Econometrics Paper

Kushal Durga, Ritvik Ellendula, Puneet Karanth, Manav Shah

How does the relationship between remote work and salary vary by company size in the global tech workforce?

July 31, 2025

Abstract

In its continuously growing and changing job market, the tech industry holds a large variety of work options, such as remote, hybrid, and in-person work in companies from small startups to tech giants such as Google and Amazon. With such a variety in remote work opportunities and company sizes, we sought to explore the effects of these variables on salaries of tech workers, specifically in the field of data science and artificial intelligence. The data used in this study comes from AI Jobs' Data Science Salary Index with about 140,000 survey results from workers in the industry. We performed a Welch's ANOVA test to determine the statistical significance and effect size of the variables, finding that the categorical variables of remote work ratio, company size, and the combination of the two are all statistically significant with notably low effect sizes. Ultimately, the low impact of the variables presents fascinating implications for the variety of work opportunities in the data science and artificial intelligence industry.

1 Introduction

In the aftermath of the COVID-19 pandemic, the world's labor markets have shifted dramatically—even several years after the incident. Specifically, there has been a drastic rise in remote work amongst organizations, regardless of company size. Employers were forced to find alternative methods of labor, and these methods of remote employment have allowed companies and their employees to overcome geographical barriers that span hundreds or even thousands of miles. Since the pandemic, one industry has stood out as a key adopter of remote work methods: technology. In respect to jobs relating to data science and artificial intelligence—data engineers, analysts, and researchers—there has been a dramatic increase in remote work. Companies of all sizes, from small startups to massive tech giants, have all been affected. This fundamental shift poses an important question: do salaries and rates differ between in-house workers and remote workers across different company sizes?

Adam Smith's compensating wage differentials theory posits that workers in jobs with undesirable characteristics will receive higher wage "premiums" to compensate for negative job aspects. Conversely, with desirable aspects of "remote "work"—eliminating commutes, providing flexibility, and saving time—remote workers may be amenable to lower wages than their in-house peers. However, the dynamics of equitable compensation are further convoluted by varying firm sizes. Larger firms historically offer higher wages due to greater capital resources derived from economies of scale, stronger bargaining power, and greater access to advanced infrastructure and technology. These firm-size wage premiums may interact with remote work in interesting ways: larger firms may offer remote roles without lowering pay, while smaller firms may offer remote work to substitute for higher wages. This research is critical as it addresses both remote work and firm size – two essential components of the modern labor market – and their association with compensation. As remote work becomes a permanent fixture in many industries, understanding the extent of its impact is critical for workers to better navigate the ever-changing business environment, enabling them to better negotiate pay and receive equitable standards. While theory provides some semblance of predicting the future economic environment, empirical evidence remains limited. Therefore, clarifying whether remote workers are systematically paid more or less, and how this varies with firm size, helps guide discussion around wage equity and labor mobility in the evolving digital economy.

Another paper, "Research on the Factors that Influence Wages—Take the Example of the Data Science Industry" by Yuyan Chen, explores a similar dataset found on the Kaggle website and originally sourced from aijobs.net (Data Science Salaries 2023). They sought to explore which factors had a relationship with pay in the data science industry, ultimately concluding that of the factors in the dataset, work experience, company location, place of residence, and company size were correlated with salaries. Their multiple regression results showed that remote ratio, a categorical variable differentiating between in-person, remote, and hybrid employees, and employment type did not have a significant correlation with wages, but they also conceded that their sample size of 3,755 data points may have caused this observation.

This paper isolates the variables of remote ratio, company size, and salary to determine if the remote ratio is associated with differences in pay and if that difference differs by company size. The dataset used covers similar variables but includes salaries from both 2024 and 2025, totaling over 132,000 data points. Ideally, this larger sample size should help determine whether the remote ratio is truly significant in this study.

2 Data

The data used within this paper is sourced from AI Jobs's Data Science Salary Index for 2025. Participants can voluntarily fill out a survey form with the following information collected: work year, experience level, employment type, job title, salary, currency, salary in USD, employee location, remote ratio, company size, and company location. Because this survey relies on volunteer responses, there exists an inherent voluntary response bias, meaning those who complete the survey may not accurately represent the population; however, this is simply a consideration, as the data cannot be adjusted to account for this. The survey is targeted to employees who work in fields related to AI and data science. For this paper, researchers isolated all data from 2025 and 2024 to control for potential inflationary effects across different time periods. The survey has collected over 140,000 entries in total, with over 132,000 entries submitted in 2024 and 2025. Data for the survey has been collected from the beginning of 2021 up to the present day, which as of this writing is July 8, 2025. The survey collects data from all nations globally but is heavily skewed toward U.S. participants, with 87 percent of responses coming from the U.S. and another 4 percent from

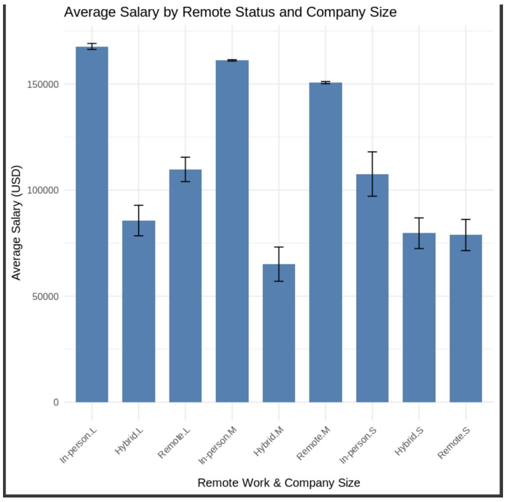

Figure 1: Nine Bar Graphs Representative of Average Salary (USD) with Differing Company Size and Remote Status

the U.K. The first variable used within this paper is the size of a company. The approximate size of the company is split into three categories: small, medium, and large. Small companies are defined as enterprises with fewer than 50 employees and contractors, medium companies have between 50 and 250 employees, and large companies are described as companies with over 250 employees. This definition falls in line with the European Commission guidelines for company size, with mid-cap companies having fewer than 250 employees and small-cap companies having fewer than 50 employees. The second variable used within this paper is the remote ratio, or the percent of time that employees work remotely, with 100 percent indicating full remote work and 0 percent suggesting only in-person work. Additionally, there are some who entered 50 percent, indicating hybrid work. The final variable is salary, which is defined as the amount of cash compensation annually for that worker. Other forms of compensation, like benefits, stock, and equity, are not factored into this figure. Before continuing, a brief overview of the descriptive statistics of this paper will be given. Firstly, the dataset is skewed toward medium-sized firms, with 129,000 of the 132,000 entries coming from employees of medium-sized companies, 46 responses from employees at smaller companies, and 3,000 from employees of large companies. The data is also skewed toward in-person workers, with 106,000 of the respondents working in person, most of the remaining workers working remotely, totaling 26,000, and a small section of around 105 responders working hybrid. As for salaries, the mean salary is $158,000 when all currency is standardized to USD, and the median salary is $147,000. The standard deviation of the dataset is $74,000. Salaries range from a minimum of $15,000 to a maximum of $800,000, resulting in a range of $785,000.

3 Methods

Our research is primarily composed of the utilization of two different statistical tests – ANOVA and Post-Hoc. ANOVA checks to ensure if the average salaries are meaningfully different between the different groups that were established on the basis of remote work and company size. On the other hand, the Post-Hoc test iterates through every pair of groups to see which salaries are different and determines not only statistical significance of a difference but the extent to which the contrast is present. ANOVA models, also known as analysis of variance models, are a statistical test that assesses the difference between the means of more than two groups. In this research, due to a plethora of different groups, an ANOVA model perfectly fits as it compares the arithmetic means across each group, providing critical insight into if the differences can be attributed to random chance or meaningful difference. There are three main assumptions that must be met to conduct a proper ANOVA test: normality, homogeneity of variance, and independence. Essentially, the data must fall into a normal distribution, the variances across groups are the same, and each value is independent from one another. However, the homogeneity of variance assumption need not be met if performing a Welch's ANOVA test, which is the specific ANOVA test used in this study. Interpreting ANOVA test results is also important. The F-statistic indicates the ratio of variance between groups to variance within groups. Larger F values correlate to greater intergroup differences. Meanwhile, the p-value determines statistical significance by indicating the probability of obtaining the observed results, with a p-value of ¡0.05 being indicative of meaningful statistical significance. At the end of the process, an ANOVA test helps determine if there's a statistically significant difference between the groups' means. The programming language primarily used to conduct this research was R—a programming language for statistical computing and data visualization. In econometrics especially, R and Python are two of the most regularly used languages. Furthermore, Google Collaboratory, a cloud-based Jupyter notebook environment, allowed multiple users to simultaneously write code, facilitating the project.

4 Results

Before analyzing the results of the ANOVA test, researchers assessed the assumptions required to validate the tests. According to Figure 2, the p-values for both the Shapiro-Wilk Normality Test and Levene's Test for Homogeneity of Variance are low enough to be considered statistically significant, meaning that the data fails both tests. However, due to the sheer size of the data, normality can still be assumed, while homogeneity need not be met for a Welch's ANOVA test. Since the data was collected from different individuals when surveyed, researchers can also assume the independence necessary to proceed with the analysis of the tests.

```
        Shapiro-Wilk normality test

data:  residuals_sample
W = 0.91532, p-value < 2.2e-16

Levene's Test for Homogeneity of Variance (center = median)
           Df F value    Pr(>F)
group       8  82.311 < 2.2e-16 ***
      132856
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Coefficient covariances computed by hccm()
```

Figure 2: Shapiro Wilk-Normality Test and Levene's Test for Homogeneity Results

Moving forward with the ANOVA test, researchers can clearly see that all three variables in the test are statistically significant due to notably low P-Values in Figure 3. Therefore, remote work, company size, and the combination of the two all have an actual impact on the salary of an employee in the data science field. This result also informs us that it is appropriate to move forward with the analysis of effect sizes of the ANOVA test.

```
Analysis of Deviance Table (Type II tests)

Response: salary_in_usd
                             Df       F    Pr(>F)
remote_work                   2 220.894 < 2.2e-16 ***
company_size                  2  39.253 < 2.2e-16 ***
remote_work:company_size      4  16.621 1.262e-13 ***
Residuals                132856
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 3: ANOVA Test Results

When looking at the effect sizes in Figure 4, we noticed that the impact of these variables determined before were quite small due to the partial eta squared values of these all being less than 0.01. With that, the significance is still present, but any results with a Tukey's test would be seen as negligible differences and have little practical significance. Therefore, any Post-Hoc test would be impractical to carry out due to the effect size of the ANOVA test.

```
# Effect Size for ANOVA (Type I)

Parameter                | Eta2 (partial) |       95% CI
-----------------------------------------------------------
remote_work              |       2.36e-03 | [0.00, 1.00]
company_size             |       4.11e-04 | [0.00, 1.00]
remote_work:company_size |       2.75e-04 | [0.00, 1.00]
```

Figure 4: ANOVA Test Effect Size

5 Conclusion

The results of the study showed that there was indeed an impact on salary amounts between remote workers and in-person workers, as well as a small discrepancy in wages between employees of small, medium, and large companies. However, both of these differences were quite small and can be considered negligible, with partial eta squared values below 0.01 indicating minimal practical significance despite statistical significance. Due to the minor discrepancy in wage when considering remote work and company size, this study implies that tech workers should not consider remote work as a liability to their wages, nor should they favor certain-sized companies in the tech industry when considering wages as a main factor. This conclusion, when recognized, may have good implications for smaller tech startups, as workers realize they have similar monetary opportunities as compared to larger tech companies, which may help staff the smaller companies that desperately need experienced, educated employees. Furthermore, these findings suggest increased labor market efficiency in the digital economy, as geographic barriers to talent acquisition are diminished when compensation remains relatively consistent across work arrangements. The implications drawn from the data analyzed within this paper are subject to several limitations. Firstly, the most critical of which is the severe sample imbalance, with 97.7 percent of observations coming from employees who worked in medium-sized companies. This does undermine the test's ability to find relationships between the different company size categories. Additionally, the voluntary response bias reported in a self-reported survey may raise concerns with the data accurately representing the population. Additionally, the geographical concentration of data in the United States limits the international application of these findings. Finally, the failure to meet homogeneity and normalcy tests, while addressed due to Welch's ANOVA and the large amount of data gathered, may raise potential statistical concerns. While our analysis identifies associations between work arrangements and compensation, establishing causal relationships requires the longitudinal research designs outlined in our future studies section. Future studies should address the limitations of this paper by tracking the same individuals over several years as they transition from remote to in-person and hybrid working environments. This would allow researchers to control for certain interchanges like productivity and salary negotiation skills. Future studies should ensure that small and large companies are equally as represented as employees from medium-sized firms. Finally, future studies should account for alternative means of compensation such as equity, benefits, and professional development, as these factors may differ between remote and in-person employment. Contrary to

Adam Smith's compensating wage differentials theory, which predicts remote workers should accept lower wages due to the desirable aspects of flexibility and eliminated commutes, our findings suggest the modern data science labor market has reached an equilibrium where remote work no longer commands a significant wage penalty or premium. This development is important for employees, as it means that they won't be negatively impacted for choosing remote work, representing a fundamental shift in how the labor market values work arrangements in the post-pandemic economy. As remote work becomes increasingly normalized, this research contributes to understanding how traditional economic theories must evolve to reflect the changing nature of work in the digital age.

6 References

Admin. (2021, June 24). Least square method. *BYJUS*. https://byjus.com/maths/least-square-method/

chopper53. (2024). Data engineer salary in 2024 [Data set]. *Kaggle*. https://www.kaggle.com/datasets/chopper53/data-engineer-salary-in-2024

Kenton, W. (n.d.). What is analysis of variance (ANOVA)? *Investopedia*. https://www.investopedia.com/terms/a/anova.asp

Qualtrics. (2024, September 26). What is ANOVA (analysis of variance) testing? *Qualtrics*. https://www.qualtrics.com/experience-management/research/anova/

Yuyan, C. (n.d.). Research on the factors that influence wages—Take the example of the data science industry [PDF]. *ResearchGate*. https://www.researchgate.net/publication/387715059_Research_on_the_Factors_that_Influence_Wages-Take_the_Example_of_the_Data_Science_Industry