WORKING PAPER | September 2026

Columbia University, Department of Economics

The Overtime Divide: The Gender Pay Gap in New York City's Municipal Workforce*

Tyler Sotomayor§ Gabriel Uceda-Sosa¶

Columbia University

September 2026

Work in progress

Abstract

Salary schedules leave substantial scope for differences in annual pay. Using New York City payroll records, we compare base salaries with realized compensation, including overtime and other pay. In the classified, active annual-salaried FY2024 sample, the male–female base-salary gap is 4.6 log points and is near zero after controlling for civil-service title. The total-compensation gap is 24 log points. Men record overtime more often than women, 60 versus 22 percent, and average roughly four times as many overtime hours. A dollar decomposition separates job composition from within-cell regular pay, overtime participation, hours, rates, and other pay. Quantile decompositions show larger compensation gaps toward the top of the pay distribution. In a FY2014–FY2025 panel, holding an overtime-rich title is associated with higher overtime within workers; the estimated increase is 38 hours smaller for women. These comparisons do not hold specific title transitions fixed or identify the effects of assignment rules. Sex is inferred from names rather than observed, and alternative classification rules cannot establish label accuracy. Statistical precision also depends on the clustering unit. The results motivate research linking payroll to validated demographic and scheduling records to distinguish job access, overtime offers, and workers' responses.

JEL Codes: J16, J31, J45, J71

Keywords: gender pay gap, Oaxaca–Blinder decomposition, overtime, public sector, civil service, unconditional quantile regression

*We are grateful to seminar participants and colleagues for helpful discussions and comments. All data used in this paper are public administrative records published on the NYC Open Data portal; replication code is available from the authors. All errors are our own. § Corresponding author: tyler.sotomayor@columbia.edu. ¶ Correspondence: gabrielucedasosa.com. Please cite as: Sotomayor, Tyler and Gabriel Uceda-Sosa (2026). "The Overtime Divide: The Gender Pay Gap in New York City's Municipal Workforce." Working Paper, Columbia University.

Contents

1 Introduction **5**

2 Related Literature **6**

Decomposition Methods 6

The Gender Gap and the Hours Margin 7

Public-Sector Pay and the NYC Workforce 7

3 Institutional Background **8**

Pay-Setting in New York City Government 8

Overtime 8

Payroll Transparency 9

4 Data **9**

Sources 9

Inferring Sex from First Names 10

Analysis Sample 11

Outcomes and Covariates 11

5 Three Facts about Municipal Pay **12**

Fact 1: Base Pay Is Nearly Sex-Neutral; Realized Pay Is Not 12

Fact 2: The Overtime Margin Is Extremely Gendered 13

Fact 3: The Gap Has Narrowed in Base, Not in Total Pay 13

6 An Accounting Framework **14**

The Pay Identity 14

Job Composition and Within-Job Differences 15

7 Empirical Methodology **15**

The Aggregate Decomposition 15

Detailed Decompositions and the Omitted-Group Problem 16

Distributional Decompositions . . . 16
Overtime Margins . . . 17
The Exact Dollar Decomposition . . . 17
Inference . . . 17
What the Decompositions Identify . . . 18

8 Results **18**
Summary Statistics . . . 18
Mean Decompositions: Base Salary . . . 19
Mean Decompositions: Total Compensation . . . 20
The Overtime Margins . . . 23
An Exact Dollar Decomposition . . . 24
How the Residual Depends on Job Granularity . . . 25
Uniformed and Civilian Services . . . 26
Across the Distribution . . . 27
Heterogeneity Across Agencies . . . 28

9 Within-Worker Overtime Patterns **29**
Panel Construction . . . 29
Worker Fixed-Effects Estimates . . . 30
Event-Time Patterns and Variance Shares . . . 31
What a Stronger Design Would Require . . . 32

10 Robustness **32**
Sex-Inference Choices . . . 32
Reference Coefficients and Estimators . . . 33
Sample Definitions . . . 33
Standard Errors and Sample Thresholds . . . 34

11 Conclusion **36**

References **37**

A The Exact Accounting Decomposition **39**

Support and Interpretation 39

B Robustness Appendix **40**

Threefold Components 40

C Empirical Appendix **41**

Sex-Inference Diagnostics 41

Occupation-Group Mapping 44

1 Introduction

A salary schedule says what a job pays. It says less about how much its holder earns over a year. New York City's civil-service payroll makes that distinction visible: the male–female gap in annual base salary is small within titles, while the gap in realized total compensation is 24 log points in our FY2024 sample. Overtime hours and other payments matter for the comparison, as do the jobs men and women hold.

The New York City Council's analysis of 2018 workforce data found that women earned about 99.6 percent of male salaries after accounting for title, agency, and demographics, and emphasized occupational segregation (New York City Council Data Team, 2020). We build on that distinction between salary differences across and within jobs by examining regular gross pay, overtime, and other compensation. Our question is how these components contribute to the difference between a posted annual salary and the earnings recorded on a payroll line.

The Citywide Payroll Data file contains 562,898 lines in FY2024, including names, agencies, civil-service titles, start dates, salary rates, hours, and pay components (New York City Office of Payroll Administration, 2026). It contains no observed sex or gender. We infer a binary sex variable from first names using New York City birth-name counts and a name dictionary. The main sample comprises 228,778 classified, active annual-salaried payroll records across 1,248 titles and 99 agencies; 54.6 percent are classified female. This selected sample does not represent every municipal worker. Classification agreement, alternative thresholds, and comparisons with unclassified records provide sensitivity checks, but no individual-level validation of the labels.

Three patterns motivate the analysis. The raw gap in log base salary is 4.6 points and falls close to zero with title fixed effects. The gap in realized compensation is much larger: 24.0 log points among employees meeting our full-year proxy. Overtime is a substantial part of that difference. In this sample, 60.3 percent of men and 22.0 percent of women record positive overtime hours; their respective averages are 220 and 56 hours. Across FY2015–FY2025, the base-salary gap narrows while the total-compensation gap remains comparatively large.

We use Oaxaca–Blinder decompositions to distinguish the contributions of observed characteristics from differences in fitted pay relationships, and recentered-influence-function regressions to study quantile gaps. A separate Shapley decomposition reconciles the dollar gap with payroll to-

tals. It reports job-cell composition and within-cell regular pay separately, along with overtime and other pay. The hourly-equivalent outcome offers another perspective, although missing or unusable hours reduce its sample, preventing a direct attribution of the difference between annual and hourly gaps to hours alone.

The FY2014–FY2025 panel adds descriptive within-worker comparisons. Holding an overtime-rich title is associated with approximately 110 more annual overtime hours for men, with an interaction of −38 hours for women. Worker fixed effects absorb time-invariant differences between people. They do not make job changes exogenous, hold the origin and destination titles fixed, or absorb common year shocks. These estimates therefore summarize changes accompanying overtime-rich employment; they do not identify an assignment effect or separate discrimination from differential offers, acceptance, and constraints on workers' time.

We interpret the decompositions as accounting exercises (Fortin et al., 2011). Their contributions depend on the covariates, the definition of a job cell, and assumptions about cells in which one group is absent. They are not estimates of how a policy would change earnings. Precision also varies: the 9.0-log-point residual in the coarse agency–occupation specification is statistically distinguishable from zero at five percent with title clustering, but not with agency-group or two-way clustering.

Bolotnyy and Emanuel (2022) show how overtime and unpaid leave produce an earnings gap among transit operators on common wage schedules. Our broader payroll setting documents related patterns across many occupations, while lacking their time-card detail. The next step is to connect these accounting patterns to the choices and constraints that generate them.

Section 2 discusses related work, and Section 3 describes pay-setting institutions. Sections 4 and 5 introduce the data and descriptive patterns. Sections 6 and 7 set out the accounting and estimation methods. Sections 8 and 9 report the cross-sectional and panel results. Section 10 examines sensitivity, and Section 11 discusses further research.

2 Related Literature

Decomposition Methods. Early contributions include R. Oaxaca (1973) and Blinder (1973); Neumark (1988) and R. L. Oaxaca and Ransom (1994) developed the reference-coefficient variants

we use, Jann (2008) standardized estimation and inference, and Yun (2005) resolved the omitted-group dependence of detailed decompositions. Distributional extensions include the reweighting approach of DiNardo et al. (1996), the conditional-quantile method of Machado and Mata (2005), and the recentered influence function regressions of Firpo et al. (2009) that we adopt; Fortin et al. (2011) survey the field and supply the interpretive framework — decompositions are accounting identities whose causal content depends on what the covariates hold fixed — that disciplines our reading of the results.

The Gender Gap and the Hours Margin. Blau and Kahn (2017) document that occupation and industry now account for more of the U.S. gender gap than human capital, and Goldin (2014) argues that the residual gap is concentrated in jobs with convex returns to long and particular hours. Cha and Weeden (2014) show that the growing wage premium to overwork slowed aggregate convergence; Cortés and Pan (2019) tie the returns to long hours to household constraints among the highly skilled; Mas and Pallais (2017) measure willingness to pay for schedule flexibility; and Bertrand et al. (2010) and Azmat and Ferrer (2017) trace how hours and career interruptions open gaps among professionals. Closest to us, Bolotnyy and Emanuel (2022) use time-card data from the Massachusetts Bay Transportation Authority to show that, in a unionized setting where the wage schedule is gender-blind by construction, the entire earnings gap among bus and train operators arises from overtime and unpaid time off. We examine related accounting patterns across a broader, selected municipal sample: across a 228,778-employee workforce spanning teachers, police officers, and engineers, rigid base schedules leave a within-title base gap of essentially zero, while the allocation of overtime regenerates a total-pay gap of twenty-four log points. Relative to their single-agency design, the citywide file trades time-card detail for comprehensiveness: we cannot observe shift-level choice sets, but we can decompose how much of the citywide divide flows through sorting across agencies and occupations versus the overtime margin within them.

Public-Sector Pay and the NYC Workforce. Gittleman and Pierce (2012) benchmark state and local government compensation; Card et al. (2016) show how firm-level wage-setting transmits to gender gaps in the private sector — a channel that civil-service schedules deliberately suppress; and Mas (2017) studies what happens when public pay becomes transparent. Closest in setting, the New York City Council's pay-equity analysis of 2018 Local Law 18 workforce data

(New York City Council Data Team, 2020) documents an unadjusted median male–female salary gap of about $21,600 that falls to roughly 99.6 percent salary parity once civil-service title, agency, and demographics are held fixed, and attributes the residual to occupational segregation. We take that within-title salary parity as a starting point rather than a finding to reproduce. Our contribution is to study the object the salary analyses set aside: the divergence between schedule-set base salary and *realized* total compensation — the overtime participation and hours margins, their distributional incidence, an exact dollar decomposition of the compensation wedge, and descriptive within-worker comparisons across overtime-rich employment states. These comparisons do not separate job sorting from overtime offers or acceptance.

3 Institutional Background

Pay-Setting in New York City Government. New York City runs the largest municipal workforce in the United States: roughly 560,000 people drew a City paycheck in fiscal year 2024, spanning police officers, firefighters, sanitation workers, teachers, social-service caseworkers, engineers, and attorneys across roughly one hundred agencies. Pay-setting in this workforce is heavily institutionalized. Employees occupy civil-service *titles* governed by New York State Civil Service Law, and base salaries attached to a title are determined by collectively bargained salary schedules — step schedules with longevity increments negotiated between the City and its municipal unions. The design of the system leaves little room for individual wage bargaining: two employees in the same title, at the same step, earn the same base salary by construction. This makes the setting an unusually clean laboratory for decomposition methods. In private-sector data, an unexplained within-job pay difference may reflect unobserved bargaining, performance pay, or firm-level wage-setting discretion (Card et al., 2016); here the base-pay schedule removes those margins almost entirely, so any total-pay divergence must arise from *assignment*: which titles men and women hold, and who works paid overtime.

Overtime. Overtime is the principal flexible margin of municipal pay. Most City employees below managerial ranks are overtime-eligible, typically at time and a half, and overtime is concentrated in the uniformed services — police, fire, correction, and sanitation — where minimum staffing rules,

emergency response, and vacancy backfilling generate large and recurring overtime budgets. Overtime hours are allocated through a mix of seniority rules, rotation lists, and supervisor discretion that varies by agency and title. As in the transit setting of Bolotnyy and Emanuel (2022), the opportunity to convert time into money is formally gender-neutral but need not be gender-neutral in practice: schedule predictability, shift timing, and family constraints shape who signs up for — and who is offered — overtime. Fiscal year 2024 City payroll records show the magnitude of this margin directly: the average full-year male employee in our sample was paid for 220 overtime hours, the average female employee for 56.

Payroll Transparency. Under the City's open-data legislation, the Office of Payroll Administration publishes an annual microdata file covering every City payroll record — name, agency, title, work borough, start date, base salary, regular gross pay, overtime hours and pay, and other pay — from fiscal year 2014 onward (New York City Office of Payroll Administration, 2026). Public-sector pay transparency of this kind has been shown to compress pay and change wage-setting behavior (Mas, 2017); here we exploit it differently, as a census of administrative pay records on which decomposition methods can be applied without survey measurement error in earnings. To our knowledge, this paper is the first academic gender-gap decomposition built on the full citywide payroll file.

4 Data

Sources. Our primary source is the Citywide Payroll Data (Fiscal Year) file published on the NYC Open Data portal (New York City Office of Payroll Administration, 2026). Each record is a payroll line — an employee–agency–fiscal-year observation, so a worker holding two concurrent appointments contributes two lines — containing first and last name, middle initial, agency, civil-service title description, agency start date, work borough, leave status as of June 30, pay basis, annual base salary rate, regular hours and regular gross pay, overtime hours and overtime pay, and "total other pay" (differentials, longevity payments, retroactive settlements, and similar items). We use fiscal year 2024 — 562,898 lines — as the main cross-section and fiscal years 2014–2025 for the trends and the worker panel of Section 9. The file carries no demographic fields, which motivates

the sex inference described next.

Inferring Sex from First Names. The file records no sex, so we infer a binary sex from first names. This is an *inference*, not observed or self-identified gender; we use "sex" for the inferred binary variable throughout and treat its error explicitly. We combine two sources. First, an empirical posterior Pr(female | name) from the City's own Popularity of Baby Names dataset (New York City Department of Health and Mental Hygiene, 2026), which reports name-by-sex birth counts for New York City cohorts; a name is classified here when it has at least 25 recorded births. Second, names not covered are classified with the `gender-guesser` dictionary, an international name–sex lexicon in wide research use. In the analysis sample, 68 percent of workers are classified by the NYC posterior and 32 percent by the dictionary. We assign female when the posterior is at least 0.90 and male when at most 0.10; ambiguous names (4.5 percent of records with a usable name) and names unknown to both sources are left unclassified. Overall 80.3 percent of FY2024 lines are classified; among active annual-salaried employees the unclassified share is 19.4 percent. Where the two sources overlap and are both confident they agree on 98.2 percent of names.

Classification Error and Coverage. Agreement between two name classifiers is not an estimate of accuracy. Both sources can make the same error, and birth cohorts and name origins in the workforce can differ from those represented in the source data. The 98.2 percent agreement rate therefore supplies neither a two-percent error bound nor a four-percent attenuation correction. We do not know whether errors are symmetric or independent of earnings and occupation.

Unclassified active annual-salaried records account for 19.4 percent of the eligible pool. They have lower mean base salary ($82,951 versus $91,449) and lower realized compensation in the full-year proxy sample, but slightly higher overtime participation (40.7 versus 39.4 percent). Their uniformed share is lower, 13.9 versus 19.5 percent (Table A2). These differences establish that coverage is selective; they do not determine the direction of bias in a sex-specific gap. Sex-specific classification rates are not observed for the unclassified records, so trimming classified outcomes cannot supply valid selection bounds from these data alone.

Alternative classification thresholds and a dictionary-only specification are sensitivity checks. Broad agency shares can be compared with patterns reported by the City Council (New York City Council Data Team, 2020), but an aggregate resemblance does not validate individual labels. A

linked sample with administrative sex or self-reported gender, appropriate to the question being studied, is needed to estimate error rates and assess their variation across groups. The near-zero within-title base-salary coefficient is not a validation test for the labels either.

Analysis Sample. We are explicit that this is not the "complete" payroll. The main analysis sample keeps FY2024 lines that are active as of June 30, paid on an annual basis, have a base salary of at least $25,000, a classifiable first name, and a valid start date: 228,778 payroll records, 54.6 percent classified female, over 1,248 titles in 99 agencies. It excludes per-diem and hourly workers (per-session teachers, poll workers, part-time staff), unclassified names, and mid-year leavers — so the conclusions apply to the retained sample. Section 10 examines selected alternative restrictions. For realized-pay outcomes we further require a full-year presence; because the half-of-base threshold used in our baseline is loose, we re-estimate the headline log-compensation decomposition under 80, 90, and 95 percent thresholds (Table 13). Education is the largest occupation group (37.9 percent), followed by policing (10.7 percent) and social services (6.0 percent); 19.5 percent of full-year employees hold uniformed titles, of whom 14.5 percent are women.

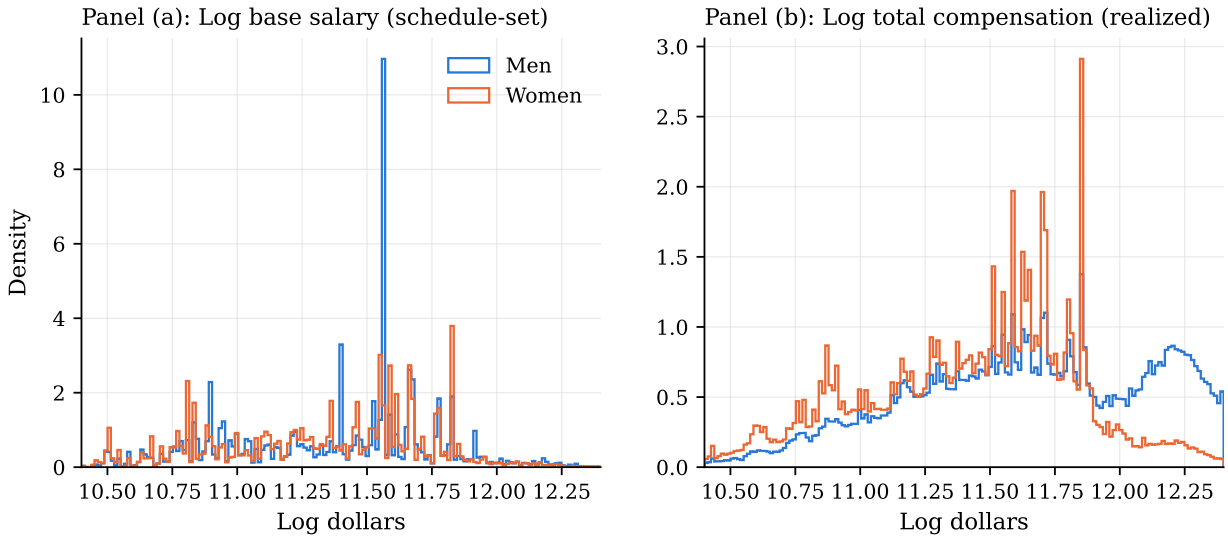
Outcomes and Covariates. Our first outcome is log annual base salary. Total compensation is the log of regular gross pay plus overtime pay plus other pay — what a worker took home. A *hourly-equivalent* outcome divides total compensation by regular plus overtime hours; it summarizes compensation per recorded hour, including overtime premiums and other pay. It is not a contractual wage rate. Only 120,394 observations have usable hours and positive hourly compensation, compared with 214,354 in the full-year annual-pay sample; differences between their gaps cannot be attributed entirely to hours. Covariates enter in nested specifications: (1) none; (2) a tenure quadratic plus borough; (3) adding agency indicators; (4) adding fourteen occupation-group indicators built from title keywords (Appendix C). Because a decomposition's unexplained share depends on how finely jobs are defined, we do not treat any single grouping as definitive: we report a granularity ladder from no controls through full 99-agency and title fixed effects (Section 8, Table 7), and read the title-fixed-effects column — identifying purely within civil-service titles — as the finest job control reported here, rather than a bound on discrimination.

5 Three Facts about Municipal Pay

Before decomposing, we document three facts that organize everything that follows.

Fact 1: Base Pay Is Nearly Sex-Neutral; Realized Pay Is Not. In fiscal year 2024 the average male employee in the analysis sample had a base salary of $93,500 against $89,745 for the average female employee — a raw log gap of 4.6 points. Realized total compensation tells a different story: $130,730 against $100,066, a raw log gap of 24.0 points, five times larger. Base salaries are prices attached to titles; realized pay adds the quantity margins — overtime and differentials. Figure 1 makes the contrast visual: the base-salary distributions of men and women overlap heavily and are punctuated by common spikes at collectively bargained schedule points, while the total-compensation distributions separate markedly in the right tail, where overtime accumulates. The spikes are not a plotting artifact: at several deciles the male and female unconditional quantiles of base salary are *identical to the dollar*, because both sexes sit on the same schedule step. Rigid schedules mechanically equalize base pay; the divergence into realized compensation is the paper's central object.

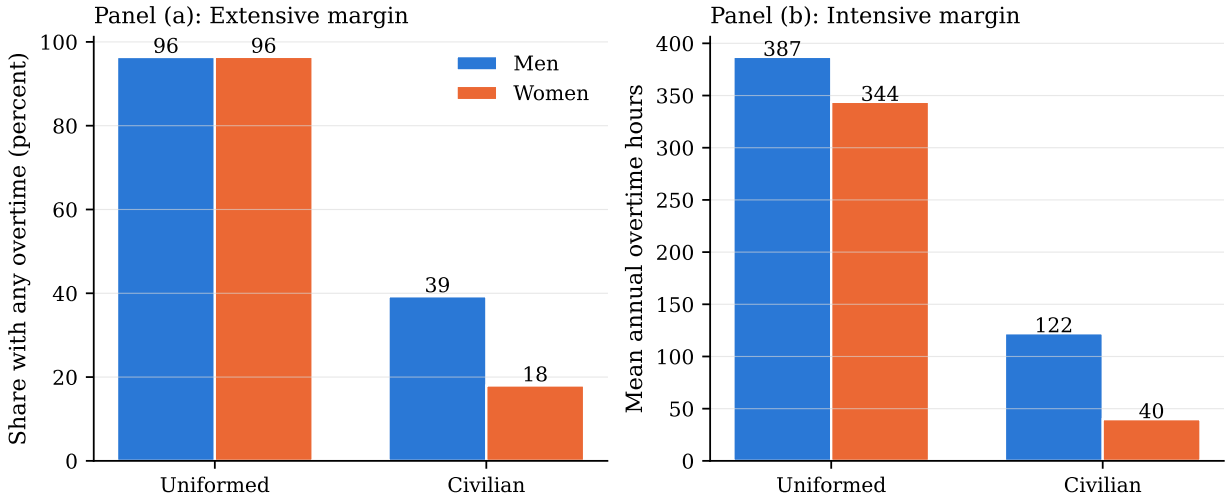
Figure 1: Base Salary Is Schedule-Set, Total Compensation Is Not



Notes: FY2024 analysis sample (Panel (a), $N = 228{,}778$) and full-year subsample (Panel (b), $N = 214{,}354$). Step-histogram densities of log base salary and log total compensation by sex. Spikes in Panel (a) are collectively bargained schedule salaries common to both genders. Source: Citywide Payroll Data (Fiscal Year), NYC Open Data.

Fact 2: The Overtime Margin Is Extremely Gendered. Among full-year employees, 60.3 percent of men worked paid overtime in FY2024 against 22.0 percent of women; the average man was paid for 220 overtime hours against 56 for the average woman, and collected $17,987 of overtime pay against $3,376 for women. Figure 2 splits both margins by uniformed status and reveals the anatomy of the divide. Inside the uniformed services, overtime is near-universal for both sexes — 96 percent of men and women alike — and the intensive-margin gap is comparatively modest (387 versus 344 hours). In the civilian workforce both margins are sharply gendered: 39 percent of men versus 18 percent of women work any overtime, and men average three times the hours. The citywide divide therefore stacks a composition effect — women are scarce in the overtime-rich uniformed services — on top of a gendered allocation of overtime among civilians. Overtime in this workforce is not an occasional perturbation: for uniformed employees it approaches a sixth of annual take-home pay.

FIGURE 2: THE OVERTIME DIVIDE, EXTENSIVE AND INTENSIVE MARGINS

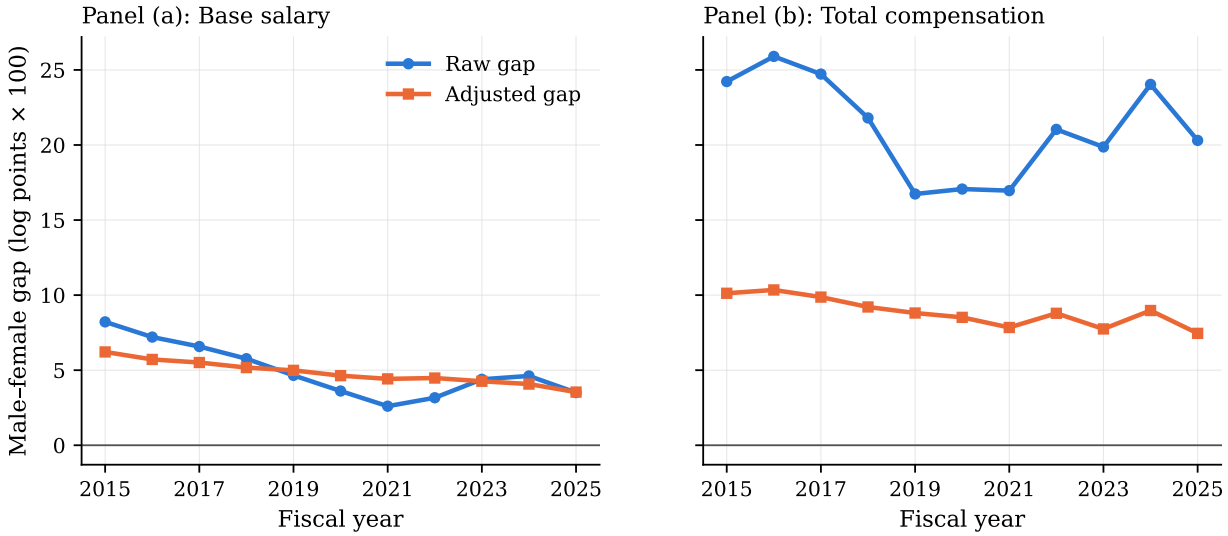


Notes: Full-year FY2024 analysis sample. Panel (a): share of employees with positive overtime hours. Panel (b): mean annual overtime hours, including zeros. "Uniformed" denotes police, fire, correction, and sanitation titles in their respective agencies. Source: Citywide Payroll Data (Fiscal Year), NYC Open Data.

Fact 3: The Gap Has Narrowed in Base, Not in Total Pay. Figure 3 plots raw and adjusted gaps for fiscal years 2015–2025. The raw base-salary gap falls from 8.2 log points in FY2015 to under 5 by FY2023 — consistent with schedule compression and compositional convergence. The total-compensation gap barely moves: 24.2 log points in FY2015 and 24.0 in FY2024. A decade

of base-pay convergence has been almost entirely offset by the stability of the overtime divide, echoing the "last-chapter" diagnosis of Goldin (2014): what remains of the gap lives in how hours are supplied and rewarded, not in posted wages.

FIGURE 3: RAW AND ADJUSTED GENDER GAPS, FY2015–FY2025



Notes: Male–female differences in mean log base salary (Panel (a)) and mean log total compensation (Panel (b)), fiscal years 2015–2025. Adjusted gaps are coefficients on a male indicator from OLS regressions controlling for a tenure quadratic, borough, agency, and occupation-group indicators, estimated year by year on the corresponding analysis sample. Source: Citywide Payroll Data (Fiscal Year), NYC Open Data.

6 An Accounting Framework

An annual salary rate and realized regular pay are different objects. The latter can vary with paid time, entry and exit dates, leave, and payroll adjustments even for workers with the same stated salary. This distinction matters when separating job composition from pay differences within jobs.

The Pay Identity. For worker i, recorded annual compensation satisfies

$$Y_i \;=\; r_i + a_i \eta_i \rho_i + d_i + z_i, \tag{1}$$

where r_i is regular gross pay, a_i indicates positive recorded overtime hours, η_i is overtime hours conditional on participation, ρ_i is overtime pay per recorded overtime hour, and d_i is other pay.

The small reconciliation term z_i retains overtime dollars on records with no positive overtime hours. For these records, $a_i = 0$ and the product is defined as zero. This accounting preserves all recorded overtime dollars without assigning an hourly rate when no positive hours are recorded.

We aggregate within title × agency × seniority-band cells. The decomposition then distinguishes the distribution of workers across cells from regular gross pay within them. Overtime participation, conditional hours, effective rates, other pay, and the reconciliation term complete the seven channels. Appendix A defines the cell moments, including an effective rate that makes the overtime product exact. The resulting dollar identity does not map term by term to a decomposition in logs.

Job Composition and Within-Job Differences. An overtime gap can reflect the jobs people hold, differences in overtime opportunities within those jobs, or differences in whether workers accept available hours. Payroll records show hours worked and payments received, without observing all the offers or constraints behind them. A within-cell contribution consequently has no unique behavioral interpretation.

Following workers across jobs removes time-invariant worker differences from a regression. It does not remove changing household circumstances, job-specific demand, or selection into particular transitions. The panel comparisons in Section 9 are descriptive. Distinguishing the underlying mechanisms requires information about offers, eligibility, seniority rules, and acceptance, of the kind available in the setting studied by Bolotnyy and Emanuel (2022).

7 Empirical Methodology

The Aggregate Decomposition. Let y_i denote a log pay outcome for employee i and let $g \in \{M, F\}$ index gender. For each group we estimate the linear pay equation below

$$y_i \;=\; X_i'\beta_g + \varepsilon_i, \qquad \mathbb{E}[\varepsilon_i \mid X_i,\, i \in g] = 0, \tag{2}$$

where X_i collects a constant, a quadratic in agency tenure, and vectors of borough, agency, and occupation-group indicators. The mean gap $\Delta \equiv \bar{y}_M - \bar{y}_F$ admits the threefold decomposition of

R. Oaxaca (1973) and Blinder (1973),

$$\Delta \;=\; \underbrace{(\bar{X}_M - \bar{X}_F)'\hat{\beta}_F}_{\text{endowments}} \;+\; \underbrace{\bar{X}_F'(\hat{\beta}_M - \hat{\beta}_F)}_{\text{coefficients}} \;+\; \underbrace{(\bar{X}_M - \bar{X}_F)'(\hat{\beta}_M - \hat{\beta}_F)}_{\text{interaction}}, \tag{3}$$

and the twofold decomposition

$$\Delta \;=\; \underbrace{(\bar{X}_M - \bar{X}_F)'\hat{\beta}^*}_{\text{explained}} \;+\; \underbrace{\bar{X}_M'(\hat{\beta}_M - \hat{\beta}^*) + \bar{X}_F'(\hat{\beta}^* - \hat{\beta}_F)}_{\text{unexplained}}, \tag{4}$$

where $\hat{\beta}^*$ is the reference coefficient vector. The specification of equation (2) also supplies our pooled reference regression that includes a gender indicator, following Jann (2008); excluding the indicator, as in Neumark (1988), changes nothing of substance (Appendix B). We report equation (4) as the headline split and the threefold components for completeness.

Detailed Decompositions and the Omitted-Group Problem. Contributions of individual covariate blocks to the explained and unexplained components are not invariant to the choice of omitted category for indicator variables (R. L. Oaxaca & Ransom, 1994). We therefore apply the normalization of Yun (2005): within every categorical block the estimated coefficients are re-expressed as deviations from the unweighted mean across all categories (including the omitted one), with the block mean absorbed into the constant. Detailed contributions reported in Section 8 are invariant to the reference category under this transformation.

Distributional Decompositions. Mean decompositions can mask heterogeneity across the pay distribution. We estimate unconditional quantile decompositions using the recentered influence function (RIF) approach of Firpo et al. (2009). For quantile τ of the group-g distribution, each observation's outcome is replaced by

$$\widehat{\mathrm{RIF}}(y_i; q_{\tau,g}) \;=\; q_{\tau,g} + \frac{\tau - \mathbf{1}\{y_i \le q_{\tau,g}\}}{\hat{f}_g(q_{\tau,g})}, \tag{5}$$

where $\hat{f}_g$ is a Gaussian kernel density estimate. Because the sample mean of the RIF equals the quantile itself, an Oaxaca–Blinder decomposition applied to group-specific RIFs decomposes the male–female quantile gap $q_{\tau,M} - q_{\tau,F}$ exactly; reference coefficients come from the pooled RIF

regression. We run the RIF analysis on total compensation only: base salaries are set on discrete schedule steps, so several unconditional deciles of base pay sit on identical mass points for men and women, and the density in the denominator of equation (5) is not well defined at those points. We therefore present the base-pay mass points descriptively in Section 5, without applying the continuous-density RIF formula to them.

Overtime Margins. Overtime enters realized pay as a participation margin and, conditional on participation, an hours-and-rate margin. A regression of overtime pay restricted to workers with positive overtime would be selected — selection into any overtime differs sharply by sex — and would not identify an intensive-margin effect. We therefore decompose overtime on the *unconditional* sample, which includes the many zeros: the participation gap via a linear-probability decomposition, overtime hours and overtime dollars including zeros, and the hourly-equivalent wage. These are separate outcome decompositions, not additive components of one regression. The dollar accounting below uses cell moments to reconcile the pay components, including observations with zero overtime hours.

The Exact Dollar Decomposition. We decompose the realized-pay dollar gap into seven Shapley contributions: job-cell composition, regular gross pay, overtime participation, conditional overtime hours, effective overtime rates, other pay, and overtime dollars recorded without positive hours. Cells are title × agency × seniority band. Appendix A defines the moments and the cross-group imputation used when a group or its overtime participants are absent. Contributions sum to the observed dollar gap. Their allocation depends on that support convention and cannot be interpreted as the earnings effect of a policy that changes one component.

Inference. We use a superpopulation interpretation of uncertainty: the observed payroll is one realization of employment and pay-setting processes. The analysis sample excludes unclassified names and several categories of employment, even though the underlying file covers the payroll broadly. Title-cluster bootstraps resample the observed classifications; they do not propagate uncertainty about whether those classifications are correct.

Headline mean specifications use 200 title-bootstrap replications and columns 1–2 use 100. The RIF results use 500 replications at the first and ninth deciles and 120 at interior deciles. Table 12

also reports record, agency-group, and approximate two-way bootstrap errors for the headline unexplained components. The total-compensation residual is 0.0898, with a title-cluster error of 0.0255 and an agency-group error of 0.0518. It is not significant at five percent under agency-group or two-way clustering. These schemes describe different dependence assumptions, so precision should be assessed alongside the point estimates.

What the Decompositions Identify. The decompositions describe differences in observed compensation and characteristics, following Fortin et al. (2011). Job controls are partly outcomes themselves (Blau & Kahn, 2017): conditioning on agency and occupation places their contribution in the explained component without establishing why men and women hold different jobs. A residual can reflect unmeasured job details, paid time, rates, or other compensation; it is not an estimate of discrimination. Worker fixed effects in Section 9 remove fixed worker differences but do not supply exogenous variation in job assignment. Name-based sex inference adds an unvalidated measurement step, discussed in Section 4.

8 Results

Summary Statistics. Table 1 reports means by gender. Men and women in the analysis sample differ modestly in base salary and tenure and enormously in overtime: the male overtime-pay mean is more than five times the female mean.

Table 1: Summary Statistics by Gender, FY2024

	Men	Women	Difference
Panel A: Pay (analysis sample)			
Base salary ($)	93,500	89,745	3,755
Total compensation ($)	130,730	100,066	30,664
Overtime pay ($)	17,987	3,376	14,611
Other pay ($)	13,257	8,001	5,256
Panel B: Overtime margins (full-year sample)			
Any overtime (share)	0.603	0.220	0.383
Overtime hours	219.8	55.5	164.3
Panel C: Covariates			
Tenure (years)	12.2	13.6	-1.4
Uniformed (share)	0.368	0.051	0.317
Education occupation (share)	0.212	0.518	-0.306
Employees (analysis sample)	103,978	124,800	
Employees (full-year sample)	96,825	117,529	

Notes: FY2024 analysis sample (active, annual-salaried, base salary $\geq$ \$25,000, sex classified, valid start date). Pay components and overtime margins are computed on the full-year subsample (regular gross pay at least half of base salary and tenure of at least one year). Source: Citywide Payroll Data (Fiscal Year), NYC Open Data.

Mean Decompositions: Base Salary. Table 2 walks the covariate ladder for log base salary. The raw gap of 4.6 log points is essentially untouched — indeed pushed *up*, to 8.7 points — by tenure and borough controls, because women in this workforce have slightly favorable observables on those dimensions. Agency controls pull the unexplained component to 6.7 points and occupation groups to 4.1. The residual is what coarse occupation cells cannot see: title composition within occupations. The title-fixed-effects estimate in Table 11 closes the ladder: within civil-service titles the base-salary gap is -0.1 log points with a standard error of 0.4 — a precisely estimated zero, consistent with schedule-set pay and with the New York City Council's finding of 99.6 percent within-title salary parity (New York City Council Data Team, 2020). The near-zero coefficient is

consistent with common salary schedules but does not validate the inferred sex labels. Section 4 and Appendix C describe the available classification diagnostics and the need for validation against observed demographics.

Table 2: Oaxaca–Blinder Decompositions of Log Base Salary

Dependent variable: log base salary				
	(1)	(2)	(3)	(4)
Total gap	0.046	0.046	0.046	0.046
	(0.068)	(0.064)	(0.066)	(0.077)
Explained	0.000	-0.041	-0.020	0.005
	(0.000)	(0.028)	(0.064)	(0.074)
Unexplained	0.046	0.087**	0.067***	0.041***
	(0.068)	(0.039)	(0.024)	(0.016)
Unexplained share	100%	189%	144%	88%
Tenure, borough	N	Y	Y	Y
Agency	N	N	Y	Y
Occupation groups	N	N	N	Y
Men	103,978	103,978	103,978	103,978
Women	124,800	124,800	124,800	124,800

Notes: Twofold decompositions of the male–female mean gap in log base salary, FY2024 analysis sample. The reference coefficient vector is from the pooled regression including a gender indicator (Jann, 2008). Cluster bootstrap standard errors (civil-service titles, 200 replications; 100 for columns 1–2) in parentheses. ***, **, and * denote significance at the 1, 5, and 10 percent levels.

Mean Decompositions: Total Compensation. Table 3 repeats the ladder for realized total compensation, where the raw gap is 24.0 log points. Tenure and borough absorb nothing. Agencies absorb 11.5 points — the male-dominated uniformed agencies are also the overtime-rich ones — and occupation groups a further 3.6. But unlike base pay, the ladder stalls: 9.0 log points (37 percent of the raw gap) remain unexplained after agency and occupation-group controls. The title-clustered standard error is small enough to reject a zero residual at five percent; agency-group and two-way errors are not (Table 12). With title fixed effects (Table 11), the estimated gap is 2.4 log points.

That remaining difference can include paid time, overtime, other pay, and regular-pay variation within titles; it does not identify differential treatment.

Table 3: Oaxaca–Blinder Decompositions of Log Total Compensation

Dependent variable: log total compensation

	(1)	(2)	(3)	(4)
Total gap	0.240***	0.240***	0.240***	0.240***
	(0.091)	(0.081)	(0.083)	(0.083)
Explained	0.000	0.004	0.115	0.151**
	(0.000)	(0.025)	(0.074)	(0.072)
Unexplained	0.240***	0.237***	0.126***	0.090***
	(0.091)	(0.062)	(0.031)	(0.028)
Unexplained share	100%	99%	52%	37%
Tenure, borough	N	Y	Y	Y
Agency	N	N	Y	Y
Occupation groups	N	N	N	Y
Men	96,825	96,825	96,825	96,825
Women	117,529	117,529	117,529	117,529

Notes: Twofold decompositions of the male–female mean gap in log total compensation (regular gross pay + overtime pay + other pay), full-year FY2024 sample. Reference coefficients from the pooled regression with a gender indicator. Cluster bootstrap standard errors (titles, 200 replications; 100 for columns 1–2) in parentheses. ***, **, and * denote significance at the 1, 5, and 10 percent levels.

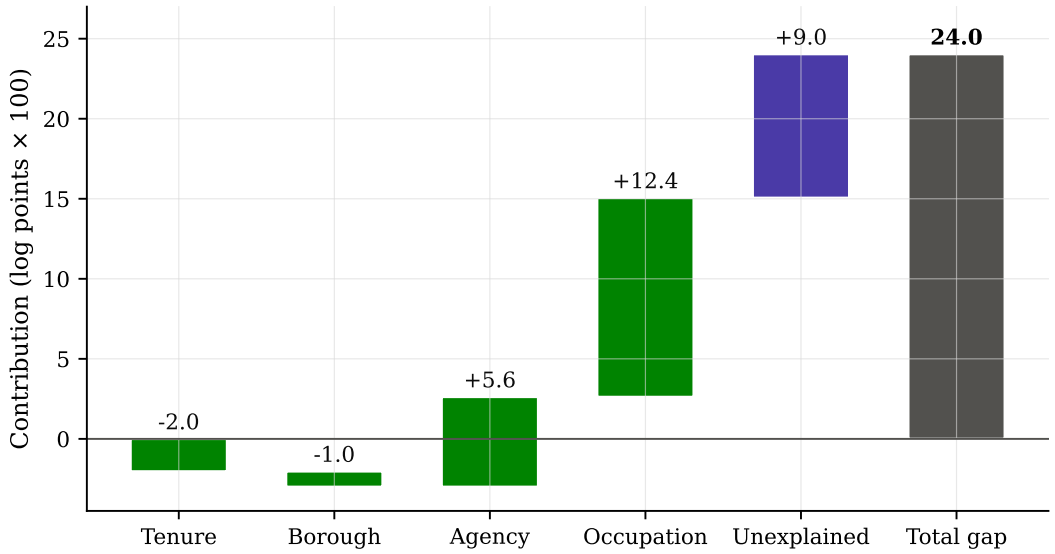
Table 4 reports Yun-normalized detailed contributions for the full specification. For total compensation, occupation sorting is the single largest explained contributor (12.4 log points), followed by agency composition (5.6); tenure works in women's favor. The detailed unexplained contributions carry the usual caveat that they depend on covariate scaling; we emphasize the block pattern rather than magnitudes. Figure 4 visualizes the explained blocks against the unexplained remainder.

Table 4: Detailed Decomposition, Full Specification

	Log base salary		Log total compensation	
	Explained	Unexplained	Explained	Unexplained
Tenure	-0.020*	0.116***	-0.020	0.171***
	(0.011)	(0.033)	(0.014)	(0.045)
Borough	-0.015***	-0.002	-0.010***	-0.004
	(0.003)	(0.013)	(0.003)	(0.014)
Agency	-0.031	0.080**	0.056	0.090**
	(0.093)	(0.036)	(0.082)	(0.038)
Occupation	0.071	-0.013	0.124**	-0.001
	(0.052)	(0.039)	(0.054)	(0.046)
Constant	0.000	-0.139***	0.000	-0.167***
	(0.000)	(0.050)	(0.000)	(0.065)
Total	0.005	0.041***	0.151**	0.090***

Notes: Detailed twofold decompositions under the normalization of Yun (2005): within each categorical block, coefficients are recentered as deviations from the unweighted category mean (including the omitted category), making block contributions invariant to the reference category. Blocks sum to the aggregate components of column (4) of Tables 2 and 3. Cluster bootstrap standard errors (titles, 200 replications) in parentheses. ***, **, and * denote significance at the 1, 5, and 10 percent levels.

FIGURE 4: DECOMPOSITION OF THE TOTAL-COMPENSATION GAP



Notes: Yun-normalized explained contributions (green) and the unexplained component (violet) for log total compensation, full specification (column 4 of Table 3); the gray bar is the raw gap. Contributions sum to the total. Units are log points × 100.

The Overtime Margins. Table 5 decomposes the overtime margins on the *unconditional* sample — including the many zeros — so the comparison includes participation differences rather than conditioning on positive overtime. Sorting across agencies and occupations explains almost all of the participation gap: of the 38.3-percentage-point gap in the probability of any overtime, 34.9 points are explained and 3.4 unexplained. It explains the bulk of the 164-hour gap in overtime hours and the corresponding dollar gap in overtime pay as well, though a smaller share of each remains within agency–occupation cells. The hourly-equivalent gap is 16.6 log points in its smaller sample of 120,394 records with usable hours, compared with 24.0 in the annual-pay sample of 214,354. Both sample composition and recorded hours affect that comparison. We cannot assign their 7.4-log-point difference entirely to hours, or treat the hourly outcome as a pure wage rate. Neither these decompositions nor the panel regressions identify the cause of the remaining overtime difference.

Table 5: Decomposing the Overtime Margins (Unconditional)

	Any OT (LPM)	OT hours	OT dollars	Log hourly comp.
Total gap	0.383***	164.3***	14611***	0.166***
	(0.059)	(29.7)	(2967)	(0.043)
Explained	0.349***	125.1***	10785***	0.087**
	(0.057)	(26.3)	(2392)	(0.041)
Unexplained	0.034***	39.2***	3825***	0.079***
	(0.011)	(10.8)	(1172)	(0.026)
Men	96,825	96,825	96,825	74,218
Women	117,529	117,529	117,529	46,176

Notes: Twofold decompositions (full specification) of four *unconditional* overtime outcomes on the full-year FY2024 sample, including workers with zero overtime: an indicator for any overtime (linear probability), annual overtime hours, overtime dollars, and the log hourly-equivalent wage (total compensation over regular plus overtime hours). Zeros are retained for overtime outcomes; the hourly outcome additionally requires usable recorded hours. Title-cluster bootstrap standard errors (100 replications for participation and hours, 80 for dollars and hourly compensation) in parentheses. ***, **, and * denote significance at the 1, 5, and 10 percent levels.

An Exact Dollar Decomposition. The log decompositions are informative but not additive in dollars. Table 6 separates the $30,664 mean compensation gap into seven contributions. Cell composition accounts for $27,478 (89.6 percent), and within-cell regular gross pay for $476 (1.6 percent). Separating these two channels matters: regular pay can differ within a title–agency–tenure cell even when salary rates follow schedules.

The overtime-hours channel contributes $1,642 (5.4 percent), effective rates $136 (0.4 percent), participation $72 (0.2 percent), and other pay $860 (2.8 percent). The reconciliation channel for overtime recorded without positive hours contributes less than one dollar within cells. Its inclusion also affects the composition calculation and ensures that the sum equals the observed gap. These allocations are conditional on the support convention in Appendix A. Cells containing both sexes account for 95.3 percent of men's records and 97.6 percent of women's records; moments for an absent group are borrowed from the observed group.

Overtime pay contributes $14,611, or 47.6 percent, of the raw difference in mean compensation.

This direct pay-component comparison answers a different question from the Shapley decomposition, which allocates part of the overtime gap to differences in job composition. The large composition contribution is descriptive. It is not an estimate of how much a policy changing occupational representation would close the gap.

Table 6: Exact Shapley Decomposition of the Dollar Gap

	Dollars	Share of gap
Job-cell composition	27,478***	89.6%
	(8,032)	
Regular gross pay within cells	476	1.6%
	(314)	
Overtime participation	72*	0.2%
	(42)	
Overtime hours	1,642***	5.4%
	(454)	
Overtime rate	136***	0.4%
	(44)	
Other pay	860***	2.8%
	(140)	
OT pay without positive recorded hours	0	0.0%
	(5)	
Total realized-pay gap	30,664	100.0%

Notes: Shapley decomposition of the mean male–female gap in realized total compensation (dollars), full-year FY2024 sample, into seven channels equalized sequentially and averaged over all orderings (Appendix A). Cells are title × agency × seniority band. Contributions sum exactly to the raw dollar gap. Cluster bootstrap standard errors (titles, 150 replications) in parentheses. ***, **, and * denote significance at the 1, 5, and 10 percent levels.

How the Residual Depends on Job Granularity. Because a decomposition's unexplained component depends on how finely jobs are defined, no single number is definitive. Table 7 traces the unexplained share of the total-compensation gap from no controls (100 percent) through agency-and-occupation groups (37 percent), full 99-agency plus occupation controls (38 percent), down to

title fixed effects (10 percent). Two lessons follow. Refining the agency partition from 25 groups to all 99 agencies barely moves the residual — occupation groups and agency-group already capture most of the between-agency variation — so the coarseness that matters is at the *title* level: only exact civil-service titles cut the residual to 2.4 log points. We report the whole ladder so the reader can see that "how much is unexplained" is inseparable from "how finely is the job defined," and read the title-fixed-effects figure as a within-title comparison, not an identified bound.

Table 7: Unexplained Share by Job Granularity

Dependent variable: log total compensation

	Unexplained	Share of raw gap
No controls	0.240	100%
Agency + occupation groups	0.090	37%
Full agency (99) + occupation	0.090	38%
Title fixed effects	0.024	10%

Notes: Unexplained component of the twofold decomposition of log total compensation, full-year FY2024 sample, as the job definition is refined. "Title fixed effects" reports the male–female within-title gap. Source: Citywide Payroll Data, NYC Open Data.

Uniformed and Civilian Services. Table 8 splits the sample. Within the uniformed services the base-salary gap is 2.8 log points and *over*-explained by tenure and borough — female officers are on average more junior — while the total-compensation gap of 7.4 points leaves 3.3 unexplained. The civilian split mirrors it at 8.5 and 5.0. The pooled 24-point gap is thus mostly a between-sector composition effect — women are under-represented in the uniformed, overtime-rich services (14.5 percent of uniformed full-year employees) — stacked on top of within-sector compensation differences. Note the distinction between uniformed *titles* and uniformed *agencies*: the NYPD's large internal gap in Figure 6 mixes officers with the department's heavily female civilian administrative workforce, which is itself a within-agency sorting margin.

Table 8: Decompositions Within Uniformed and Civilian Services

	Uniformed		Civilian	
	Base	Total comp.	Base	Total comp.
Total gap	0.028	0.074	0.020	0.085
	(0.055)	(0.065)	(0.058)	(0.063)
Explained	0.034	0.040	-0.002	0.035
	(0.034)	(0.042)	(0.057)	(0.058)
Unexplained	-0.006	0.033	0.022	0.050**
	(0.026)	(0.054)	(0.016)	(0.020)
Men	35,674	35,674	61,151	61,151
Women	6,032	6,032	111,497	111,497

Notes: Twofold decompositions within the uniformed subsample (police, fire, correction, and sanitation titles; controls: tenure quadratic and borough) and the civilian subsample (full specification), full-year FY2024 sample. Cluster bootstrap standard errors (titles, 100 replications) in parentheses. ***, **, and * denote significance at the 1, 5, and 10 percent levels.

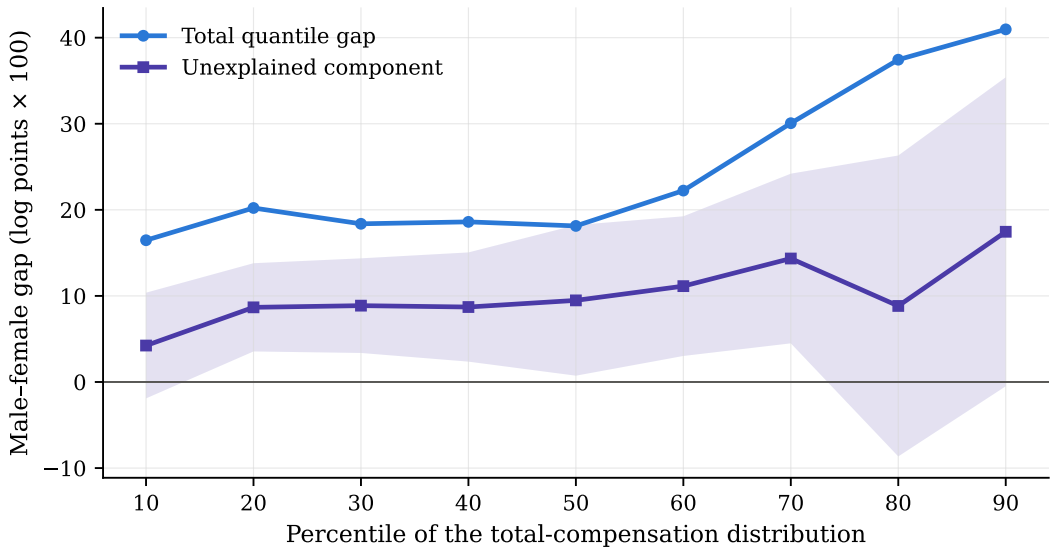
Across the Distribution. Table 9 and Figure 5 report RIF decompositions of quantile gaps in total compensation. The gap rises steeply with the quantile — about 16 log points at the first decile, 41 at the ninth — and the unexplained component is also larger at the top, from an insignificant 4 log points at the bottom to 18 at the top. The larger upper-tail gap is consistent with the concentration of overtime and other pay among high earners, but the quantile decompositions do not isolate a single mechanism. Work by Cha and Weeden (2014) and Goldin (2014) motivates examining the returns to hours and their interaction with job composition.

Table 9: RIF Decompositions of Total Compensation by Quantile

	$\tau = 0.10$	$\tau = 0.30$	$\tau = 0.50$	$\tau = 0.70$	$\tau = 0.90$
Quantile gap	0.165*	0.184	0.181	0.301**	0.410***
	(0.085)	(0.131)	(0.144)	(0.134)	(0.063)
Explained	0.122	0.095	0.086	0.157	0.235**
	(0.095)	(0.126)	(0.109)	(0.122)	(0.115)
Unexplained	0.042	0.089***	0.095**	0.143***	0.175*
	(0.033)	(0.031)	(0.048)	(0.055)	(0.092)

Notes: Unconditional quantile decompositions (Firpo et al., 2009) of the male–female gap in log total compensation at selected quantiles τ, full specification. RIFs are constructed within gender at each gender's own quantile with Gaussian kernel densities; reference coefficients from the pooled RIF regression. Title-cluster bootstrap standard errors (500 replications at the first and ninth deciles, 120 at interior deciles) in parentheses. ***, **, and * denote significance at the 1, 5, and 10 percent levels.

Figure 5: Quantile Gaps and Their Unexplained Component

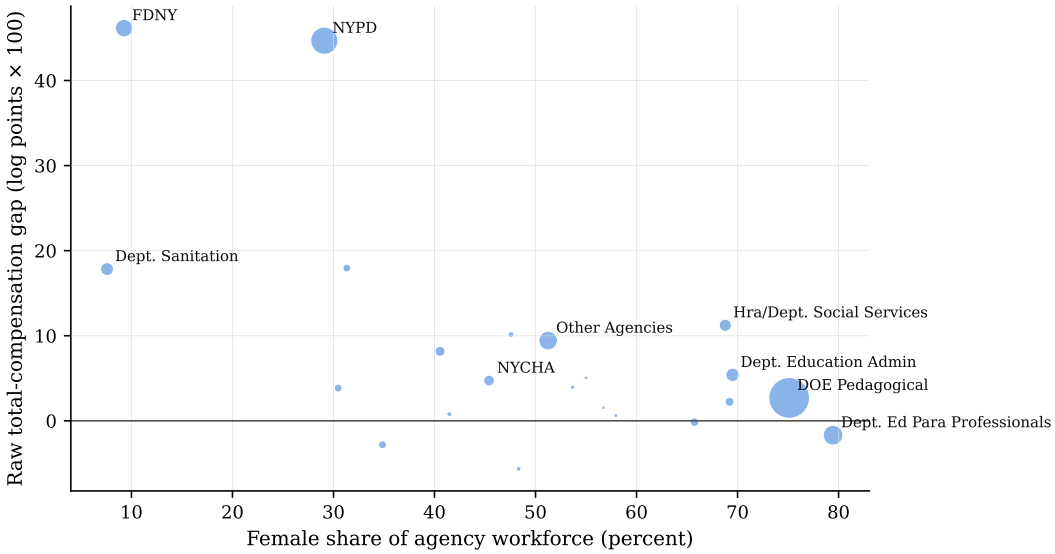


Notes: Male–female gaps in log total compensation across percentiles of the distribution (blue) and the unexplained component of the RIF decomposition at each decile (violet), with a 95 percent pointwise confidence band from the title-cluster bootstrap. Units are log points × 100.

Heterogeneity Across Agencies. Figure 6 plots each large agency's raw total-compensation gap against its female workforce share. The overtime-rich uniformed agencies sit in the upper-left —

male-dominated with large internal gaps — while the education cluster combines a large female majority with modest internal gaps. The between-agency pattern visualizes the composition effect that the decompositions attribute to agency and occupation sorting.

FIGURE 6: AGENCY FEMALE SHARE VERSUS INTERNAL PAY GAP



Notes: Each bubble is one of the 25 largest agencies (plus the residual group), full-year FY2024 sample; bubble area is proportional to headcount. The vertical axis is the agency's raw male–female gap in mean log total compensation; the horizontal axis is its female share. Agencies with fewer than 500 sampled employees are omitted.

9 Within-Worker Overtime Patterns

The panel follows changes in overtime as workers' job classifications change. Worker fixed effects remove time-invariant differences between workers. The remaining variation is still observational: promotions, staffing needs, health, family circumstances, and earnings opportunities can change alongside a title. We use the panel to describe those changes without treating them as an experiment in job assignment.

Panel Construction. The matching key combines normalized last name, first name, middle initial, and recorded agency start date. It identifies a worker spell only insofar as those fields remain stable and distinguish people. A changed start date or name can split a person's history; coincident names and dates can combine people. We do not replace the recorded date with an earliest citywide

hire date. Multiple lines within a key and year are collapsed to the record with the highest regular gross pay, summing pay and overtime hours. In the baseline, we exclude every key that ever has multiple lines in a year, not just the duplicate years. Retaining those keys is a sensitivity check in Appendix C.

The baseline hours regression contains 266,412 keys observed for at least two years and 1,804,927 person-years after restricting to classified, active, annual-salaried records and valid tenure. A title is overtime-rich when its mean annual overtime in the pooled panel is at least 100 hours. That mean includes the worker's own observations and subsequent years; it is a sample-defined classification, not a predetermined treatment.

Worker Fixed-Effects Estimates. We estimate

$$\mathrm{OT}_{it} = \alpha_i + \beta \mathrm{Rich}_{j(i,t)} + \gamma(\mathrm{Rich}_{j(i,t)} \times \mathrm{Female}_i) + X'_{it}\delta + \varepsilon_{it}, \tag{6}$$

where X_{it} contains a tenure quadratic. Standard errors are clustered by worker key. The specification includes neither year fixed effects nor origin–destination title-pair controls. The interaction is estimated from changes in broad overtime-rich status; it does not compare women and men making an identical title transition. The sex main effect is absorbed by α_i, so the regression does not estimate a within-worker sex pay gap.

Table 10 reports an association of 109.7 additional annual overtime hours with overtime-rich employment for men (standard error 4.2). The female interaction is −38.3 hours (standard error 6.1). Corresponding associations are \$7,247 in overtime pay with a $-\$2,083$ interaction, and 29.7 percentage points in participation with a −9.8-point interaction. These describe different within-worker patterns for the two groups across the transitions they actually make. Differences in their origin and destination titles, timing, and changing circumstances remain possible explanations.

The hourly-equivalent coefficient is −0.050 and the female interaction is 0.030. These are changes associated with overtime-rich status, not a zero sex gap in hourly pay. Total compensation per recorded hour also combines regular rates, overtime premiums, other payments, and measurement of hours. The regression cannot establish that the annual compensation divide is entirely a quantity effect.

Table 10: Worker Fixed-Effects Associations with Overtime-Rich Employment

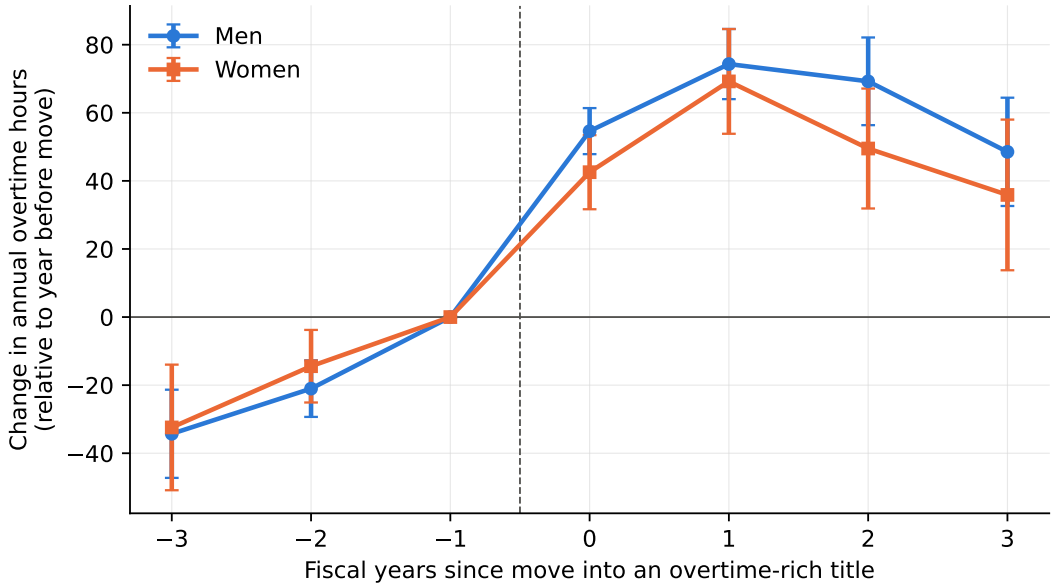
	OT hours	OT pay ($)	Any OT	Log hourly comp.
Holding an OT-rich title	109.7***	7,247***	0.297***	-0.050***
	(4.2)	(300)	(0.008)	(0.003)
× Female	-38.3***	-2,083***	-0.098***	0.030***
	(6.1)	(388)	(0.013)	(0.005)
Worker fixed effects	Y	Y	Y	Y
Tenure quadratic	Y	Y	Y	Y
Workers	266,412	266,412	266,412	229,556
Person-years	1,804,927	1,804,927	1,804,927	1,568,564

Notes: Estimates of equation (6), FY2014–FY2025. OT-rich denotes a title with pooled-panel mean annual overtime of at least 100 hours. Each column includes worker fixed effects and a tenure quadratic, but no year effects or origin–destination title-pair controls. Samples vary with outcome availability. Worker-clustered standard errors in parentheses. ***, **, and * denote significance at the 1, 5, and 10 percent levels.

Event-Time Patterns and Variance Shares. Figure 7 plots mean changes in overtime around a first observed transition into an overtime-rich title, relative to the preceding year. Both groups' overtime rises before the transition. The series are unadjusted averages; sample composition varies by event year, no formal parallel-trends test is reported, and the groups need not make the same transitions. Men's increase is larger in the plotted post-move years, but the difference varies rather than widening steadily. The figure supports a descriptive pattern, not a causal event-study interpretation.

Worker fixed effects alone absorb 71.8 percent of overtime-hour variance; title effects alone absorb 45.3 percent; jointly they absorb 72.7 percent. These are overlapping measures of fit. Their shares are not additive and do not identify how much overtime is caused by worker traits or jobs.

FIGURE 7: OVERTIME AROUND A MOVE INTO AN OVERTIME-RICH TITLE



Notes: Unadjusted mean change in annual overtime hours relative to the fiscal year before the first observed transition into an overtime-rich title, by inferred sex. The pool contains 2,591 worker keys with a transition and a valid preceding-year observation; observations available at other event years vary. Bars are 95 percent confidence intervals for the plotted means. The dashed line marks the transition.

What a Stronger Design Would Require. Comparisons within the same origin–destination title pair, with common year shocks accounted for, would narrow the sources of heterogeneity. Defining richness using earlier years or a separate sample would avoid classifying jobs partly from the outcomes being studied. Neither step alone makes moves exogenous. A credible assignment design would need a rule change, threshold, or other source of variation, together with evidence about selection and pre-existing trends. Scheduling records could then distinguish eligibility, offers, and acceptance instead of inferring those margins from paid hours.

10 Robustness

Table 11 collects the main sensitivity checks; the sex-inference diagnostics are in Appendix C.

Sex-Inference Choices. The headline decompositions change little when the posterior threshold rises from 0.90 to 0.99 or when only the dictionary is used. These checks show stability across the specified rules. They do not establish classification accuracy: both sources may make correlated

errors, and changing a threshold also changes the sample. No validated sex-specific selection rates are available for unclassified records, so we do not present trimming exercises as selection bounds.

Reference Coefficients and Estimators. Replacing the pooled-with-indicator reference vector by the Neumark (1988) pooled regression without the indicator moves the explained/unexplained split by well under one log point. The threefold components reported alongside the twofold results in Section 8 bracket the twofold split from both sides. Our implementation reproduces the `statsmodels` Oaxaca–Blinder three-fold components to numerical precision on the shared specification.

Sample Definitions. Including part-year annual employees raises the total-compensation gap mechanically (part-year women are over-represented among new hires) but moves the unexplained component little. Extending to hourly employees — for whom the pay "schedule" is an hourly rate — yields a small raw wage-rate gap, consistent with schedule-set pay. Because our baseline full-year threshold (regular gross at least half of base salary) is loose, Table 13 re-estimates the realized-pay decomposition under 80, 90, and 95 percent thresholds; the raw gap ranges from 0.237 to 0.252 and the unexplained component from 0.086 to 0.090 across the reported thresholds. Finally, the title-fixed-effects estimates bracket the exercise: a zero within-title base gap and a 2.4-log-point within-title total-compensation gap.

Table 11: Robustness of the Headline Decompositions

	Log base salary		Log total compensation	
	Gap	Unexplained	Gap	Unexplained
Baseline	0.046	0.041***	0.240***	0.090***
	(0.077)	(0.016)	(0.083)	(0.028)
Posterior threshold 0.99	0.045	0.034**	0.253***	0.080***
	(0.073)	(0.016)	(0.075)	(0.021)
Dictionary imputation only	0.048	0.041***	0.241**	0.090***
	(0.064)	(0.014)	(0.095)	(0.027)
Neumark pooled reference	0.046	0.030	0.240	0.075
Including part-year employees	–	–	0.207**	0.085***
			(0.083)	(0.026)
Hourly workers (log wage rate)	0.030	0.012	–	–
	(0.078)	(0.014)		
Title fixed effects (male coef.)	-0.001	–	0.024***	–
	(0.004)		(0.007)	

Notes: Each row re-estimates the full-specification twofold decomposition (column 4 of Tables 2 and 3) under the indicated variation. "Title fixed effects" reports the male–female within-title gap (male coefficient from a regression absorbing title effects and a tenure quadratic), with title-clustered standard errors. Cluster bootstrap standard errors (titles, 100 replications) in parentheses elsewhere. ***, **, and * denote significance at the 1, 5, and 10 percent levels.

Standard Errors and Sample Thresholds. We report four bootstrap schemes in Table 12: records, titles, agency groups, and the approximate two-way combination. The agency grouping contains 22 clusters in the full-year sample. The 0.090 total-compensation residual is more than three title-cluster standard errors from zero but only about 1.7 agency-group standard errors from zero. It is not significant at five percent with agency-group or two-way clustering; a normal approximation gives an agency-group two-sided p-value of roughly 0.08. With few agency groups, that approximation warrants caution. The standard errors also condition on fixed inferred sex labels and do not measure classification uncertainty. Table 13 reports the sample-threshold comparisons.

Table 12: Inference Under Alternative Clustering

	Log base salary	Log total comp.
Unexplained (point)	0.0408	0.0898
Standard error under:		
Record bootstrap	(0.0015)	(0.0015)
Clustered on title	(0.0162)	(0.0255)
Clustered on agency group	(0.0254)	(0.0518)
Two-way (title, agency group)	(0.0241)	(0.0510)

Notes: Unexplained component of the full-specification decomposition and its standard error under four inference schemes. Bootstrap standard errors, 80 replications per resampling scheme. Agency clusters use the 22 agency groups in the main specification. The two-way variance is approximated by title variance plus agency-group variance minus their intersection variance. The observation is a payroll line; the cross-section is not collapsed using the panel's primary-line rule.

Table 13: Realized-Pay Gap Under Full-Year Thresholds

	N	Raw gap	Unexplained
Regular gross $\geq$ 50% of base	214,354	0.240	0.090
Regular gross $\geq$ 80% of base	207,757	0.237	0.087
Regular gross $\geq$ 90% of base	195,326	0.239	0.086
Regular gross $\geq$ 95% of base	174,003	0.252	0.087

Notes: Twofold decomposition of log total compensation, full specification, as the full-year presence threshold (regular gross pay as a fraction of stated base salary) is tightened. Source: Citywide Payroll Data, NYC Open Data.

11 Conclusion

The small within-title base-salary gap in New York City's payroll coexists with a much larger gap in annual compensation. In our FY2024 full-year proxy sample, men earn about $30,664 more than women on average; overtime pay accounts for roughly $14,611 of that raw difference. The accounting exercise separates the contribution of job composition from realized regular pay and overtime within cells. It does not establish how much a change in hiring or scheduling rules would close the gap.

The panel adds evidence that overtime changes with workers' job histories, with smaller increases associated with overtime-rich employment for women. Those associations leave open differences in specific job transitions, economic conditions, and workers' circumstances. The precision of the cross-sectional residual also depends on the clustering unit. Name-based classification remains a separate limitation that agreement between two classifiers cannot resolve.

The next stage has two parts. First, a linked demographic sample would allow us to assess classification and coverage errors by occupation and name origin. Second, payroll linked to scheduling records would reveal who was eligible for overtime, who received an offer, how much notice they received, and who accepted. We would study a documented change in scheduling or eligibility rules, define the affected jobs before the change, and compare outcomes with suitable unaffected jobs while examining prior trends.

That approach would distinguish access to paid hours from the ability to accept them. It could also measure whether a rule improves workers' control over their time. As Bolotnyy and Emanuel (2022) emphasize in their transit setting, a smaller earnings gap need not imply greater worker welfare if it comes at the cost of valued flexibility. The payroll patterns make both earnings and control over hours worth studying together.

References

Azmat, G., & Ferrer, R. (2017). Gender gaps in performance: Evidence from young lawyers. *Journal of Political Economy*, *125*(5), 1306–1355.

Bertrand, M., Goldin, C., & Katz, L. F. (2010). Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American Economic Journal: Applied Economics*, *2*(3), 228–255.

Blau, F. D., & Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, *55*(3), 789–865.

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources*, *8*(4), 436–455.

Bolotnyy, V., & Emanuel, N. (2022). Why do women earn less than men? Evidence from bus and train operators. *Journal of Labor Economics*, *40*(2), 283–323.

Card, D., Cardoso, A. R., & Kline, P. (2016). Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women. *Quarterly Journal of Economics*, *131*(2), 633–686.

Cha, Y., & Weeden, K. A. (2014). Overwork and the slow convergence in the gender gap in wages. *American Sociological Review*, *79*(3), 457–484.

Cortés, P., & Pan, J. (2019). When time binds: Substitutes for household production, returns to working long hours, and the gender wage gap among the highly skilled. *Journal of Labor Economics*, *37*(2), 351–398.

DiNardo, J., Fortin, N. M., & Lemieux, T. (1996). Labor market institutions and the distribution of wages, 1973–1992: A semiparametric approach. *Econometrica*, *64*(5), 1001–1044.

Firpo, S., Fortin, N. M., & Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica*, *77*(3), 953–973.

Fortin, N., Lemieux, T., & Firpo, S. (2011). Decomposition methods in economics. In O. Ashenfelter & D. Card (Eds.), *Handbook of labor economics* (pp. 1–102, Vol. 4A). Elsevier.

Gittleman, M., & Pierce, B. (2012). Compensation for state and local government workers. *Journal of Economic Perspectives*, *26*(1), 217–242.

Goldin, C. (2014). A grand gender convergence: Its last chapter. *American Economic Review*, *104*(4), 1091–1119.

Jann, B. (2008). The Blinder–Oaxaca decomposition for linear regression models. *The Stata Journal*, *8*(4), 453–479.

Machado, J. A. F., & Mata, J. (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics*, *20*(4), 445–465.

Mas, A. (2017). Does transparency lead to pay compression? *Journal of Political Economy*, *125*(5), 1683–1721.

Mas, A., & Pallais, A. (2017). Valuing alternative work arrangements. *American Economic Review*, *107*(12), 3722–3759.

Neumark, D. (1988). Employers' discriminatory behavior and the estimation of wage discrimination. *Journal of Human Resources*, *23*(3), 279–295.

New York City Council Data Team. (2020). Understanding pay equity in the municipal workforce [Analysis of 2018 Local Law 18 workforce data; accessed July 2026].

New York City Department of Health and Mental Hygiene. (2026). Popularity of baby names [Accessed July 2026].

New York City Office of Payroll Administration. (2026). Citywide payroll data (fiscal year) [Accessed July 2026].

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, *14*(3), 693–709.

Oaxaca, R. L., & Ransom, M. R. (1994). On discrimination and the decomposition of wage differentials. *Journal of Econometrics*, *61*(1), 5–21.

Yun, M.-S. (2005). A simple solution to the identification problem in detailed wage decompositions. *Economic Inquiry*, *43*(4), 766–772.

A The Exact Accounting Decomposition

Aggregating equation (1), group workers into cells c defined by title, agency, and tenure bands of at most five years, over five through fifteen years, and over fifteen years. For inferred sex g, mean compensation can be written as

$$\bar{Y}_g = \sum_c \pi_{gc}[r_{gc} + a_{gc}\eta_{gc}\rho_{gc} + d_{gc} + z_{gc}].$$

Here π_{gc} is the group's cell share; r_{gc} and d_{gc} are cell means of regular gross and other pay; a_{gc} is the share recording positive overtime hours; and η_{gc} is mean hours among those participants. Define ρ_{gc} as their mean overtime pay divided by their mean overtime hours. It is a ratio of means, not a mean of individual rates. Finally, z_{gc} is mean overtime pay on records without positive overtime hours, including zeros for other workers. These definitions reproduce the group means exactly, including payroll records whose overtime dollars and hours do not align.

The seven channels are $\pi, r, a, \eta, \rho, d, z$. Job composition π is separate from regular gross pay r; fixed salary schedules do not justify holding realized regular pay constant within a cell. Let $v(S)$ be mean pay using men's moments for channels in set S and women's for all others. For a channel k among the set K of seven channels, its Shapley contribution is

$$\phi_k = \sum_{S \subseteq K \setminus \{k\}} \frac{|S|!(|K| - |S| - 1)!}{|K|!}[v(S \cup \{k\}) - v(S)].$$

This equals the average contribution across all 7! replacement orderings and gives $\sum_k \phi_k = \bar{Y}_M - \bar{Y}_F$. We evaluate the equivalent coalition formula and check reconciliation against the observed means in both the original sample and every bootstrap replication.

Support and Interpretation. When a sex is absent from a cell, its cell share is zero and its unobserved within-cell moments are set to the other sex's observed moments. If a sex is present but has no overtime participants, its undefined conditional hours and rate are likewise borrowed from the other sex. When neither group participates, both conditional moments are set to zero. Observed group means are unaffected, but counterfactual means and channel contributions depend on this convention. The convention attributes differences associated with single-sex cells to composition

rather than extrapolated within-cell pay differences. It should not be read as evidence that the missing group's potential pay equals the observed group's pay.

We resample titles with replacement 150 times, rebuilding cells and applying the same support convention in each replication. The standard errors are conditional on the name classifications and the support convention. As with other decomposition methods (Fortin et al., 2011), the arithmetic is exact but its components are not identified policy effects: changing who holds a job may also change hours, staffing, and pay.

B Robustness Appendix

Threefold Components. Table A1 reports the threefold decomposition — endowments, coefficients, and interaction, from the female coefficient vector's perspective — for every specification of the covariate ladder and both headline outcomes. The coefficients component, the threefold analogue of the unexplained part, tracks the twofold results in Section 8 closely; the interaction terms are sizable only where male and female coefficient vectors differ most (agency blocks), which is why we prefer the pooled-reference twofold split as the headline. Replacing the pooled-with-indicator reference by the Neumark (1988) variant is examined in Table 11.

Table A1: Threefold Decompositions, All Specifications

	(1)	(2)	(3)	(4)
Panel A: Log base salary				
Endowments	0.000	-0.049	-0.055	-0.023
	(0.000)	(0.035)	(0.063)	(0.071)
Coefficients	0.046	0.082*	0.045*	0.025
	(0.068)	(0.044)	(0.024)	(0.015)
Interaction	0.000	0.013	0.057***	0.044***
	(0.000)	(0.018)	(0.019)	(0.015)
Panel B: Log total compensation				
Endowments	0.000	-0.017	0.035	0.092
	(0.000)	(0.031)	(0.069)	(0.072)
Coefficients	0.240***	0.223***	0.080***	0.056**
	(0.091)	(0.067)	(0.028)	(0.025)
Interaction	0.000	0.035*	0.125***	0.093***
	(0.000)	(0.020)	(0.028)	(0.025)

Notes: Threefold Oaxaca–Blinder components (eq. 3): endowments $(\bar{X}_M - \bar{X}_F)'\beta_F$, coefficients $\bar{X}_F'(\beta_M - \beta_F)$, and interaction $(\bar{X}_M - \bar{X}_F)'(\beta_M - \beta_F)$. Columns are the covariate ladder of Tables 2 and 3. Cluster bootstrap standard errors (titles) in parentheses. ***, **, and * denote significance at the 1, 5, and 10 percent levels.

C Empirical Appendix

Sex-Inference Diagnostics. This appendix collects the diagnostics for the first-name sex inference introduced in Section 4.

Outcomes of Unclassified Workers. We compare records with and without classified names in Table A2. The sample is active annual-salaried employees. Unclassified records account for 19.4 percent of this eligible pool. They have lower mean base salary and total compensation, slightly higher overtime participation, and a lower uniformed share. Their sex-specific outcomes are unobserved, so these comparisons cannot establish whether exclusion increases or reduces a

measured male–female gap.

Table A2: Outcomes of Classified and Unclassified Workers

	Classified	Unclassified
Base salary ($)	91,449	82,951
Total compensation ($)	111,630	98,064
Any overtime (share)	0.394	0.407
Uniformed (share)	0.195	0.139
Employees	228,834	55,185

Notes: Active annual-salaried FY2024 employees with base salary of at least $25,000. "Unclassified" are records whose first name is not assigned a sex at the 0.90 posterior threshold. Base salary uses all eligible records; total compensation, overtime, and uniformed shares use the regular-gross-at-least-half-base subset. Source: Citywide Payroll Data, NYC Open Data.

Benchmarking Inferred Sex Shares. Table A3 reports inferred female shares by agency. The reported shares vary substantially across agencies: 9.5 percent in the Fire Department, 7.7 percent in Sanitation, 29.2 percent in the Police Department, and 75.0 percent in the pedagogical education agency. Agency shares mix uniformed and civilian titles and should not be equated with shares in uniformed occupations. These broad patterns are a contextual comparison with the Council's administrative analysis (New York City Council Data Team, 2020), not an individual-level validation sample.

Table A3: Inferred Female Share by Agency

Agency	*N*	Inferred female share
Dept Of Ed Pedagogical	78,885	0.750
Police Department	36,512	0.292
Other Agencies	18,493	0.515
Dept Of Ed Para Professionals	18,350	0.792
Fire Department	15,175	0.095
Department Of Education Admin	9,129	0.698
Department Of Sanitation	8,245	0.077
Hra/Dept Of Social Services	7,748	0.679
Nyc Housing Authority	6,030	0.450
Department Of Correction	5,012	0.406
Admin For Children'S Svcs	4,456	0.686
Dept Of Health/Mental Hygiene	3,888	0.657
Dept Of Parks & Recreation	3,149	0.352
Dept Of Environment Protection	3,140	0.309

Notes: Inferred female share among main-sample employees, largest agencies. Source: Citywide Payroll Data, NYC Open Data.

The Worker Panel and Ambiguous Matches. The baseline excludes all years of a worker key that has multiple payroll lines in any year. Retaining those keys produces 393,460 keys in the hours regression. The overtime-rich coefficient is 107.4 hours (standard error 4.0), compared with 109.7 in the baseline; the female interaction is −44.6 hours (standard error 5.8), compared with −38.3. The descriptive pattern survives this change, although neither sample resolves all identity-linkage errors or selection into job changes.

Classification Rates and Error. Table A4 reports classification rates by occupation group. The 98.2 percent agreement between confident name classifiers measures agreement rather than truth: shared errors are possible. The threshold and dictionary checks in Table 11 assess sensitivity to those classification rules. They provide no measured error rate, attenuation correction, or worst-case selection bound. Validating labels and coverage requires demographic information beyond

names.

Table A4: Name Classification Rates by Occupation Group

Occupation group	*N*	Match rate	Female share (matched)
Education	104,919	0.827	0.745
Other	69,801	0.772	0.539
Police	30,048	0.815	0.257
Social Services	18,799	0.734	0.677
Fire	8,805	0.929	0.023
Management	8,093	0.812	0.452
Health	7,023	0.819	0.378
Sanitation	6,901	0.918	0.038
Analyst/Finance	6,303	0.767	0.545
Correction	5,247	0.728	0.381
Legal	5,020	0.876	0.549
Engineering	4,559	0.761	0.175
IT	3,695	0.744	0.269
Clerical	3,480	0.750	0.756
Trades/Labor	1,326	0.868	0.067

Notes: Active annual-salaried FY2024 employees with base salary of at least $25,000. "Match rate" is the share of records assigned a binary sex label at the 0.90 posterior threshold; "female share" is computed on assigned records.

Occupation-Group Mapping. Occupation groups are assigned by the first matching keyword rule applied to the civil-service title description, in the priority order listed in Table A5. Rules are applied to upper-cased titles; employees matching no rule are assigned to "Other." The mapping and all processing code ship with the replication package.

Table A5: Keyword Rules Defining Occupation Groups

Group	Title keywords (regular expression fragments)
Police	POLICE, DETECTIVE, SERGEANT
Fire	FIREFIGHTER, FIRE MARSHAL, FIRE PROTECTION, PILOT
Correction	CORRECTION
Sanitation	SANITATION
Education	TEACHER, PEDAGOG, EDUCATION, PRINCIPAL, SCHOOL
Health	NURSE, MEDICAL, HEALTH, CLINIC, PSYCH, PHYSICIAN
Legal	ATTORNEY, LAW CLERK, LEGAL, COUNSEL, PARALEGAL
Engineering	ENGINEER, ARCHITECT, SURVEYOR, INSPECTOR
IT	COMPUTER, SYSTEMS, PROGRAMMER, NETWORK
Analyst/Finance	ANALYST, ACCOUNT, AUDITOR, ECONOMIST, BUDGET
Clerical	CLERK, SECRETARY, TYPIST, CASHIER
Social Services	SOCIAL, CASEWORKER, CHILD PROTECT, ELIGIBILITY
Management	COMMISSIONER, DIRECTOR, MANAGER, CHIEF, EXECUTIVE
Trades/Labor	MECHANIC, ELECTRICIAN, PLUMBER, CARPENTER, LABORER

Notes: Abbreviated keyword lists; the full regular expressions are in `code/03_sample.py` of the replication package. Earlier rules take priority over later ones.