American Journal of Student Research

Research Article

How Unemployment Trends in California Vary by Scale: Where Linear Modeling Simplicity Obscures Structural Complexity

Sean Ting

Walnut High School. 400 Pierre Rd, Walnut, CA 91789, USA

ABSTRACT

This study explores long-term unemployment trends in California across two geographic scales: county and state levels. While state-wide unemployment rates are often used as an indicator for economic health, this research paper highlights how such indicators can hide local and regional variations. Using California labor force and unemployment data from 1990 to 2025, this paper critiques the limitations of linear models that assume uniform trends across different geographical scales. The findings reveal that several counties were not representative of the state's averages, deviating significantly from state-wide averages. Counties, such as Imperial County and Colusa County, significantly deviated from California state-wide averages with Imperial County having the most significant deviation, with a mean deviation of 15.59 percentage points throughout the period of 35 years. The study also found that simplistic models like linear regression models were not accurate at predictions for both the state and county levels, with R-squared values of less than 0.14. The findings highlight the importance of economic policies that account for local conditions to ensure more equitable regional support. The results also show that simplistic models can risk misinformed economic decision-making.

Keywords: Unemployment, Spatial Scales, Regional Disparities, Linear Model Limitations, Temporal Trends

INTRODUCTION

Unemployment is widely regarded as a critical and accurate indicator of economic performance and social well-being (7). However, the interpretation of unemployment data is highly dependent on the spatial and temporal scales at which it is being analyzed. In a large, geographic and economically diverse area like California (10), state-level unemployment statistics often fail to capture the whole picture, complexity, and variation that exists throughout counties in the state. These statistics conceal local labor distress, fluctuations tied to specific areas and industries, or persistent unemployment in particular areas (5).

This study addresses the need for a more focused understanding of labor market dynamics in California, by investigating unemployment trends across multiple geographic scales and over an extended historical period of 25 years. An example of these models concealing regional disparities in California is the agriculture-heavy Imperial Valley, which cannot be fully captured by linear

---

Corresponding author: Sean Ting, E-mail: ssting0708@gmail.com.

Received June 27, 2025**; Accepted** July 24, 2025
https://doi.org/10.70251/HYJR2348.34109117

models or state-level averages (1). Understanding how employment behaves across different scales is essential for developing effective policies that address both the general welfare and the needs of vulnerable communities.

This research paper pursues two primary objectives. First, it examines unemployment at two spatial scales: county and state, observing any notable differences and how the interpretation of unemployment may differ depending on the scale of geography. This comparison approach helps us uncover any spatial disparities in labor market conditions; disparities that would have been concealed when only observing the basic state-level averages. Regional differences are driven by several factors, including geographical area, industrial composition, degree of urbanization, and demographic shifts, which shape local employment and markets (9). The second objective of the study is to evaluate the effectiveness of linear regression models in capturing unemployment trends across geographic scales and assess their limitations.

METHODS AND MATERIALS

The dataset used in this study is the *Local Area Unemployment Statistics (LAUS) Annual Average,* which is a record of unemployment and labor data for California and its counties from 1990 to 2025 (6). This dataset serves as the basis of the analysis conducted throughout this paper and contains data that allows us to compare different regions within the state. The dataset includes each area name (ex, "Los Angeles County," "Orange County") and its geographic designation. The main variables used in the dataset are the total labor force, total employment, and unemployment rate, which are reported for each county in the state throughout the whole duration from 1990 to 2025 and are clearly labeled.

The analysis first begins with spatial patterns, which are visualized using line graphs and choropleth maps. First, line graphs of annual employment rates are illustrated to provide more context to the differences in historical unemployment rates between county and state-level data. Then, choropleth maps help illustrate unemployment rates across the state, allowing for a visual comparison of labor market disparities in different geographical regions that would have been concealed by solely using state-level data. Furthermore, the study then conducted a deeper analysis of select counties that significantly diverged from state-level trends.

In addition, graphs were fitted with linear models, and illustrated with recession periods to highlight when the predicted trends diverge from observed labor market behavior in economic downturns. To evaluate how well linear regression models reflect real-world economic dynamics, the model was applied to the data, and the trend lines were overlaid on the graphs with periods of economic downturn and recession.

$$y_1 = \beta_0 + \beta_1 x_1 + \varepsilon_1 \qquad [1]$$

The linear trend model in my graphs took the standard form as shown above (11). y_1 represents the predicted unemployment rate, β_0 represents the y-intercept when x is 0, β_1 represents the change in unemployment rate for every year, x is the year, and ε is the error term. To get the β_1 in the equation, I used the Ordinary Least Squares Formula for it, which is:

$$\beta_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \qquad [2]$$

The Ordinary Least Squares regression equation was used to account for all the data points within the dataset, rather than choosing two specific points to obtain a slope, which would be inaccurate. The Ordinary Least Squares

Table 1. Variables in the LAUS Dataset

Variable	Type	Definition
Area Name	Categorical	The name of the geographic region; ex: Los Angeles County
Area Type	Categorical	The type of geographic category: State, County
Year	Numeric	The year in which the data is for.
Labor Force	Numeric	The total number of people available to work.
Employment	Numeric	The total number of people in the labor force who are employed.
Unemployment	Numeric	The total number of people in the labor force who are unemployed.
Unemployment Rate	Numeric	The percentage of the labor force that is unemployed.

regression models were implemented separately for each geographic scale. This allows for a direct comparison of model performance across counties regarding the statewide trend. Cross-validation was not used because the goal of this research is not to assess the predictive capability of the models, but rather to analyze historical unemployment trends (2). The Ordinary Least Squares regression equation is one of the most widely used methods in linear regression, and by applying it to this situation, we can reveal the subtle differences in scale (4). x_i represents the year the unemployment data was observed, $\overline{x}$ represents the mean of all the observed years, y_i represents the observed employment rate, $\overline{y}$ represents the mean unemployment rate across all of the observed years, and Σ represents the sum of all the values. Each model was trained using annual employment data through the common time frame in the data.

These graphs display both the predicted linear models and actual unemployment rates, with the blue shaded bands being the years of recession. R-squared was calculated to determine fit. Then, the residuals were plotted through the course of the temporal period in the data to determine where the models failed to align with the actual unemployment rates. The residuals are the differences between predicted and actual unemployment rates from the linear models (8).

$$e = y - \hat{y} \qquad [3]$$

The e represents the residual value, y is actual value, and $\hat{y}$ is the expected value. This residual analysis reveals where the model succeeds or fails, showing greater volatility and mismatch in the models (7).

The findings demonstrate how spatial aggregation can smooth over local variation, potentially distorting trend interpretations in cases with simple model complexity and inconsideration of geographic scales.

RESULTS

Each of the described methods is applied to the dataset to uncover the limitations of linear models. The detailed results of these approaches are presented in the following subsections.

Unemployment Rates Across Geographic Scales: Why It Matters

Figure 1 provides an overview of average unemployment rates across the state. A consistent pattern is visible, where county-level averages are always higher than state averages. Smaller counties have more local variation, so the state data should be lower accordingly. The county data captures more volatile spikes, often magnifying the severity of downturns in economically vulnerable areas. The result is a notable spread in unemployment rates. In Figure 1, the graph displays three major unemployment spikes: the mild recession in the 1990s, the 2008 financial

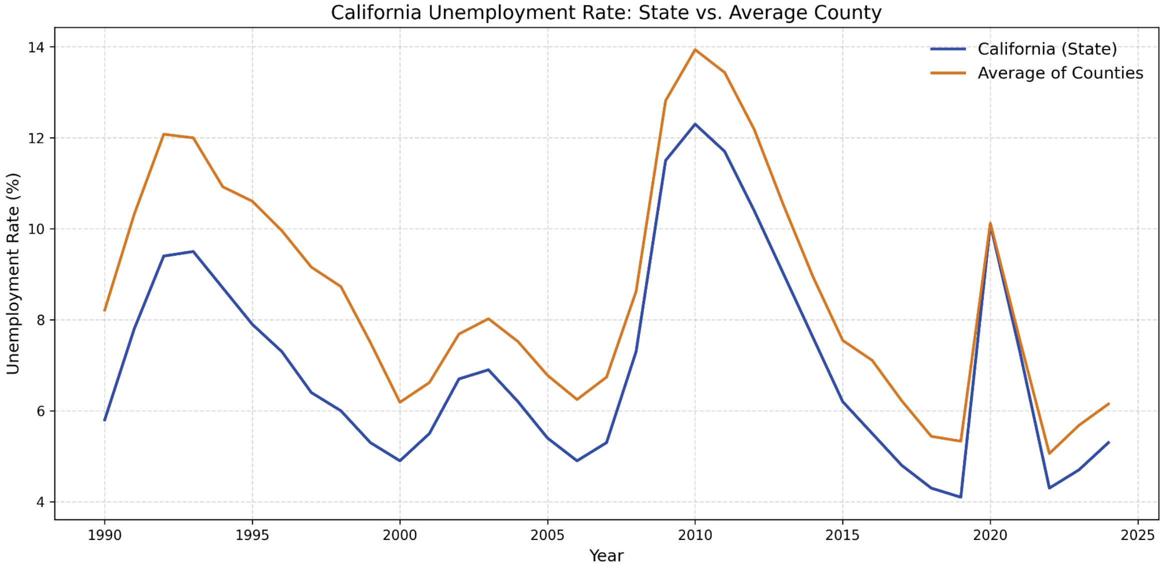

Figure 1. Overlaid Unemployment Rate Across Geographic Scales in California (1990-2025).

crisis, and the COVID-19 pandemic in 2020, where each of these spikes can be seen in both spatial scales; however, the magnitude of the increase is sharper at the county level. These graphs highlight the importance of spatial scale when analyzing the labor market.

One common misconception is the question of why there is a difference in unemployment rates despite the data encompassing all counties in the state. The misconception can be explained because the state-level unemployment rate is calculated using a weighted average (the larger the region, the more weight). In contrast, the county-level unemployment rates are calculated on an unweighted scale, which means that each county has the same influence on the unemployment rate regardless of its size. The smaller high-unemployment counties can skew the average more than they would in the state-level unemployment rate, which explains the differences.

The choropleth maps of California county-level unemployment rates for 2008 and 2020 reveal clear differences in distress. Each choropleth map presents a picture of the unemployment, showing various regional disparities. In Figure 2, the 2008 map illustrates the impact of the Great Recession, which triggered widespread job losses throughout the state. The counties that were affected the hardest were Imperial County, with over 22% as its unemployment rate, and Colusa County with a 13.8% unemployment rate.

Figure 3 illustrates the unemployment rate across the state during 2020, the year of the COVID-19 pandemic, where the effects had impacted the entire state, as reflected by the widespread unemployment spikes across California. However, despite the many variations in resilience, which saw the northernmost counties have a decrease in unemployment rate compared to the 2008 recession, the two counties, Imperial County and Colusa County, remained consistent with the highest unemployment rate, with Imperial having a 23.2% unemployment rate and Colusa County having a 15.9% unemployment rate.

Imperial County and Colusa County: Persistent Epicenters of Unemployment

The choropleth map consistently highlights Imperial County and Colusa County as chronic epicenters of joblessness during economic crises. Imperial County, located in California's southern area, has an overreliance on seasonal agricultural labor and limited industrial diversification (1). Employment is incredibly volatile, driven by seasonal cycles, and is often temporary. Colusa County faces the same challenges; its economy is also heavily dependent on agriculture, which is vulnerable to labor reductions, especially during economic crises. Their persistence as areas with high unemployment highlights the geographic economic unevenness in California and underscores the need for targeted economic policies to

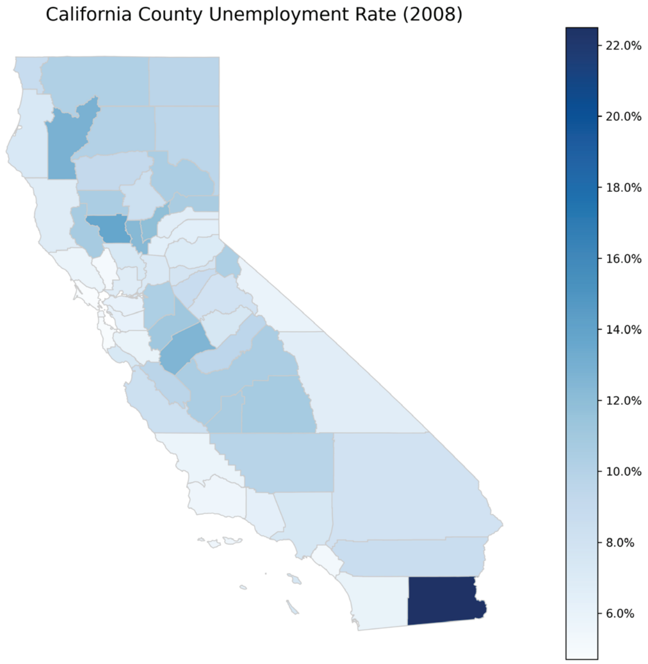

Figure 2. California County Unemployment Rate in 2008.

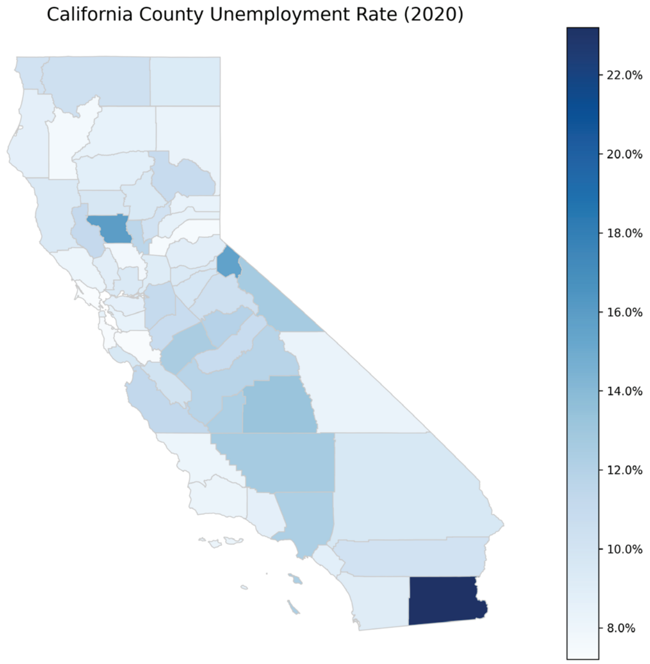

Figure 3. California County Unemployment Rate in 2020.

alleviate these counties' overreliance on single-sector industries and temporary jobs. This could not have been discovered with solely state-level data and without considering different geographic scales, which highlights the importance of examining regional disparities at various scales to help those specific areas.

Inadequacy of Linear Models in Analyzing Unemployment Trends

In this next section, the reliability of linear models will be tested in their ability to predict unemployment trends in California regardless of scale. As a result, the study will be using the historical data from 1990-2025 combined with the results of the linear model to find the county with the largest standard deviation from the state average rate (Imperial County) and the county with the smallest standard deviation (Nevada County) along with the state unemployment rate data to see if linear models are good predictors of unemployment rate. Those results are displayed in Table 2. The counties with the largest and smallest standard deviations were chosen to represent benchmarks in determining if the models were good fits, depending on the severity of standard deviation. With that in mind, the county that had the largest standard deviation was 3.899 deviations from the state, while the county with the smallest standard deviation was 0.336 deviations from the state averages.

Through the graphs, we can already start to see some

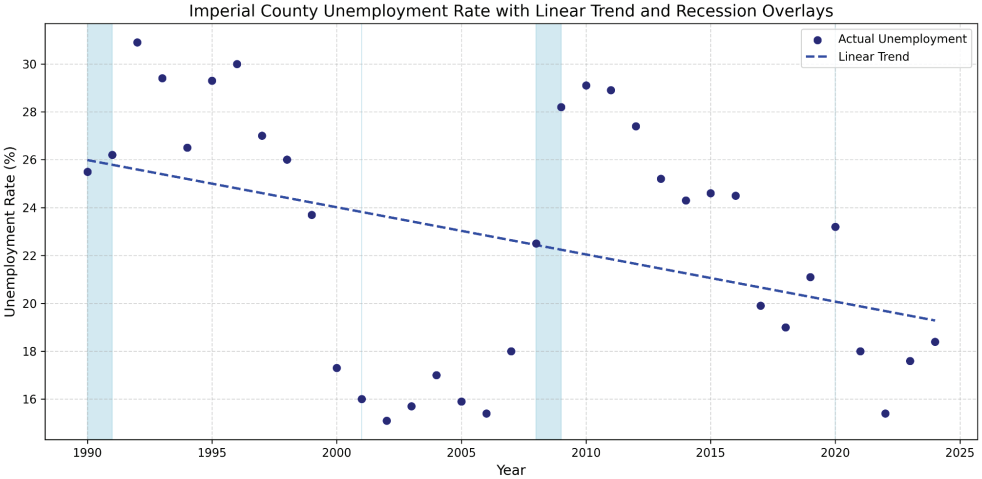

Figure 4. Imperial County's Unemployment Rate Overlaid with Linear Trend and Recessions (1990-2025).

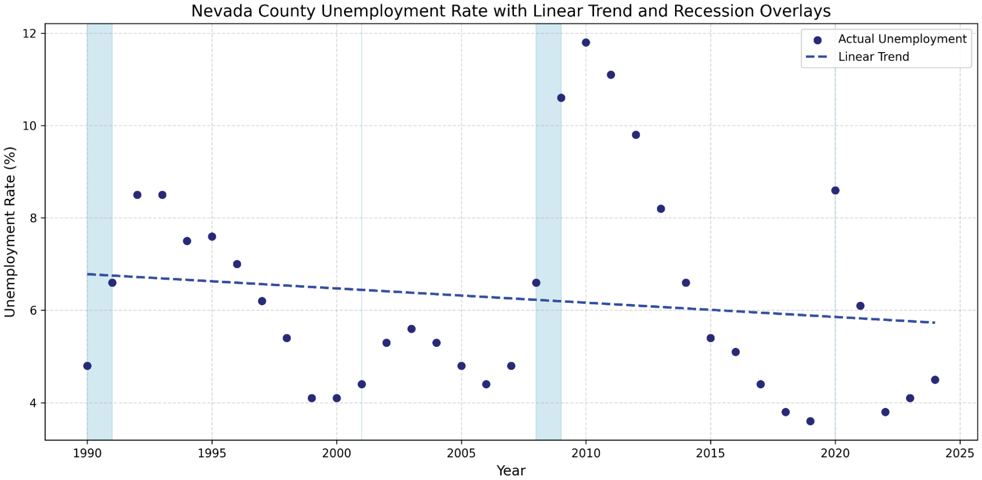

Figure 5. Nevada County's Unemployment Rate Overlaid with Linear Trend and Recessions (1990-2025).

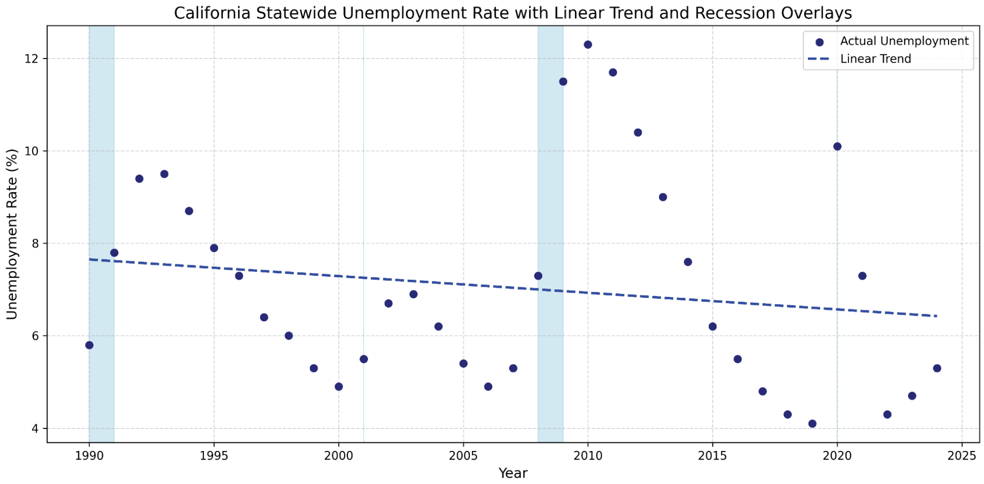

Figure 6. California's Unemployment Rate Overlaid with Linear Trend and Recessions (1990-2025).

Table 2. Results from Linear Trend Models Across Counties with Varying Deviation from State-Level

Name	R^2	Variance	SD	Mean	Deviation from State
Imperial County	0.1539797608	25.723395918	5.0718237271	22.634285714	3.8994860990
Nevada County	0.0203861680	4.7761632653	2.1854434939	6.2571428571	0.3363795604
California	0.0265651809	4.9823346938	2.2321143998	7.0371428571	0

of the shortcomings of the linear regression model. There are various periods during which the unemployment rate changes in a non-linear manner, which significantly affects its fit. In Figure 5, Nevada County exhibits a relatively stable unemployment trend, with the linear model tracking closer, with smaller residuals because of its diversified economy, which helps cushion the impact of downturns. As a result, the deviations from the graphs are smaller, with the peaks in 2010 and 2020 (2020 is not labeled because the recession lasted only 2 months), and an unemployment deviation of around 8% (3). Figure 4 depicts Imperial County, which, on the other hand, illustrates the limitations of a linear model in rural, industry-dependent regions because of its agriculture-based economy and sensitivity. The deviations in the periods of recession are much more profound. During periods of economic downturn, Imperial County exhibited much more volatility in deviation from 18% at the start of the 2008 recession to 29% right after, marking an 11% increase. The COVID-19 pandemic also significantly impacted the unemployment rates in Imperial County. For Figure 6, California displays minor deviations in periods of recession, like in 2008, where it deviated around 4% because state-level data captures a significantly larger population compared to the counties and has more minor variance.

The linear model highlights significant differences with regional and economic disparities shown in Table 2. R-squared, which measures the proportion of variance that the model can explain, is relatively low across all three geographic scales. Imperial County has an R-squared of approximately 0.154, Nevada County has an R-squared of roughly 0.02, and California has an R-squared of approximately 0.027, indicating that linear models fail to capture the full complexity of the unemployment trends. Imperial County has by far the highest rate; however, the linear model still only explains about 15.4% of the variation in the model, indicating its unsuitability. When examining variance and standard deviation, Imperial County again stands out (SD: 5.072, Variance: 25.723) like previously, with its numbers considerably higher than both Nevada County (SD: 2.185, Variance: 4.776) and California (SD: 2.232, Variance: 4.982), showing its high volatility. These findings demonstrate that simple linear regression models are inadequate for modeling California's unemployment trends even if applied to different geographic scales and across different sides of the unemployment rate scale.

Residual Plots From Linear Model

The residual plots from the linear unemployment model indicate that the smaller regions tend to have larger and more erratic residuals compared to the state-wide level. In Figure 8, Imperial County displays this as it has the most erratic residuals. All of the models exhibit large residuals during recessionary periods, such as in 2008 and 2000. However, concerning the purpose of this study, which is to demonstrate that linear models are not a viable option, the focus is on whether the residuals are scattered around zero, which would indicate that the model would be a good fit. However, in all three regions, we see noticeable deviations that suggest model misfit. These residual plots illustrate the theme that linear regression models cannot model unemployment because they fail to adapt to the complexities of actual real-world economic behavior, such as periods of economic turmoil.

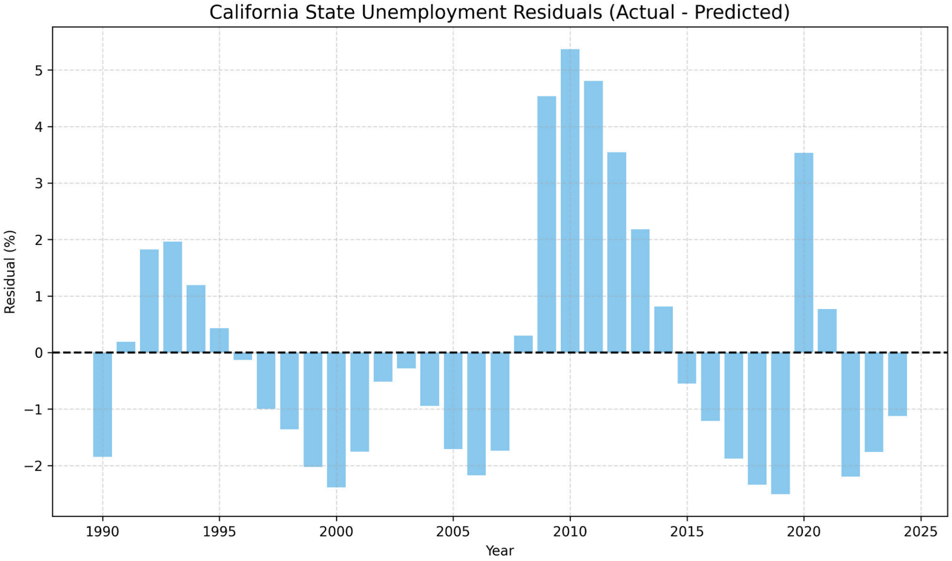

Figure 7. Residuals from Linear Model of the Unemployment Rate in California (1990-2025).

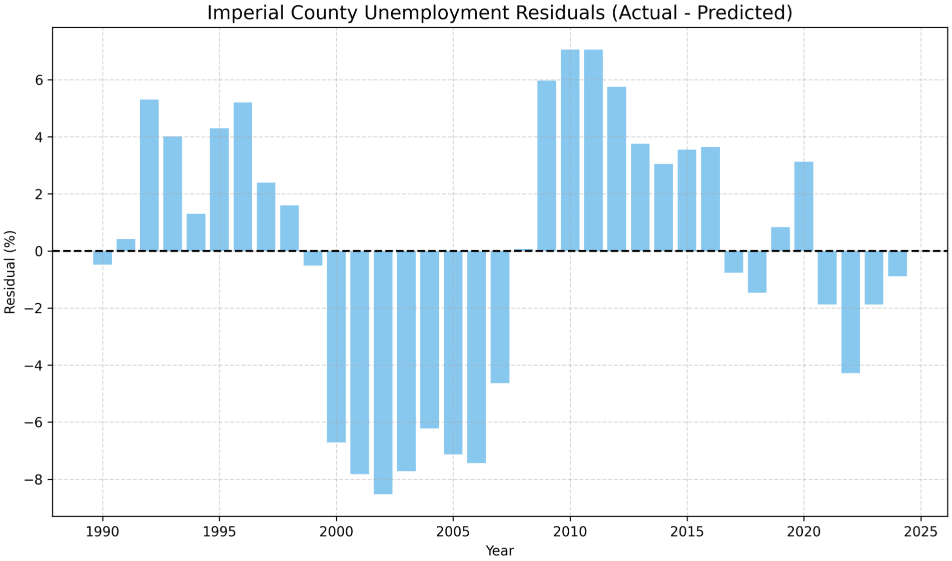

Figure 8. Residuals from Linear Model of the Unemployment Rate in Imperial County (1990-2025).

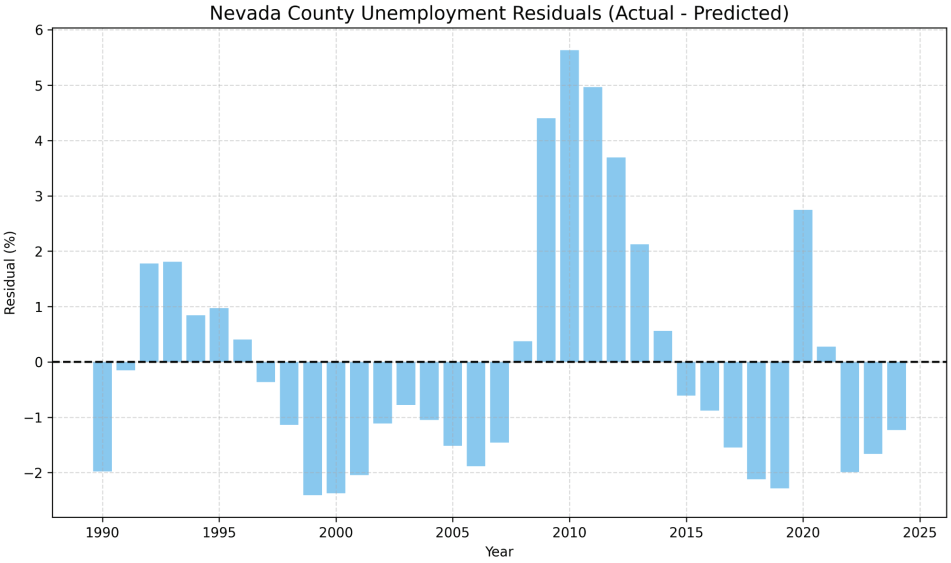

Figure 9. Residuals from Linear Model of the Unemployment Rate in Nevada County (1990-2025).

CONCLUSION

This study set out to examine how unemployment trends vary across different geographic scales in California and to assess the efficacy of linear regression models in capturing these trends. Through the study, it became evident that both the structure of unemployment and the effectiveness of linear modeling are dependent on the geographic scale. One key takeaway from this study is that local labor markets do not behave uniformly at larger geographic scales. In conclusion, this study demonstrates that state-level trends are not effective in shaping accurate labor market policies and must be informed by more specific spatially sensitive analysis. The findings show that while state-level trends do provide a general picture in non-recession periods, they often conceal substantial regional disparities. The individual counties offer critical insights into regional labor behavior, and using purely state-level data would overwrite these insights. This study also demonstrated the limitations of linear models when used to analyze and model historical unemployment data. In summary, this study reinforces the idea that state-level labor market data trends analysis is too general to guide effective economic policies that could help both metropolitan areas and smaller counties, and more complex methods and models should be used to model unemployment data in California accurately.

ACKNOWLEDGEMENT

I extend my greatest appreciation to the Bureau of Labor Statistics for their continued work in maintaining and providing public access to the Local Area Unemployment Statistics (LAUS) program. The availability of this large-scale dataset has been incredibly critical to my research.

DECLARATION OF CONFLICTS OF INTEREST

The author declares that there are no conflicts of interest regarding the publication of this article.

REFERENCES

1. Adams J. Fighting for the farm: Rural America transformed. *University of Pennsylvania Press*. 2002. ISBN: 978-0812218305. https://doi.org/10.9783/9780812201031
2. Bates S, Hastie T & Tibshirani R. Cross-Validation: What Does It Estimate and How Well Does It Do It? *Journal of the American Statistical Association*. 2023; 119 (546): 1434–1445. https://doi.org/10.1080/01621459.2023.2197686
3. Center on Budget and Policy Priorities. (2024, April 3). *Chart book: Tracking the recovery from the pandemic recession*. https://www.cbpp.org/research/economy/tracking-the-recovery-from-the-pandemic-recession (accessed on 2025-06-06)

4. Kuchibhotla AK, Brown LD & Buja A. Model-free Study of Ordinary Least Squares Linear Regression. 2018. arXiv:1809.10538.
5. Kuhn M, Manovskii I & Qiu X. The geography of job creation and job destruction. *Federal Reserve Bank of Minneapolis, Institute Working Paper.* 2024; 85. DOI: 10.21034/iwp.85.
6. Local Area Unemployment Statistics (LAUS), annual average - dataset - California open data. (2019, September 26). https://data.ca.gov/dataset/local-area-unemployment-statistics-laus-annual-average
7. Murphy KM & Topel R. Unemployment and nonemployment. *The American Economic Review.* 1997; 87 (2): 295–300.
8. Pierce DA & Schafer DW. Residuals in generalized linear models. *Journal of the American Statistical Association.* 1986; 81 (396): 977-986. https://doi.org/10.1080/01621459.1986.10478361
9. Porter M. The economic performance of regions. *Regional Studies.* 2003; 37 (6–7): 549-578. https://doi.org/10.1080/0034340032000108688
10. Price T & Smith S. Understanding unemployment across California counties. *Economic Inquiry.* 2002; 40 (1): 12–30. DOI: 10.1093/ei/40.1.12
11. Zou KH, Tuncali K & Silverman SG. Correlation and simple linear regression. *Radiology.* 2003, June 1; 227 (3): 617–622. https://doi.org/10.1148/radiol.2273011499