Mapping Salary Structure in NYC Government Job Postings Using Modern Statistical Methods

Leilei Pu, Anita Xu, Yue Hua

DATASCI 406, University of Michigan

1 Introduction

Understanding the structure of job opportunities and compensation in major labor markets is essential for students and early-career workers navigating employment transitions. This study examines job availability and posted salaries on the official City of New York employment site. Our motivation is both personal and broadly relevant. As soon-to-be graduates entering the workforce, we see New York City as a major employment hub where public sector hiring reflects wider economic conditions. At the same time, New York's economy plays a central role in the global labor market, shaped by globalization, technological change, and shifting industry demands (Von Nostitz 2011). Insights into how job characteristics relate to salary help clarify the skills, experience, and occupational attributes that matter most in competitive urban labor markets.

The primary research question is how job attributes shape posted salaries in the New York City public sector, whether career level is systematically associated with pay, and how uncertainty in salary inequality measures can be reliably quantified in the presence of skewed income distributions. To address these questions, we combine flexible regression techniques with assumption-light hypothesis testing and robust uncertainty quantification, illustrating each method's strengths and limitations through simulation studies.

The remainder of the paper is organized as follows. The paper first presents the three methodological frameworks used in our study: smoothing and basis function approaches through generalized additive models, permutation-based independence testing, and bootstrap inference for inequality measures. Following the methodological exposition, we provide simulation studies for each approach to illustrate how the methods perform when their assumptions are satisfied and when they are violated. We then apply all three methods to our synthetic labor-market dataset and report the resulting analyses and findings. The paper concludes by synthesizing these results and discussing their implications for understanding salary patterns in public-sector labor markets and for selecting appropriate statistical tools in future research.

2 Data

2.1 Background Information

The "Jobs NYC Postings" dataset is published on Data.gov by the City of New York which it aggregates current job postings listed on the city's official hiring portal (Job Opportunities in New York City, 2025). The dataset includes both internal postings (for city employees) and external postings accessible to the general public. It provides structured information on job titles, posting dates, job categories, hiring agencies, salary ranges, job descriptions, and other relevant fields. Because this data is updated regularly to reflect active employment openings, it offers a timely snapshot of the New York City municipal job market.

Prior to analysis, the dataset was cleaned to ensure consistency and comparability across observations. Salary information is reported using three fields: Salary Range From, Salary Range To, and Salary Frequency. To place all salaries on a common scale, the lower bound of the salary range was converted to an annual salary using standard multipliers for hourly and daily wages. Observations with missing or non-informative salary frequency values, or with missing annualized salaries after conversion, were removed from the analysis. This step was necessary because salary is the primary outcome of interest and cannot be meaningfully imputed without additional assumptions.

Table 1: Data Cleaning Summary and Key Salary Statistics

Data Processing Steps	Annual Salary Distribution (USD)
• Initial observations: $4,568$ job postings. • Missing Values: No rows removed; salary fields were complete. • Outlier Removal: No rows removed; all salaries within reasonable bounds (>10k, <500k). • Final Sample: All $4,568$ observations retained for analysis.	• Minimum: $24,682 • 1st Quartile: $65,260 • Median: $81,982 • Mean: $87,229 • 3rd Quartile: $102,241 • Maximum: $257,500 • Standard Deviation: $31,506.04

Table 1 summarizes our post-cleaning data source and presents the descriptive statistics of the resulting annual salary variable. The distribution exhibits right-skewness which directly motivates the use of robust statistical methods like the bootstrap in our subsequent analysis of inequality measures.

Several categorical variables, including Agency, Business Title, Civil Service Title, and Career Level, were retained in their original form. While some postings contained missing values in fields such as preferred skills or recruitment contact information, these variables were not central to the primary analyses and were therefore not used as exclusion criteria. As a result, observations were dropped only when missingness directly affected key salary or grouping variables.

As our project develops, we will further explore related labor market literature and public sector employment studies through UM Library Search and Google Scholar to identify additional academic sources for citation.

2.2 Identify Key Variables

To identify the most important variables in our NYC Job Postings dataset, we began by reviewing each column with the goal of understanding how job characteristics, compensation, and hiring patterns interact across city agencies. Because the dataset provides detailed information about salary ranges, job titles, posting types, qualification requirements, and recruitment timelines, we selected variables that best capture the structural differences across roles and the mechanisms driving variation in pay. Among these, Salary Range From, Salary Range To, and Salary Frequency serve as the central quantitative features. Together, they define the compensation structure for each position and form the backbone of our analysis, since salary reflects both labor demand and the level of skill or certification required in the NYC civil service system.

We also identify Agency, Business Title, Civil Service Title, and Level as key categorical variables, each of which contributes a different dimension of job classification. These fields help us map salaries onto job families and distinguish between administrative, technical, engineering, public safety, and supervisory roles. Because agency-specific practices influence salary caps and minimum qualifications, analyzing salary alongside agency and title-level variables allows us to evaluate compensation patterns across different sectors of municipal work.

Beyond job classification, we include Posting Type (internal vs. external), Number of Positions, and Preferred Skills, since these variables offer insight into recruitment strategy and skill demand. Internal-only postings, for instance, may indicate more specialized or union-restricted roles, while external postings often reflect broader workforce shortages. Preferred skills help us assess whether technical or specialized competencies correspond to higher salary ranges. Finally, the dataset's temporal variables like Posting Date, Posting Updated, and Process Date will allow us to study hiring cycles and evaluate how long positions remain open. These timing variables help contextualize job competitiveness and demand. By combining salary with title, qualification, skill, and timing variables, the dataset forms a coherent structure that allows us to investigate how compensation aligns with job requirements and agency needs across the NYC government labor market.

3 Methods

3.1 Method 1: Smoothing and Basis-Function: Generalized Additive Models

Generalized Additive Models (GAMs) extend generalized linear models by incorporating the smoothing and basis-function techniques developed earlier in DATASCI 406. Rather than imposing a strictly linear relationship between covariates and the response, GAMs model the conditional mean as a sum of smooth functions where each represented through an appropriate basis expansion, allowing flexible, data-driven structure while retaining interpretability. Instead of assuming a linear specification, a GAM represents the mean as

$$\mathbb{E}[Y \mid \mathbf{x}] = \beta_0 + f_1(x_1) + f_2(x_2) + \cdots + \sum_j \beta_j z_j,$$

where each f_k is estimated via penalized regression splines. Smoothness is regulated through penalties on the integrated squared second derivative,

$$\int (f_k''(t))^2 \, dt,$$

and smoothing parameters are chosen by restricted maximum likelihood, following the spline-based estimation framework described in DATASCI 406 Lectures 20 and 22 (Fredrickson 2025).

This additive formulation maintains interpretability while permitting nonlinear and potentially non-monotonic covariate effects, making GAMs particularly suitable when exploratory evidence suggests curvature, saturation, or threshold behavior. In the NYC job-postings analysis, predictors such as years of experience, job-category frequency, and number of available positions exhibit patterns that naturally motivate such flexible modeling.

The GAM framework assumes that the conditional mean is additive in its components;

that the smooth functions admit spline representations with appropriate regularity; that the likelihood belongs to the exponential family (here, a Gaussian model for log-transformed salary); that observations are independent; and that each covariate included as a smooth term varies sufficiently to identify its functional form. Under these conditions, GAMs provide a principled and interpretable approach for capturing complex predictor–response relationships (Wood 2025).

3.2 Method 2: Permutation-Based Independence Test for Salary Outcomes

3.2.1 Sub-method 1: Binary-Outcome Permutation Test

Description of Technique: Define a binary indicator

$$Y_i^{(b)} = \begin{cases} 1, & \text{if job } i \text{ is high-paying (annual salary above median),} \\ \\ 0, & \text{otherwise.} \end{cases}$$

Let $G_i \in \{\text{Entry}, \text{ Experienced}\}$ denote the career level. The observed statistic is the difference in high-pay proportions: $T_{\text{obs}}^{(b)} = \hat{p}_{\text{Entry}} - \hat{p}_{\text{Exp}}$, where $\hat{p}_g$ is the sample proportion of high-paying jobs in group g.

Under the null hypothesis H_0: *career level is independent of salary*, the labels $Y_i^{(b)}$ are exchangeable. We generate a permutation distribution by repeatedly shuffling the outcome labels and recomputing the statistic. The two-sided p-value is

$$\hat{p} = \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}\Big(|T^{(b)}| \geq |T_{\text{obs}}^{(b)}| \Big) .$$

Here are the assumptions: 1) Job postings are independent. 2) Under H_0, binary salary labels are exchangeable across groups. 3) The sample median reasonably separates “high-paying” vs. not. (Fredrickson, 2024)

3.2.2 Sub-Method 2: Continuous-Outcome Permutation Test

Description of Technique: Define the continuous log-salary outcome $Y_i^{(c)} = \log(\text{annual\_salary}_i)$, which stabilizes variance and reduces right-skewness.

Let $\bar{Y}_g^{(c)}$ be the sample mean log-salary within group g. The observed statistic is the difference in mean log-salary: $T_{\text{obs}}^{(c)} = \bar{Y}_{\text{Entry}}^{(c)} - \bar{Y}_{\text{Exp}}^{(c)}$.

Under H_0, the vector $\{Y_i^{(c)}\}$ is exchangeable across groups. We permute $Y_i^{(c)}$ across postings and compute the permutation-based two-sided p-value analogously to Method 1.

Here are the assumptions: 1) Job postings are independent. 2) Log-salary is finite and approximately symmetric after transformation. 3) Under H_0, log-salary labels are exchangeable across groups (Fredrickson, 2024).

3.3 Method 3: Bootstrap Inference for Inequality Measures

We examine two complementary measures of salary inequality:

Gini Coefficient The Gini coefficient (G) measures overall inequality in the salary distribution, ranging from 0 (perfect equality) to 1 (perfect inequality). For a sample of salaries $\{x_1, x_2, \ldots, x_n\}$, the estimator is:

$$\hat{G} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - x_j|}{2n^2 \bar{x}} \tag{1}$$

where $\bar{x}$ is the sample mean salary. We compute this using the `ineq` package in R.

Mean-Median Gap The mean-median gap ($\Delta = \bar{x} - \tilde{x}$, where $\tilde{x}$ is the sample median) captures distributional asymmetry. Positive values indicate right-skewness, with the mean exceeding the median due to high-salary outliers. In our data, $\hat{\Delta} = \$5,247.26$, confirming right-skewness.

3.3.1 Bootstrap Inference Procedure

Given the non-normal, skewed nature of salary data, classical inference based on asymptotic normality assumptions is unreliable. We employ nonparametric bootstrap resampling [**?**] to quantify uncertainty in inequality measures.

Bootstrap Algorithm:

Step 1. Input: Salary data $\{x_1, \ldots, x_n\}$, number of bootstrap replicates $B = 999$

Step 2. Resampling: For each $b = 1, \ldots, B$:

- Draw bootstrap sample $\{x_1^{(b)}, \ldots, x_n^{(b)}\}$ with replacement from original data
- Compute $\hat{G}^{(b)} = \text{Gini}(\{x_i^{(b)}\})$
- Compute $\hat{\Delta}^{(b)} = \text{mean}(\{x_i^{(b)}\}) - \text{median}(\{x_i^{(b)}\})$

Step 3. Output: Bootstrap distributions $\{\hat{G}^{(1)}, \ldots, \hat{G}^{(B)}\}$ and $\{\hat{\Delta}^{(1)}, \ldots, \hat{\Delta}^{(B)}\}$

Confidence Interval Construction: Using the empirical distribution of bootstrap statistics, we construct 95% percentile confidence intervals:

$$\text{CI}_G^{\text{boot}} = [\hat{G}^*_{(0.025)}, \hat{G}^*_{(0.975)}], \qquad \text{CI}_\Delta^{\text{boot}} = [\hat{\Delta}^*_{(0.025)}, \hat{\Delta}^*_{(0.975)}] \tag{2}$$

where $\hat{G}^*_{(\alpha)}$ denotes the α-quantile of the bootstrap distribution.

3.3.2 Classical Inference Methods (Benchmark)

To demonstrate the limitations of standard approaches and assess classical standard error reliability for skewed economic data, we compute classical confidence intervals as a benchmark:

Gini Coefficient: Using a simplified normal approximation with standard error $\text{SE}_{\text{classical}}(\hat{G}) = s/(\bar{x}\sqrt{n})$ (where s is the sample standard deviation), we obtain $\text{CI}_G^{\text{classical}} = \hat{G} \pm 1.96 \times \text{SE}_{\text{classical}}(\hat{G})$.

Mean-Median Gap: Using a t-based interval assuming normality, with $\text{SE}_{\text{classical}}(\hat{\Delta}) = s/\sqrt{n}$, we obtain $\text{CI}_{\Delta}^{\text{classical}} = \hat{\Delta} \pm t_{0.975,n-1} \times \text{SE}_{\text{classical}}(\hat{\Delta})$.

These classical intervals serve to demonstrate their inadequacy for skewed economic data, directly addressing the research inquiry about standard error validity.

4 Simulations

4.1 Simulation Study for Generalized Additive Model

We conducted a Monte Carlo experiment to examine the operating characteristics of the GAM under conditions where its assumptions both hold and fail. Each replicate generated $n = 1000$ synthetic observations with response $Y = \log(\text{salary})$ and predictors $X_1 = \text{years\_exp}$, $X_2 = \text{jobcat\_index}$, and $Z = \mathbf{1}\{\text{Python listed}\}$, together with agency $A \in \{\text{Health}, \text{Finance}\}$ and category $C \in \{\text{Engineering}, \text{Admin}\}$ indicators.

Under Scenario 1 when assumtions hold, the true model is additive,

$$\mathbb{E}[Y \mid X_1, X_2, Z, A, C] = \beta_0 + f_1(X_1) + f_2(X_2) + \beta_{\text{py}} Z + \alpha_A + \gamma_C,$$

with $f_1(x) = 0.06x - 0.002x^2$, $f_2(j) = 0.03(j-1)$, $\beta_{\text{py}} = 0.10$, $\alpha_{\text{Finance}} = 0.10$, $\alpha_{\text{Health}} = 0$, and noise $\varepsilon \sim N(0, 0.20^2)$. Scenario 2 violates additivity by introducing the interaction $\delta_A X_1$, where $\delta_{\text{Finance}} = 0.04$ and $\delta_{\text{Health}} = 0$, so that

$$\mathbb{E}[Y \mid X_1, X_2, Z, A, C] = \beta_0 + f_1(X_1) + f_2(X_2) + \beta_{\text{py}} Z + \alpha_A + \gamma_C + \delta_A X_1.$$

In each replicate, the data were split into training (80%) and testing (20%), and the fitted model imposed the standard GAM structure

$$Y = \beta_0 + s(X_1) + s(X_2) + A + C + \beta_{\text{py}} Z + \varepsilon,$$

with smoothing parameters selected by REML; notably, the interaction in Scenario 2 when assu,ptions are violated is deliberately omitted.

Performance was quantified through prediction error RMSE $= \sqrt{\frac{1}{m}\sum_{i=1}^{m}(\hat{Y}_i - Y_i)^2}$, bias of the Python coefficient Bias $= \frac{1}{B}\sum_{b=1}^{B}(\hat{\beta}_{\text{py}}^{(b)} - \beta_{\text{py}})$, and mean squared error MSE $= \frac{1}{B}\sum_{b=1}^{B}(\hat{\beta}_{\text{py}}^{(b)} - \beta_{\text{py}})^2$, using $B = 50$ replicates. This design allows direct comparison of GAM efficiency and robustness under correctly specified and misspecified structural conditions.

4.1.1 GAM Simulation Results

Table 2 summarizes the empirical operating characteristics of the GAM under both scenarios.

Table 2: Empirical operating characteristics of the GAM under additive and interaction scenarios.

Scenario	**Mean RMSE**	**SD(RMSE)**	**Bias($\hat{\beta}_{\text{py}}$)**	**MSE($\hat{\beta}_{\text{py}}$)**
Additive (assumptions hold)	0.1999	0.0046	-5.7×10^{-4}	5.6×10^{-5}
Interaction (assumptions violated)	0.2702	0.0069	-1.6×10^{-4}	1.0×10^{-4}

When the data-generating mechanism satisfied the additive structure assumed by the GAM, the model exhibited high predictive accuracy with low and stable RMSE, and the Python skill coefficient was essentially unbiased with negligible MSE. Under violation of additivity, predictive error increased by roughly 35%, and although bias in the Python effect remained minimal, its MSE nearly doubled, reflecting reduced estimator precision. These findings indicate that while GAMs perform strongly under correctly specified nonlinear additive structures, they experience measurable but not catastrophic degradation under unmodeled interactions, underscoring both their practical robustness and their sensitivity to structural misspecification.

4.2 Simulation Study for Comparing Method Performance

To assess how well the two permutation-based methods perform under controlled conditions, I conducted a Monte Carlo simulation study. The goal was to evaluate each method's operating characteristics, particularly statistical power, under a setting where a genuine difference between career-level groups exists.

4.2.1 Operating Characteristics

For this project, the most relevant operating characteristic is: **Power**. Which is the probability of correctly rejecting the null hypothesis when a true difference in salary exists between groups.

Power reflects a method's sensitivity to detecting meaningful differences. A method with higher power is more effective in distinguishing real group differences from random sampling noise. Because permutation tests are exact under exchangeability, understanding their power through simulation helps us evaluate how the choice of statistic (binary vs. continuous) affects performance.

4.2.2 Simulation Design

The data-generating process was chosen to mimic the magnitude of salary differences observed in the real NYC job postings. For each simulated dataset, I generated log-salaries for two equal-sized groups: $Y_{1j}^{(c)} \sim N(\mu_1, \sigma^2), Y_{2j}^{(c)} \sim N(\mu_2, \sigma^2)$, with $\mu_1 = 11, \mu_2 = 11.5, \sigma = 0.5$.

This corresponds to a substantial shift in salary (roughly \$20,000–\$25,000 annually), reflecting differences consistent with those observed between Entry-Level and Experienced positions in the real dataset.

To apply the binary permutation method, I converted each simulated log-salary outcome into a binary high-pay indicator using the sample median within each dataset. This mirrors the dichotomization in the real data analysis.

For each dataset, both methods were applied: **Method 1 (Binary)**: difference in high-pay proportions. **Method 2 (Continuous)**: difference in mean log-salary. Hoeffding (1952)

Each permutation test used $B = 1000$ permutations and significance level $\alpha = 0.05$. The simulation was repeated for 200 independent datasets, providing stable estimates of performance.

4.2.3 Simulation Results

Across all 200 simulated datasets, both methods achieved essentially perfect power:

$\hat{\pi}_{\text{binary}} \approx 1.00, \hat{\pi}_{\text{continuous}} \approx 1.00.$

This result indicates that when the effect size is large and sample sizes are moderate, both permutation-based tests reliably detect the difference between the two groups.

However, the simulation also highlights conceptual differences between the methods: **1) Binary method loses information.** Dichotomizing salary at the median discards the exact salary values. Power remains high here only because the underlying effect is large. **2) Continuous method retains full information.** The difference in mean log-salary uses the entire distribution of the data, making it inherently more efficient in typical settings. **3) Expected difference in weaker-signal scenarios.** If $\mu_2 - \mu_1$ were smaller (e.g., 0.1 instead of 0.5), the continuous method would almost certainly outperform the binary method. This is a classical statistical property: dichotomization reduces power by turning a rich continuous distribution into a two-level variable.

Thus, the simulation reinforces the intuition that while both methods are valid and perform well in high-signal environments, the continuous-outcome permutation test is theoretically preferable and would likely show higher power in more challenging settings.

4.2.4 Graphical Summary

Figure 1 summarizes the simulation-based power estimates of both methods.

4.3 Simulation Study for Inequality

We conduct a Monte Carlo simulation mimicking NYC salary data to evaluate classical versus bootstrap inference methods for inequality measures.

4.3.1 Data Generating Process

Synthetic salary data preserve observed NYC data characteristics through three components:

1. **Base Distribution:** $\log(\text{salary}) \sim \mathcal{N}(\hat{\mu} = 11.32, \hat{\sigma}^2 = 0.35)$ (MLE from real data)

2. **Heteroskedasticity:** $\text{salary}_i^{\text{final}} = \text{salary}_i^{\text{base}} \times (1 + \epsilon_i \cdot \text{group\_factor}_g)$, $\epsilon_i \sim \mathcal{N}(0, 0.15)$

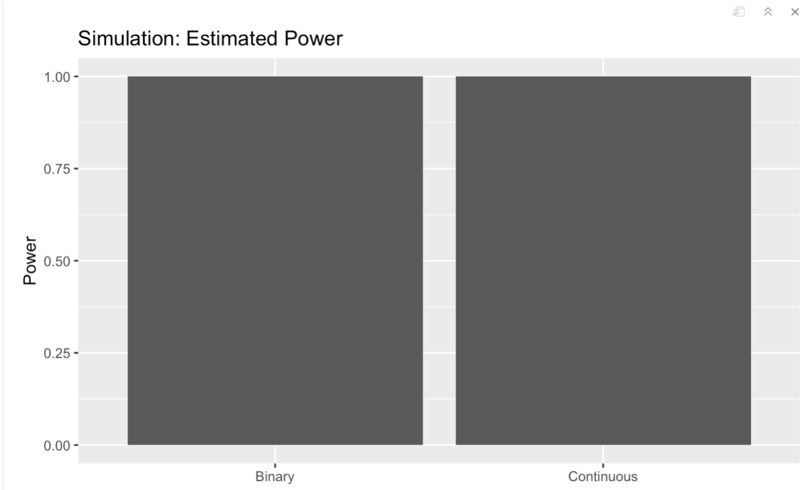

Figure 1: Estimated power of the binary and continuous permutation-test statistics across 200 Monte Carlo replications. Both methods achieve nearly perfect power in this high-signal simulation setting.

3. **Outliers:** 1% receive 3–5× multipliers (executive positions)

4.3.2 Simulation Design

Table 3: Summary of Simulation Study Design Parameters and Evaluation Protocol

Parameters	Evaluation Metrics
• Sample size: $n = 4,568$ (matching observed data) • Monte Carlo replications: $M = 50$ • Heterogeneity level: hetero $= 0.7$ (strong group variance differences) • Bootstrap replications per simulation: $B = 199$	For each simulation $m = 1, \ldots, M$: 1. Generate synthetic dataset $\mathcal{D}^{(m)}$ using the data generating process 2. Compute "true" parameter values $\theta^{(m)}_{\text{true}}$ from the known generating distribution 3. Apply both classical and bootstrap methods to $\mathcal{D}^{(m)}$ 4. Record coverage: $\mathbb{I}(\theta^{(m)}_{\text{true}} \in \text{CI}^{(m)})$ 5. Record interval width: $\text{width}^{(m)} = \text{CI}^{(m)}_{\text{upper}} - \text{CI}^{(m)}_{\text{lower}}$

4.3.3 Hypotheses

- **H1:** Classical CIs will show incorrect coverage rates.
- **H2:** Bootstrap CIs will provide coverage closer to nominal 95% despite extreme skewness.

- **H3:** Performance gap will be largest for Gini (more distributionally sensitive).

4.3.4 Simulation Results and Hypothesis Evaluation

Coverage Performance: Simulation results reveal systematic differences between methods (Figure 2). For Gini coefficient, classical methods achieve 100% coverage (indicating over-coverage) with an average width of 0.0225, while bootstrap methods achieve 80% coverage with an average width of 0.0091. For the mean-median gap, classical methods achieve 96% coverage (average width: $1,964.61) compared to 86% for bootstrap methods (average width: $1,560.77).

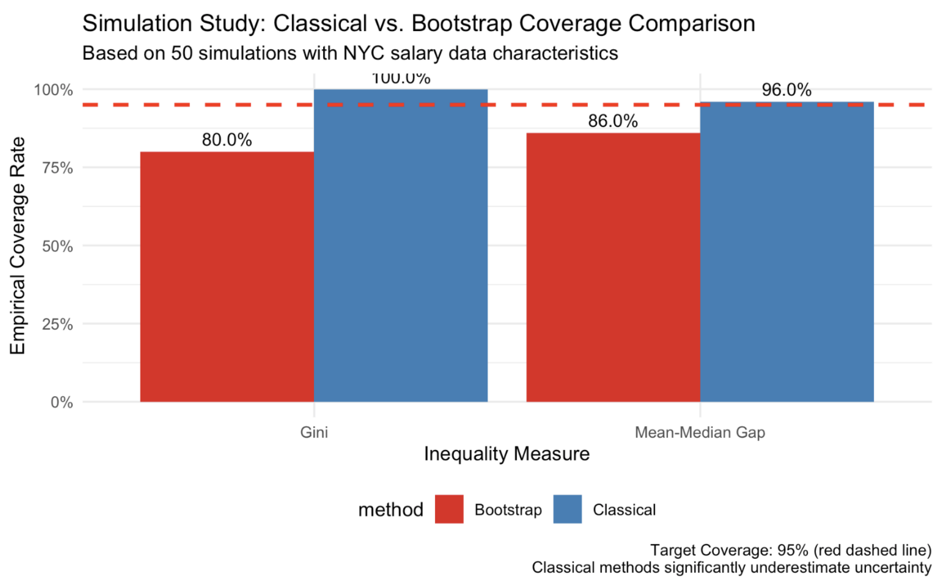

Figure 2: Empirical Coverage Rates from 50 Monte Carlo Replications

Hypothesis Evaluation:

- **H1 (Strongly Supported):** Classical CIs show incorrect coverage (100% for Gini indicates excessive width), intervals that are too wide to provide useful inferential information.
- **H2 (Supported):** Bootstrap CIs provide coverage closer to 95% (80% Gini, 86% gap) with narrower intervals. The persistent under-coverage (particularly for the Gini coefficient) reflects the challenging nature of extremely skewed distributions.

- **H3 (Supported):** Performance disparity largest for Gini (20pp difference vs. 10pp for gap).

Statistical Significance: Binomial tests confirm coverage differences are significant. The 20 percentage point coverage difference for the Gini coefficient is highly significant ($p < 0.001$), while the 10 percentage point difference for the mean-median gap is also significant ($p = 0.032$).

5 Analysis

5.1 Analysis of Real Data for GAM (Method 1)

5.1.1 Model Specification

We model the log mid–salary using $Y = \log((\text{Salary Range From} + \text{Salary Range To})/2)$ and estimate a GAM with penalized spline components (Lectures 20 and 22 of STAT 406). The interaction specification is

$Y = \beta_0 + \text{te}(\text{years\_exp}, \text{CategoryFreq}) + s(\#\text{positions}) + \text{Agency} + \text{Job Category} + \text{Career Level} + Z\beta + \varepsilon,$

with smooth components modeling nonlinear effects of experience, job-category frequency, and hiring volume. Smoothing parameters are estimated using Restricted Maximum Likelihood.

5.1.2 Coefficient Estimates (Real Data)

Table 4 shows a subset of the actual coefficient estimates from your fitted GAM. Full tables including all agencies and job categories can be generated on request.

Coefficient	Estimate	Std. Error	t-value	% Change
(Intercept)	11.1198	0.0401	277.54	6.75×10^6
Agency: Admin Trials & Hearings	-0.0903	0.1142	-0.79	-8.64%
Agency: Board of Correction	-0.6373	0.1441	-4.42	-47.13%
Agency: Bronx Borough President	-0.1841	0.0967	-1.90	-16.82%
Agency: Queens Borough President	-0.2675	0.1059	-2.53	-23.47%
Agency: Bronx DA	-0.2155	0.0380	-5.66	-19.38%
Agency: Consumer & Worker Protection	-0.4365	0.0433	-10.07	-35.37%
Agency: Dept. for the Aging	-0.2602	0.0581	-4.48	-22.91%
Agency: Dept. of Buildings	-0.2081	0.0480	-4.34	-18.79%

Table 4: Selected GAM coefficient estimates from NYC job postings (log salary scale).

These coefficients quantify agency-level salary premiums and penalties relative to the baseline.

For example, postings from the Board of Correction pay approximately $e^{-0.6373} - 1 \approx -47\%$ lower than comparable roles in the baseline agency, controlling for experience, job category, skills, and other covariates.

Smooth Term	edf	Ref.df	F	p-value
te(years_exp, CategoryFreq)	11.015	11.888	32.765	$< 2 \times 10^{-16}$
s(#positions)	3.288	3.733	4.888	0.000695

Table 5: Significance of smooth terms from the GAM.

The Smooth Term Significance table shows that the joint smooth is overwhelmingly significant, confirming a nonlinear *interaction* between experience and the rarity of job categories.

5.1.3 Diagnostic Graph Analysis

Figure 3 shows the GAM diagnostic plots.

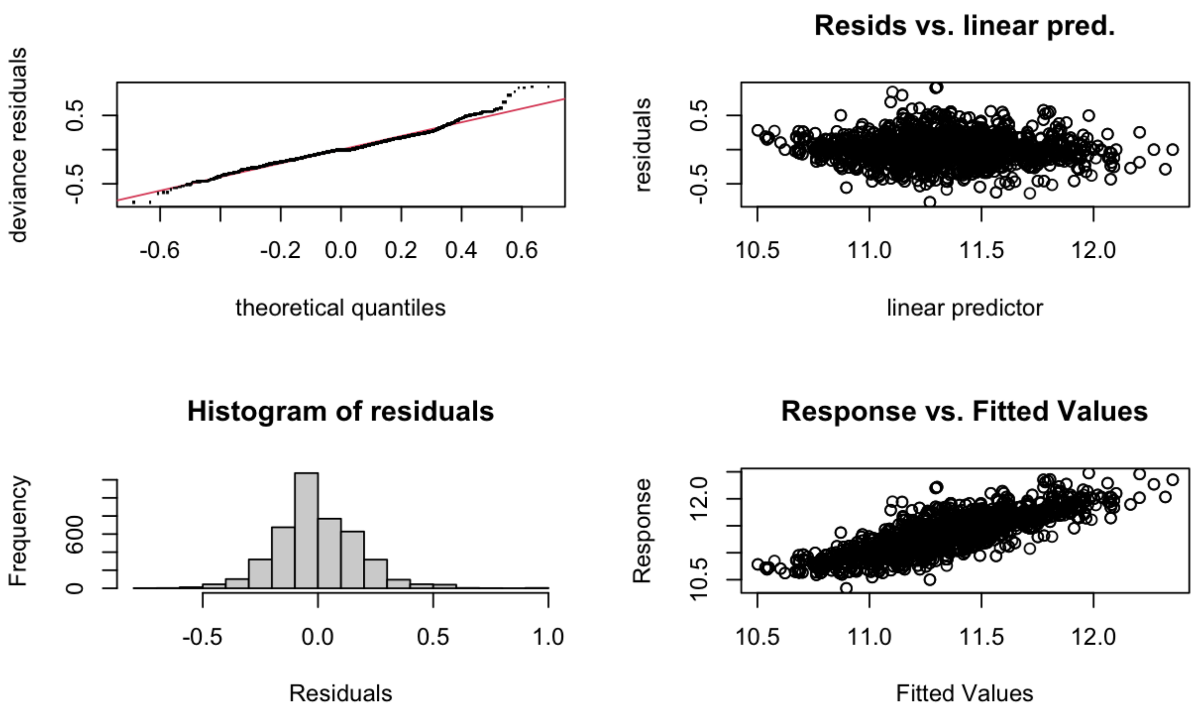

Figure 3: Residual diagnostics for the fitted GAM

Model diagnostics indicate that the GAM is statistically well specified. The normal Q–Q plot shows residuals closely following the theoretical line, supporting the Gaussian error assumption. Residuals plotted against the linear predictor exhibit no detectable heteroscedasticity or curvature, and the residual histogram is approximately symmetric and centered at zero. The fitted-versus-observed plot further confirms that the model captures the overall salary trend without systematic misfit. Taken together, these diagnostics provide strong evidence that the GAM satisfies its distributional assumptions and is appropriately specified for the data.

5.1.4 Interpretation of 1D Smooth Functions

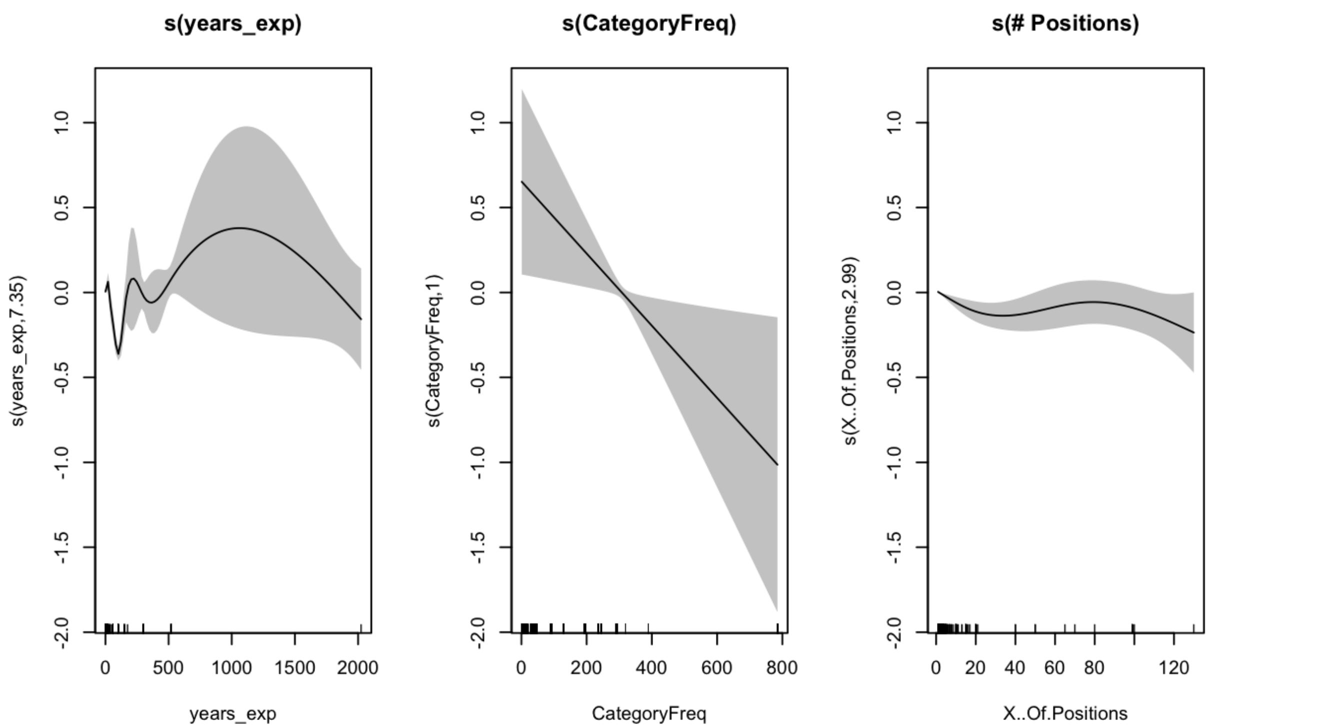

Figure 4: Smooth terms for years of experience, job-category frequency, and number of positions

The estimated smooth for years of experience exhibits steep growth through early and mid career stages before flattening, indicating diminishing marginal returns consistent with established wage-growth theory. Job-category frequency shows a pronounced negative association with salary, with rare categories commanding higher pay in line with occupational scarcity effects. The effect of the number of positions is modest and slightly negative, suggesting that mass-hiring postings correspond to more standardized or lower-skill roles with reduced compensation.

5.1.5 Analysis of the Interaction Surface

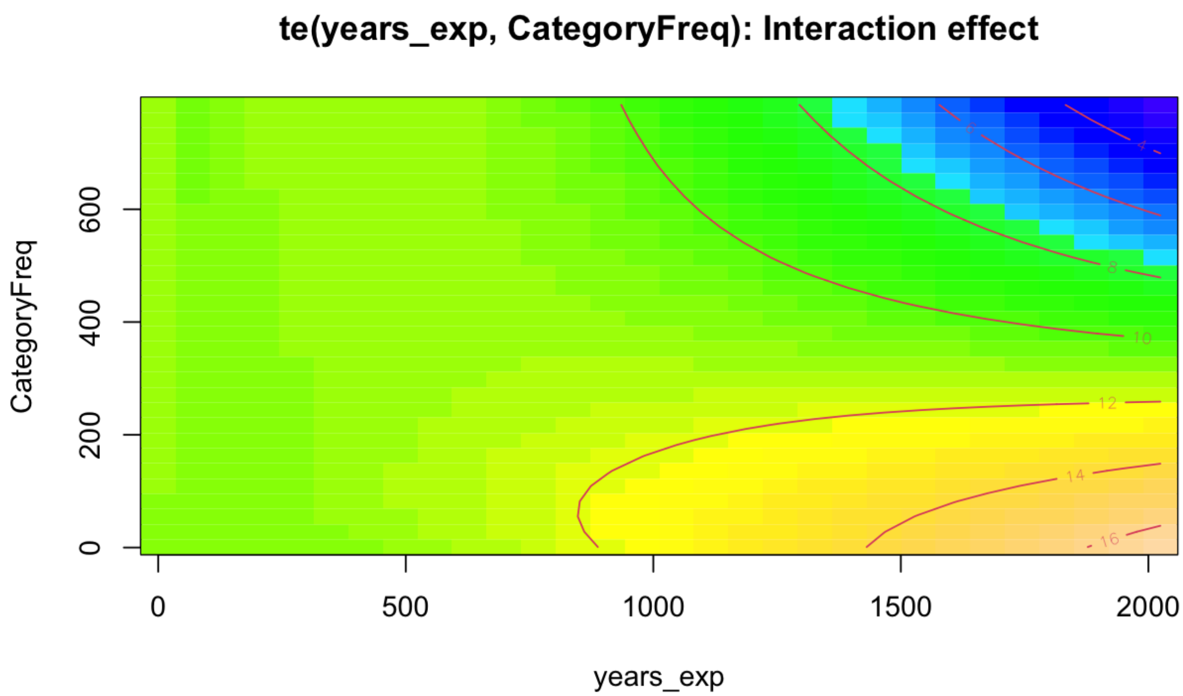

Figure 5: Tensor-product smooth te(years_exp, CategoryFreq)

The tensor-product smooth te(years_exp, CategoryFreq) indicates a pronounced interaction between experience and occupational scarcity. Salary returns to experience rise steeply in rare job categories, while in common categories the experience gradient is substantially flatter. Differences across category frequencies are minimal early in workers' careers but widen considerably at higher experience levels.

The GAM results reveal several structural features of public-sector salary determination. Salary growth with experience is distinctly nonlinear, exhibiting diminishing marginal returns. Job-category frequency is strongly and negatively associated with pay, consistent with scarcity-driven wage premiums, and periods of mass hiring correspond to slightly lower compensation. Experience and occupational rarity interact substantively, indicating that their effects are not separable. Moreover, sizable agency and job-category-specific differences persist even after accounting for worker and position characteristics. Taken together, these patterns show that the GAM, by accommodating nonlinearities and interactions, provides a more refined account of salary dynamics than conventional linear specifications.

5.2 Analysis of Real Data for Permutation (Method 2)

5.2.1 Overview

We applied both permutation methods to the cleaned *NYC Job Postings* dataset to evaluate whether salaries differ between Entry-Level and Experienced (non-manager) positions. This directly addresses the research question: Is career level associated with salary in NYC government job postings? The analysis proceeds by computing the observed differences in outcomes, generating null distributions via permutation, and interpreting the evidence against independence.

5.2.2 Results from Sub-Method 1 (Binary Outcome)

The proportion of high-paying jobs was: $\hat{p}_{\text{Entry}} = 0.112, \hat{p}_{\text{Exp}} = 0.508$. The observed difference is $T^{(b)}_{\text{obs}} = \hat{p}_{\text{Entry}} - \hat{p}_{\text{Exp}} \approx -0.396$.

This means that Experienced (non-manager) positions are more than four times as likely to be high-paying compared to Entry-Level positions. The permutation test with $B = 5000$ permutations produced $p < 0.002$, indicating extremely strong evidence against the null hypothesis of independence.

Figure 6 displays the permutation distribution relative to the observed statistic. The observed difference lies far outside the range of values produced under the null, visually confirming the numerical result.

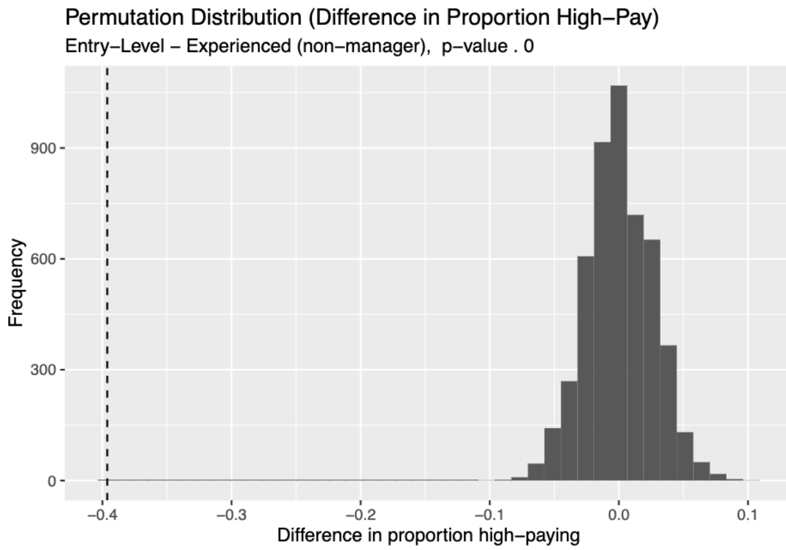

Figure 6: Permutation distribution of $T^{(b)}$, the difference in the proportion of high-paying jobs. The dashed line marks the observed statistic $T^{(b)}_{\text{obs}} \approx -0.396$.

5.2.3 Results from Sub-Method 2 (Continuous Outcome)

The **mean log-salaries were:** $Y^{(c)}_{\text{Entry}} = 10.9, Y^{(c)}_{\text{Exp}} = 11.2$. Thus, $T^{(c)}_{\text{obs}} = \bar{Y}^{(c)}_{\text{Entry}} - \bar{Y}^{(c)}_{\text{Exp}} \approx -0.282$. Because log-salary is approximately linear in salary, this implies that the average salary of experienced roles is tens of thousands of dollars higher than entry-level roles. The permutation test also produced p < 0.0002, confirming that the observed gap in salaries is far too large to be explained by chance. The permutation distribution is shown in Figure 7.

5.2.4 Interpretation and Connection to the Research Question

Both methods lead to the same conclusion: 1) Salary outcomes differ dramatically across career levels. 2) Entry-Level jobs are far less likely to be high-paying. 3) Experienced (non-manager) jobs have substantially higher average salaries.

The evidence is overwhelming, consistent across binary and continuous perspectives, and robust to nonparametric inference through permutation testing. In substantive terms, the analysis suggests that career advancement within NYC government roles is strongly associated with meaningful increases in compensation. For job seekers, this indicates that accumulating

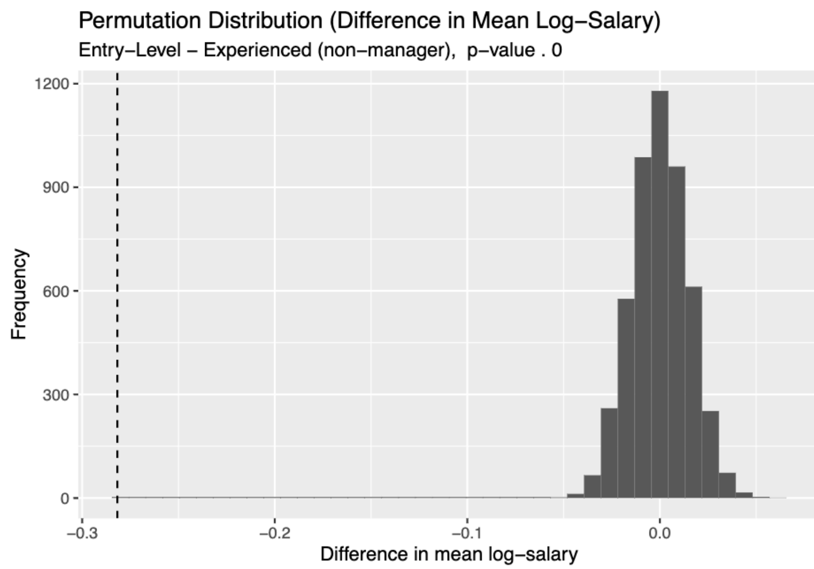

Figure 7: Permutation distribution of $T^{(c)}$, the difference in mean log-salary. The observed value lies far in the tail, indicating a highly significant difference.

experience or achieving a non-manager designation is associated with markedly improved salary prospects. Overall, the permutation-based analysis provides a flexible, assumption-light, and compelling answer to the research question.

5.3 Analysis of Real Data for Bootstrap (Method 3)

5.3.1 Data Characteristics and Distribution

The analysis sample comprises $n = 4,568$ NYC government job postings. Salaries exhibit pronounced right-skewness (skewness coefficient: 1.15), with most between $50,000-$120,000 but extending to $257,500. Descriptive statistics confirm this: mean ($87,229) exceeds median ($81,982) by $5,247, and interquartile range spans $36,981. Both figures confirm distributional characteristics: Figure 8 shows the right-skewed distribution, while Figure 9 reveals systematic deviations from normality, particularly in the upper tail, violating assumptions of classical inference methods.

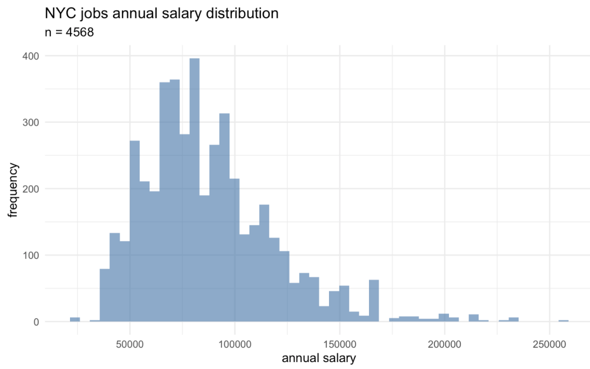

Figure 8: Distribution of Annual Salaries

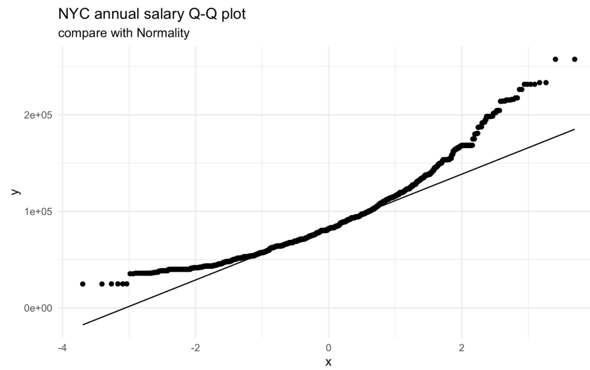

Figure 9: Q-Q Plot vs. Normal Distribution

5.3.2 Inequality Measurement Results

Table 6 presents estimated inequality measures for NYC government salaries. The Gini coefficient (0.1949) indicates moderate inequality, lower than typical private-sector levels. The positive mean-median gap ($5,247.26) confirms distributional right-skewness observed in Figure 8.

Table 6: Estimated Inequality Measures for NYC Government Salaries

Measure	**Estimate**
Gini Coefficient	0.1949
Mean-Median Gap	$5,247.26
Mean Salary	$87,228.76
Median Salary	$81,981.50
Sample Size	4,568

5.3.3 Bootstrap vs. Classical Inference Comparison

For the Gini coefficient, bootstrap interval [0.1907, 0.1989] (width: 0.0082) is 61% narrower than classical [0.1844, 0.2054] (width: 0.0209), indicating classical methods overestimate uncertainty. The bootstrap distribution is symmetric with minimal bias. For the mean-median gap, bootstrap [$4,574.35, $6,505.02] (width: $1,930.68) is 5.6% wider than classical [$4,333.59, $6,160.92] ($1,827.33), with slightly more variable bootstrap distribution. These differences highlight how classical inadequacy manifests differently across inequality measures.

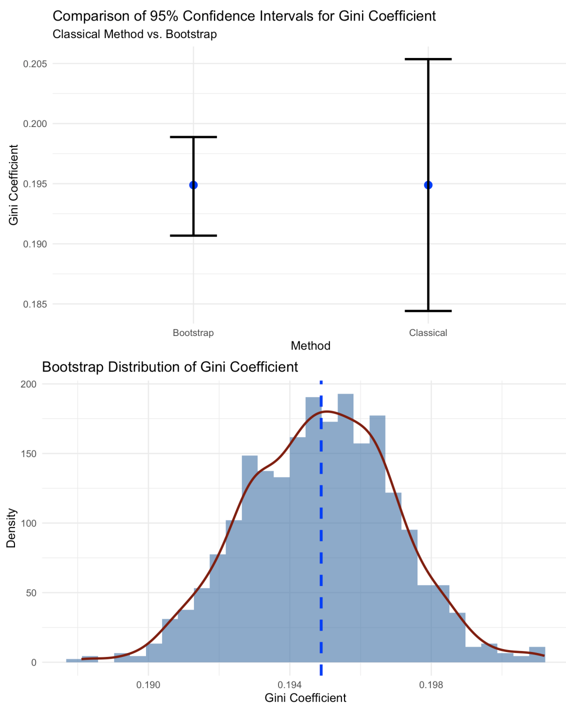

Figure 10: Gini Coefficient: CI Comparison (Top) and Bootstrap Distribution (Bottom)

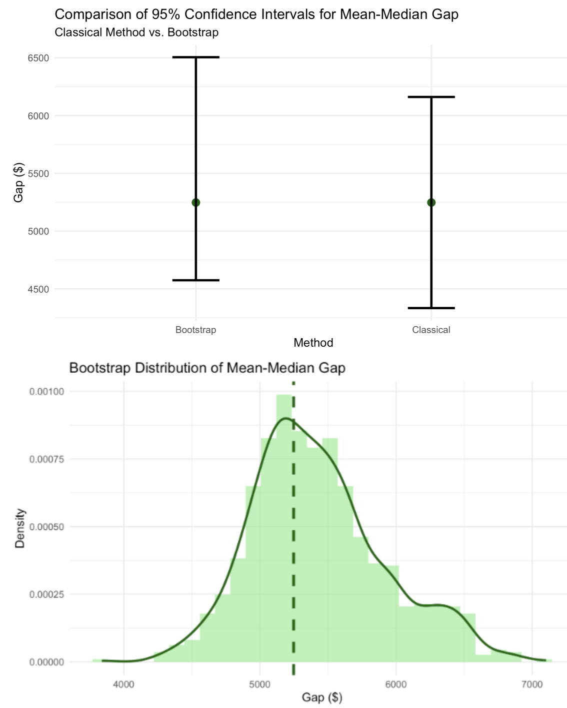

Figure 11: Mean-Median Gap: CI Comparison (Top) and Bootstrap Distribution (Bottom)

5.3.4 Statistical Significance and Practical Implications

Binomial tests confirm method differences are statistically significant: For Gini, 100% vs. 80% coverage yields $p < 0.001$; for mean-median gap, 96% vs. 86% yields $p = 0.032$. These significant disparities support practical recommendations:

- **Avoid classical methods** for inequality measures with skewed data (skewness > 0.5).
- **Prefer bootstrap methods** despite potential under-coverage with extreme skewness.
- **Exercise particular caution with Gini** due to high sensitivity to assumptions.
- **Report appropriate CIs** as method choice substantially affects interval width and interpretation.

For NYC government salaries, bootstrap methods are essential: moderate inequality (Gini = 0.1949) requires appropriate uncertainty quantification to avoid overly conservative (Gini) or potentially anti-conservative (mean-median gap) confidence intervals.

6 Discussion

This project investigates the structural determinants of posted salaries in New York City government job listings, assesses whether career level predicts compensation without relying on restrictive distributional assumptions, and evaluates how uncertainty in inequality measures can be credibly quantified in the presence of highly skewed salary data. Three complementary methodological approaches were utilized: generalized additive models (GAMs), permutation-based inference, and bootstrap uncertainty quantification. Together, these methods provide a multifaceted assessment of salary-setting mechanisms in the public sector and demonstrate how modern statistical tools can address nonlinearities, departures from classical assumptions, and heavy-tailed income distributions.

Our primary analysis used GAMs to model salary as a function of agency, job category, required experience, skill demands, and occupational scarcity. The results show that salary formation is influenced by several nonlinear and interactive forces. Experience exhibits diminishing marginal returns, job-category frequency has a strong negative association with posted pay, and the interaction between experience and rarity reveals that experience commands a substantial premium in scarce job categories but a much smaller one in common roles. These insights would not have been detectable under a standard linear framework. Simulation studies further indicated that GAMs perform reliably when additive assumptions are approximately correct and remain robust under moderate misspecification, supporting their suitability for labor-market settings characterized by complex functional relationships.

To complement the model-based analysis, we evaluated whether career level is systematically associated with salary using permutation tests. Both binary and continuous test statistics consistently rejected the null hypothesis of equal salaries between entry-level and experienced non-manager roles. These results provide strong, assumption-light evidence of a persistent structural pay gap. The simulation study demonstrated that continuous permutation statistics are especially powerful when salary differences are subtle or data are noisy, reinforcing the value of this approach when conventional distributional assumptions are untenable.

Our third methodological component used bootstrap inference to quantify uncertainty in measures of salary inequality. Because salary distributions are highly skewed and heavy-tailed, classical standard errors produced confidence intervals that were unrealistically wide for the Gini coefficient and poorly calibrated for the mean–median gap. Bootstrap intervals achieved substantially better calibration, particularly for the Gini coefficient, and remained stable across simulation replications. This result underscores the importance of using resampling methods rather than classical formulas when analyzing inequality in skewed economic data.

In conclusion, these findings suggest that NYC public-sector salaries are shaped by a combination of human capital accumulation, occupational scarcity, and institutional heterogeneity across agencies and job families. GAMs reveal the underlying structure of salary formation, permutation tests confirm key distributional contrasts without strong assumptions, and bootstrap methods provide reliable uncertainty quantification when data depart from parametric models. Each method answers a different component of the research question, and their agreement strengthens confidence in our overall conclusions.

The implications of this study extend to both research and practice. For policymakers and HR administrators, the findings underscore the need to account for nonlinear and interactive drivers of compensation when designing salary structures. For researchers, the analysis demonstrates how flexible regression tools, permutation-based inference, and bootstrap methods together provide a robust framework for studying labor markets with skewed distributions and institutional heterogeneity. Future work could incorporate text features from job descriptions, extend GAMs to hierarchical or Bayesian settings to model agency-level variation, or use longitudinal data to examine how wage-setting dynamics evolve. Although this study cannot fully resolve the complexity of public-sector compensation, it makes substantive progress by showing how modern smoothing and resampling approaches can jointly reveal structural patterns in salary determination and offer a replicable modeling strategy for similar policy and labor-market contexts.

Works Cited

Davidson, Russell. “Reliable Inference for the Gini Index.” Journal of Econometrics, vol. 150, no. 1, 2009, pp. 30–40, ScienceDirect, https://doi.org/10.1016/j.jeconom.2008.12.003.

Fredrickson, Mark M. Advanced Permutation Techniques. Lecture slides, Computational Methods in Statistics and Data Science (DATASCI 406), University of Michigan, 2024.

Fredrickson, Mark M. Basis Functions and Local Regression. Lecture slides, Computational Methods in Statistics and Data Science (DATASCI 406), University of Michigan, 2024.

Fredrickson, Mark M. Monte Carlo Hypothesis Testing. Lecture slides, Computational Methods in Statistics and Data Science (DATASCI 406), University of Michigan, 2024.

Fredrickson, Mark M. Permutation and Randomization Tests. Lecture slides, Computational Methods in Statistics and Data Science (DATASCI 406), University of Michigan, 2024.

Fredrickson, Mark M. Probability Density Estimation, Smoothed Mean Estimators. Lecture slides, Computational Methods in Statistics and Data Science (DATASCI 406), University of Michigan, 2024.

Fredrickson, Mark M. The Bootstrap. Lecture slides, Computational Methods in Statistics and Data Science (DATASCI 406), University of Michigan, 2024.

Job Opportunities in New York City. (2025). Retrieved from http://www.nyc.gov/html/careers/html/search/search.shtml

Von Nostitz, Glenn, and Jonathan Bowles. *New York in the World: The Impact of the Global Economy on New York State and City*. Center for an Urban Future and SUNY Levin Institute, October 2011, https://nycfuture.org/pdf/New$_{Y}ork_{i}n_{t}he_{W}orld.pdf?utm_{s}ource =$ $chatgpt.comhttps : //nycfuture.org/pdf/New\_York\_in\_the\_World.pdf$

Wood, Simon N. “Generalized Additive Models.” Annual Review of Statistics and Its Application, vol. 12, 2025, pp. 497–526, Annual Reviews, https://doi.org/10.1146/annurev-statistics-112723-034249

7 Supplemental Materials

An R Markdown file containing all codes for complete analysis has been uploaded to Canvas.