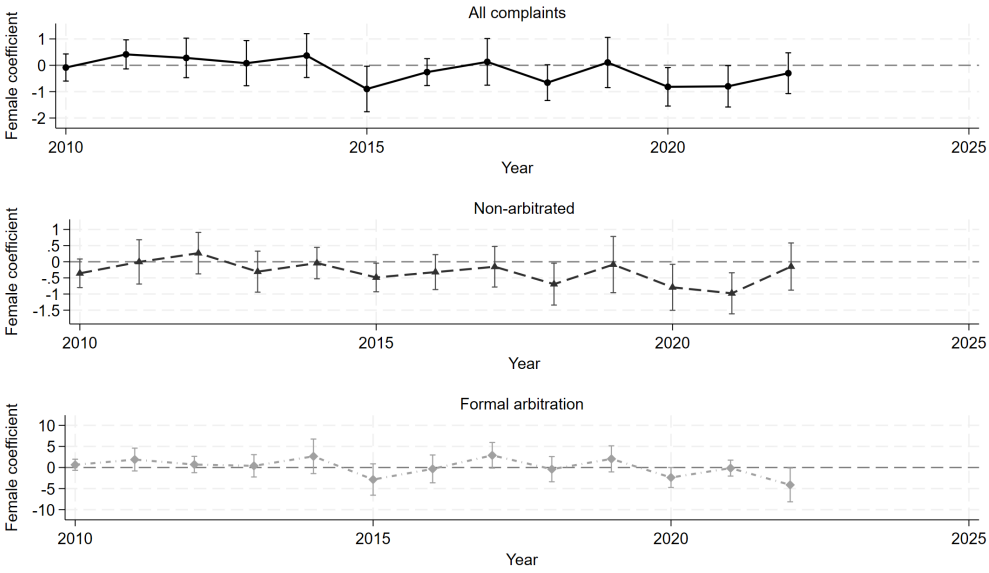


Figure K1: Time Trends in Gender Gaps by Dispute Channel

Notes: The figure plots annual female coefficients and confidence intervals by channel. The formal-arbitration series has wider confidence intervals and does not track a stable pattern relative to the pooled and non-arbitrated series.

Table K1: Time Trends in Gender Gaps by Dispute Channel

	All complaints		**Non-arbitrated**		**Formal arbitration**	
Specification	Estimate	SE	Estimate	SE	Estimate	SE
Female coefficient	0.128	0.226	-0.146	0.157	1.066	0.730
Female × year	-0.051	0.036	-0.025	0.031	-0.170	0.149
2010–2014	0.279	0.213	-0.120	0.134	1.578	0.673
2015–2019	-0.383	0.205	-0.430	0.228	0.069	0.871
2020–2022	-0.679	0.267	-0.553	0.206	-1.586	1.274
N used	99,612 / 42,912 / 17,652		74,849 / 29,899 / 14,992		23,808 / 12,793 / 2,577	

Notes: The linear-trend rows report the female main effect and the Female × year slope from the pooled 2010–2022 specification within each channel. The period rows report separate female coefficients for 2010–2014, 2015–2019, and 2020–2022. The 'N used' row reports the linear-trend sample size followed by the early-, mid-, and late-period sample sizes for each channel. The all-complaints and non-arbitrated estimates drift more negative over time. In formal arbitration, the early-period coefficient is positive and statistically significant, but the later-period coefficients are imprecise and do not trace a stable pattern.

Appendix L. Placebo Tests

Table L1 reports the fake-cutoff, rule-date, and exposure placebos. The fake-cutoff exercises are not clean nulls, so these are best read as diagnostics rather than as additional identification tests.

Table L1: Placebo and Falsification Tests

Test	Coefficient	Estimate	SE	p-value
Panel A. Fake damages cutoffs				
Female share at fake cutoff $75,000 (above)	above	0.098	0.028	0.001
Recovery jump at fake cutoff $75,000 (Female:above)	Female:above	13.851	4.671	0.004
Female share at fake cutoff $150,000 (above)	above	0.085	0.049	0.085
Recovery jump at fake cutoff $150,000 (Female:above)	Female:above	7.159	2.847	0.015
Panel B. Rule-date composition placebo				
Female share at the 2011 all-public rule cutoff (`post_rule`)	`post_rule`	-0.044	0.032	0.175
Panel C. Male-exposure placebo and reference				
Lead female hire share (`Avg_M_PenExp`)	`Avg_M_PenExp`	0.001	0.005	0.879
Lead female hire share (`Avg_F_PenExp`)	`Avg_F_PenExp`	-0.000	0.004	0.909
Lead complaint rate (`Avg_M_PenExp`)	`Avg_M_PenExp`	-0.003	0.008	0.724
Lead complaint rate (`Avg_F_PenExp`)	`Avg_F_PenExp`	-0.006	0.003	0.107
Lead advisor growth (`Avg_M_PenExp`)	`Avg_M_PenExp`	-0.034	0.037	0.356
Lead advisor growth (`Avg_F_PenExp`)	`Avg_F_PenExp`	-0.032	0.015	0.028

Notes: Panel A reports fake-cutoff diagnostics from the formal-arbitration sample. These placebo cutoffs are not clean nulls, so they should be read as cautionary diagnostics rather than as reassuring falsification checks. Panels B and C report the rule-date composition placebo and the male-exposure placebo/reference comparison from the institutional analysis.

Appendix M. Formal Cross-Channel Contrast Tests

Table M1 reports the pooled interaction specifications that formally compare the female coefficient across non-arbitrated complaints and formal arbitration. Panel A shows a parsimonious common-slopes model and a version that allows the baseline covariates to vary by channel. Panel B traces how the implied non-arbitrated, cross-channel, and formal-arbitration coefficients move as controls and fixed effects are added.

Table M1: Formal Cross-Channel Contrast Tests

Panel A. Pooled interaction specifications

	Common slopes (1)	**Channel-specific slopes** (2)
Female (non-arbitrated effect)	-0.255^{**}	-0.280^{***}
	(0.103)	(0.103)
Female × Formal arbitration	0.738	0.833^{*}
	(0.470)	(0.466)
Implied female effect in formal arbitration	0.482	0.553
	(0.462)	(0.455)
Baseline controls × Formal arbitration	No	Yes
Complaint Type FE	Yes	Yes
Firm × Year FE	Yes	Yes
State × Year FE	Yes	Yes
Observations	99,674	99,674
Firm clusters	1,331	1,331

Panel B. Coefficient path for the common-slopes pooled interaction

	Non-arbitrated effect	**Cross-channel difference**	**Implied formal-arbitration effect**
Raw	−0.243	−0.154	−0.398
	(0.202)	(0.605)	(0.573)
+ Controls	-0.238^{*}	0.653	0.415
	(0.137)	(0.479)	(0.461)
+ Complaint type FE	-0.238^{*}	0.653	0.415
	(0.137)	(0.479)	(0.461)
+ Firm × Year FE	-0.302^{***}	0.840^{*}	0.537
	(0.0885)	(0.457)	(0.451)
+ State × Year FE	-0.255^{**}	0.738	0.482
	(0.103)	(0.470)	(0.462)

Notes: Panel A reports pooled interaction regressions estimated on the full 99,674-observation claimant-recovery sample. The Female coefficient is the non-arbitrated effect. The Female × Formal arbitration coefficient is the difference relative to the non-arbitrated channel. The implied formal-arbitration effect is the sum of those two coefficients. Column (1) constrains the baseline covariate slopes to be common across channels. Column (2) interacts every baseline covariate with the formal-arbitration indicator, making it the pooled interaction specification closest to the split-sample Table 4 regressions. Panel B traces the coefficient path for the common-slopes interaction. The raw, controls, and complaint-type-FE rows use 102,347 observations and 2,427 firm clusters. The firm-by-year-FE row uses 99,682 observations and 1,331 firm clusters. The final row adds state-by-year fixed effects and uses 99,674 observations and 1,331 firm clusters. Robust standard errors clustered at the firm level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Appendix N. Reconciliation with Egan–Matvos–Seru

Table N1 compares the paper's complaint-only termination setup with the closest feasible analogue to Egan et al. (2022). The overlap-period EMS-style punishment interaction is positive and statistically significant once the advisor-year panel is rebuilt around next-year firm separation, but the corresponding complaint-only female coefficient remains negative and imprecise.

Table N1: Reconciliation with the Egan–Matvos–Seru Job-Separation Design

	Female	Complaint	Female × Complaint	SE	p-value	Obs.	Clusters
Panel A. EMS-style advisor-year punishment analogue							
2010–2015 overlap, firm × year FE	-0.0044^{***}	0.0518^{***}	0.0081	(0.0068)	0.232	4,387,314	18,461
2010–2015 overlap, firm × year × county × role FE	-0.0056^{***}	0.0479^{***}	0.0152^{**}	(0.0064)	0.018	4,387,314	18,461
2010–2022 full sample, firm × year × county × role FE	-0.0031^{**}	0.0498^{***}	0.0017	(0.0052)	0.751	10,186,862	23,599
Panel B. Complaint-only restrictions							
	Female			**SE**	**p-value**	**Obs.**	**Clusters**
2010–2015 complaint rows only	−0.0075			(0.0087)	0.392	35,419	2,169
2010–2015 complaint rows with reported damages	−0.0123			(0.0086)	0.154	30,726	1,563

Notes: This table reconciles the paper's complaint-side employment results with Egan et al. (2022). Panel A rebuilds the outcome as an advisor-year job-separation measure analogous to their design: the dependent variable equals one if the advisor is not observed at the same firm in year $t+1$. The key regressor is a misconduct proxy equal to one when the advisor receives a complaint disclosure in year t, and the coefficient of interest is *Female* × *Complaint*. The local data do not contain the full EMS disclosure universe or their detailed license matrix, so the closest feasible role cell is built from the advisor-year panel using county, firm, year, and the broad role indicators 'IA', 'BD', and 'IAStts'. Panel B shows that simply restricting the complaint-only sample to rows with reported damages does not reproduce the EMS estimand. Duplicate disclosure rows are first collapsed to advisor × firm × year × county × role cells. Standard errors are clustered at the firm level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Appendix O. Recovery Decomposition and Zero-Payment Diagnostics

Table O1 decomposes the baseline claimant-recovery index into an extensive margin for any positive recovery, a conditional-intensive margin for paid cases only, and a PPML specification in dollar levels. It also reports the symmetric resolution-winsorization check discussed in the main text.

Table O1: Recovery Decomposition and Zero-Payment Diagnostics

Outcome	Female coef.	SE	p-value	Obs.	Firms
Panel A. All complaints					
Any positive recovery	−0.010	(0.011)	0.361	99,674	1,331
Conditional log resolution	−0.086	(0.032)	0.007	46,217	935
Baseline claimant-recovery index	−0.152	(0.120)	0.203	99,674	1,331
PPML resolution dollars	−0.072	(0.162)	0.659	95,185	1,051
Panel B. Non-arbitrated complaints					
Any positive recovery	−0.024	(0.009)	0.008	74,905	1,003
Conditional log resolution	−0.109	(0.039)	0.005	32,808	659
Baseline claimant-recovery index	−0.294	(0.094)	0.002	74,905	1,003
PPML resolution dollars	−0.055	(0.171)	0.749	71,103	783
Panel C. Formal arbitration					
Any positive recovery	0.043	(0.052)	0.406	23,814	787
Conditional log resolution	0.031	(0.090)	0.733	12,603	608
Baseline claimant-recovery index	0.431	(0.555)	0.438	23,814	787
PPML resolution dollars	−0.067	(0.158)	0.670	20,905	641

Notes: This table decomposes the baseline claimant-recovery outcome into an extensive margin for any positive payment, an intensive margin for conditional log resolution among paid cases, and a PPML specification in dollar levels. The baseline claimant-recovery index is the paper's main outcome, $R = \ln(1 + \text{Resolution}) - \ln(1 + \text{Damages})$. Conditional-log-resolution rows are estimated only on disputes with positive recovery. The PPML rows use the resolution amount in levels. Across the three channels, the non-arbitrated female coefficient is negative on both the extensive and conditional-intensive margins, whereas the formal-arbitration coefficients remain imprecise. Symmetric winsorization of the resolution amount leaves the split-sample baseline estimates essentially unchanged: the non-arbitrated coefficient is −0.293 (s.e. 0.094) and the formal-arbitration coefficient is 0.438 (s.e. 0.556).

Appendix P. Award-Sample Inference Diagnostics

Table P1 reports two-way firm-and-case clustering and family-wise multiple-testing diagnostics for the main matched-award interaction family. The point is not that these interactions become stronger under alternative inference, but that they remain exploratory once the small-sample environment is treated more cautiously.

Table P1: Award-Sample Interaction Diagnostics

Interaction	One-way p	Two-way p	BH q	Bonf. p	Obs.	Firms	Cases
Claim revision × judgment	0.057	0.057	0.205	0.341	1,206	169	692
Claimant attorney repeat	0.068	0.069	0.205	0.411	986	141	548
Female chair	0.300	–	0.481	1.000	945	151	497
Repeat arbitrator	0.321	–	0.481	1.000	1,543	200	886
Judgment effect	0.668	0.670	0.670	1.000	1,557	200	892
Public-share severity	0.670	–	0.670	1.000	1,317	173	743

Notes: One-way p-values are the firm-clustered values reported in the award-side interaction family. Two-way p-values cluster simultaneously by firm and matched award case where that extension was completed. BH q-values are Benjamini–Hochberg false-discovery-rate adjustments across the six headline award interactions, and Bonf. p reports the Bonferroni-adjusted family-wise p-value. The two nominally marginal interactions in the main text remain essentially unchanged under two-way clustering, but neither survives family-wise adjustment. Wild-cluster bootstrap p-values are not reported for these absorbed-interaction specifications.

Appendix Q. Paper-Sample Entry Audit

Table Q1 documents entry into the paper's fully coded claimant-recovery sample. In the raw file, excluded complaint rows combine missing claimed damages with missing channel coding, so the restriction is broader than a clean damages-reporting filter. Female-advisor complaints enter that paper sample at a slightly higher rate than male-advisor complaints (86.0% versus 83.8%), so the restriction is not selecting only unusually strong female cases.

Table Q1: Paper-Sample Entry Audit

Statistic	Value
Total complaint rows in raw file	121,791
Paper claimant-recovery sample	102,347
Excluded raw-file rows	19,444
Mean enters paper sample	0.840
Mean enters sample, female	0.860
Mean enters sample, male	0.838

Notes: This table audits entry into the paper's fully coded claimant-recovery sample before the singleton-dropping fixed-effect absorption that yields the final pooled regression sample of 99,674 observations. In the raw complaint file, the excluded rows are exactly the observations with both missing claimed damages and blank arbitration coding. For that reason, the restriction is broader than a clean damages-reporting filter: it is best described as entry into the paper sample with fully coded complaint-channel and damages information. Female-advisor complaints enter that paper sample at a slightly higher rate than male-advisor complaints (86.0% versus 83.8%). The raw file therefore supports a descriptive sample-entry audit, but not a cleaner conditional selection test on damages reporting alone.

Appendix R. Observable Arbitration-Proneness and Common Support

Table R1 restricts attention to complaints that look observably close to the arbitration boundary based on baseline covariates and fixed effects. The goal is not to claim random channel assignment, but to show whether the non-arbitrated female coefficient survives once obviously low-arbitration complaints are set aside.

Table R1: Observable Arbitration-Proneness and Common-Support Checks

Panel A. Female coefficient in non-arbitrated complaints

	Full sample	**Overlap band**	**Top half risk**	**Top quartile risk**
Female	-0.294^{***}	-0.245^{**}	-0.304^{**}	-0.177
	(0.094)	(0.109)	(0.137)	(0.180)
p-value	0.002	0.024	0.027	0.327
Observations	74,905	48,364	29,216	12,487
Firm clusters	1,003	849	631	358
Mean predicted arbitration risk	0.218	0.257	0.347	0.405

Panel B. Pooled interaction in the common-support band

	Non-arbitrated effect	**Cross-channel difference**	**Implied arbitration effect**
Coefficient	-0.224^{*}	0.666	0.442
	(0.116)	(0.594)	(0.579)
p-value	0.053	0.263	0.445
Observations	67,842	67,842	67,842
Firm clusters	1,105	1,105	1,105
Mean predicted arbitration risk	0.273	0.273	0.273
Arbitration share	0.282	0.282	0.282

Notes: Predicted arbitration risk is constructed from a linear probability model of formal-arbitration entry on advisor experience, IA registration, log claimed damages, missing-damages status, complaint-type fixed effects, firm-by-year fixed effects, and state-by-year fixed effects, estimated without advisor gender. The overlap band is the common-support region defined by the intersection of the channel-specific 5th-to-95th percentile ranges of that predicted-risk index. The top-half-risk and top-quartile-risk columns restrict the non-arbitrated sample to increasingly arbitration-prone complaints based on the pooled predicted-risk distribution. Panel B re-estimates the pooled interaction model from Appendix Table M1 within the common-support band while allowing baseline covariates to vary by channel. The non-arbitrated female coefficient remains negative in the overlap and top-half-risk subsamples, but the cross-channel interaction remains statistically imprecise. Robust standard errors clustered at the firm level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Appendix S. Arbitration Outcome Composition and Panel Structure

These tables report the formal-arbitration diagnostics summarized in Section V: whether cases involving female advisors sort into different arbitration outcome types, whether the female coefficient shifts at the $50,000 and $100,000 procedural thresholds, whether judgment cases show a sharper pattern than the full arbitration sample, and whether panel size or arbitrator composition carries the female coefficient. None of them does.

Table S1: Outcome Type Structure Within Formal Arbitration

	Settled	**Judgment**	**Denied**	**Dismissed**
	(1)	(2)	(3)	(4)
Female	-0.00905	-0.0190	-0.0310	-0.00116
	(0.0170)	(0.0140)	(0.0492)	(0.00344)
Experience	0.000333	0.00154^{***}	-0.00404^{***}	5.78×10^{-5}
	(0.000329)	(0.000463)	(0.00121)	(0.000122)
IA registration	0.0123^{***}	-0.0582^{***}	0.175^{***}	-0.00311^{**}
	(0.00336)	(0.00794)	(0.0201)	(0.00156)
Log damages	-0.00295	0.00619^{*}	-0.0534^{***}	0.00356^{***}
	(0.00225)	(0.00344)	(0.00751)	(0.00117)
Positive damages	0.0211	-0.0297	0.517^{***}	-0.0367^{***}
	(0.0246)	(0.0426)	(0.116)	(0.0130)
All FEs	Yes	Yes	Yes	Yes
Observations	21,079	21,079	21,079	21,079
R^2	0.525	0.496	0.697	0.381

Notes: OLS estimates with outcome-type indicators as dependent variables in the formal-arbitration sample. These regressions are limited to the 21,079 arbitration rows in the outcome file with non-missing status coding, rather than the full 23,814-row arbitration claimant-recovery sample. The near-zero female coefficients do not support systematic sorting of women's and men's arbitration rows into different outcome categories once the baseline controls and fixed effects are applied. Robust standard errors clustered at the firm level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table S2: Gender Effects Across Damage Bins in Formal Arbitration

	Claimant recovery (R)
Female	0.205
	(0.851)
\$50k–\$100k bin	1.037**
	(0.483)
\$100k+ bin	1.763***
	(0.509)
Female × \$50k–\$100k	0.110
	(0.970)
Female × \$100k+	0.648
	(0.829)
Experience	0.0404***
	(0.0137)
IA registration	−1.336***
	(0.218)
Log damages	−0.510***
	(0.160)
Positive damages	−5.577***
	(1.459)
All FEs	Yes
Observations	23,814
R^2	0.739

Notes: The formal-arbitration sample is split into claimed-damages bins below \$50,000, between \$50,000 and \$100,000, and above \$100,000. The omitted category is the lowest-damages bin. Robust standard errors clustered at the firm level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table S3: Judgment Awards in the Matched-Award Sample

	Judgment specification (1)	**Judgment × female arbitrator share** (2)
Female	−0.154	−0.194
	(0.507)	(0.548)
Judgment	−9.378***	−9.168***
	(0.248)	(0.292)
Female × Judgment	−0.177	−0.165
	(0.612)	(0.662)
Female arbitrator share		0.425
		(0.509)
Judgment × Female arbitrator share		−0.671
		(0.605)
Experience	0.0265***	0.0237**
	(0.00943)	(0.0118)
IA registration	−1.197***	−1.043***
	(0.396)	(0.401)
Log damages	−1.028***	−1.007***
	(0.0297)	(0.0269)
All FEs	Yes	Yes
Observations	1,750	1,657
R^2	0.900	0.903

Notes: These regressions use the matched-award subsample where the FINRA award PDF could be linked to the advisor record. Column (1) starts from 2,793 matched advisor rows and uses 1,750 after baseline-covariate screening and fixed-effect absorption. Column (2) further requires non-missing female-arbitrator-share coding and uses 1,657 rows. Robust standard errors clustered at the firm level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table S4: Panel Composition and Claimant Recovery in Matched Awards

	Three-arbitrator panel	**Female arbitrator share**
	(1)	(2)
Female	−0.669	−0.482
	(1.194)	(1.004)
Three-arbitrator panel	−0.641	
	(0.526)	
Female × three-arbitrator panel	0.542	
	(1.663)	
Female arbitrator share		−0.352
		(0.642)
Female × female arbitrator share		0.820
		(2.853)
Experience	0.0321**	0.0182
	(0.0154)	(0.0182)
IA registration	0.275	0.440
	(0.332)	(0.381)
Log damages	−1.127***	−1.121***
	(0.0470)	(0.0520)
All FEs	Yes	Yes
Observations	1,750	1,657
R^2	0.697	0.700

Notes: These matched-award regressions test whether the advisor-gender margin varies with panel size or female representation on the panel. Column (1) starts from 2,793 matched advisor rows and uses 1,750 after baseline-covariate screening and fixed-effect absorption. Column (2) further requires non-missing female-arbitrator-share coding and uses 1,657 rows. Robust standard errors clustered at the firm level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.