# Measuring informality in California

Allegra Saggese

December 2025

# 1. Introduction

*Research question.*    How large is the informal economy in the United States? While the US is an engine of growth with strong formalized economy across a number of advanced and highly productive sectors, the economy has struggled with increasing inequality and a number of economic shocks. Macroeconomic shocks are beyond the control of ordinary workers and even policymakers, but individuals may adapt to these conditions through reverting back to informal work during tough times. This idea, previously developed and coined *The Hugo Effect* captures the idea that during recessionary periods, people undertake more informal work.

*Answer to the question.*    I use the case study of California to explore a consumption-tax-based measurement for quantifying informal economic activity in the economy. California is an ideal setting as one of the largest economies in the world. Additionally, it has a variety of taxes at the state and local level, applied to consumer goods. It has a mix of both urban and rural settings, as well the highest poverty rate in the United States, at 17.7% in 2023 (California Budget & Policy Center 2024).

*Positioning in the literature.*    Measuring the informal economy is not a new exercise, and there currently exist several methods to answer this question in a developed economy context.

# 2. Methods

*Data.*    Data is collected from both state-level sources and federal sources. Both contain state level data, although, where further disaggregation at the county level is available, it is used to refine estimates.

*Methods.*    The key equation we find is one which estimates the current tax compliance, or $\gamma_t$. In each year, $t$, we find estimate the potential tax base, $\tau_{it}^{potential}$ at the county level, and compare it to the reported tax collected, $\tau_t^{real}$. Note that $\gamma$ cannot be further disaggregated to the county level due to data limitations. The most comprehensive and accurate measure of the real tax base is the recorded estimated gross sales, which includes all consumer spending sales receipts, provided by the US Treasury FRED data collection. From this data we are able to estimate:

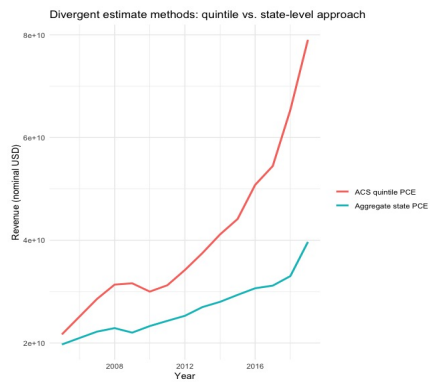$$\gamma_t = \frac{\sum \tau_t^{real}}{\sum \tau_{it}^{potential}}$$

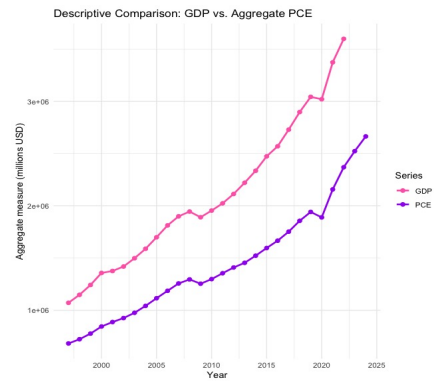FIGURE 1. Both estimation methods (transaction-based, PCE-based) for tax base



FIGURE 2. Total consumer spending (PCE) and GDP in California

*Descriptive.* Results are able to gather variation at the county-level, although our final results will be presented in aggregate at the state level.



FIGURE 3. PCE (2023)



FIGURE 4. Median income (2023)



FIGURE 5. Revenue estimate (2023)



FIGURE 6. Local sales tax (2023)

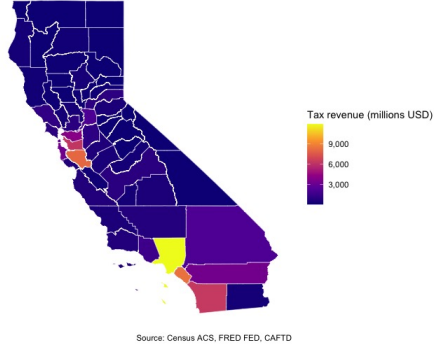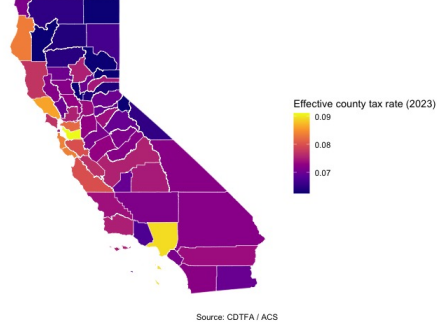*Projected data.* We start by estimating $\tau_{it}^{potential}$ or the sum of all taxable consumer spending in the state. Steps are as follows:

- Using BEA quintile level disposable income and expenditure estimates for California, we classify each sector of spending within the survey as taxed or non-taxed.

$$S_{q,t} = \mathbb{1}\{\tau_t \neq 0\}$$

- The percentage of the share of expenditure (μ) that is taxable is then calculated for each quintile, across all years in our sample. This gives the figure for the estimated share of spending by household, depending on quintile.
- Concurrently, each counties in California is then aggregated for the same time period,

3

and given the distribution around the annual median income in California, counties are assigned to quintiles ($q$).

$$\sum S_{q\mathbb{1}\{\tau_t \neq 0\},t} \rightarrow \frac{\sum S_{q\mathbb{1}\{\tau_t \neq 0\},t}}{\sum exp_{q,t}} = \mu_{q,t}$$

- Each county ($i$) is then matched with their quintile share for taxable expenditure, giving us a percentage of total personal consumption expenditures that are likely to be taxable in a particular county. This disaggregation provides more richness to our investigation in that we imagine areas with higher income may spend more of their disposal (a higher share) on luxury or taxed goods, as opposed to lower income counties.
- We then merge based on a *county · year* pairing with our personal consumption expenditure (PCE) data. This data is provided by the American Community Survey (ACS) in both one-year and five-year study increments. It is a cross-sectional panel data that provides either five-year rolling averages or one year samples, providing per capita PCE ($\phi$) at the county level for all of California. The two samples are combined where counties with less than a population of 60,000 individuals are present. This included a list of twelve counties, and as such, the five-year survey data was then used for those missing counties. Additionally 2020 data was not collected in the one-year survey, and the five-year average was then used to estimate the PCE at the county level for this year.
- Given all data were able to be merged, the final data four county disaggregation, the local effective tax rates, were provided. Local tax rates ($\upsilon$) were only available for the most recent year (2025), and backfilling was required. The median effective tax rate ($\Upsilon$) for California was available in time series format, and this rate of change was used to backfill county level rates. To create a linear trend, I took the distribution of the 2025 local tax rate, around the median and calculated the percentage point difference between a particular county and the median. This was then used as the distance to the median to follow along the same state-level growth trend for all tax rates. This ensured certain counties were not overweighted in specific periods.
- Finally, county-level PCE, was combined with:
    - the incidence of tax rate for a county was matched given their quintile ranking across California
    - the population data ($\rho$) was then merged, to provide an aggregate sum of taxable transactions
- The final results was to get a value for the taxable base for a specific county in each year in our sample, given the county's relative income quintile and unique annual rate of tax, inclusive of broad base state consumer taxes and any adjustments to the local

tax code.

The resulting equation for the **estimated tax base** (spending subject to tax) looks like the following:

$$\mu_{q,t} \cdot (\phi_{i,t} \cdot \rho_{i,t}) \cdot \mathbb{1}[q] = \beta_{i,t}$$

The resulting equation for **estimated tax revenue** from consumer spending looks like the following:

$$\mu_{q,t} \cdot (\phi_{i,c,t} \cdot \rho_{i,t}) \cdot \upsilon_{i,t} \cdot \mathbb{1}[q] = \Pi_{i,t}$$

This provides us with the total personal consumption of individuals at the county level, with a share of their disposal income subject to taxes, depending on the county's relative income, taxed at the local effective rate. From this we get an estimate for the total tax base, and the revenue ($\Pi_{it}$) that should result from each county. Note that $\mathbb{1}[q]$ is an indicator function for the vector of the quintiles possible for each county. From this we are able to estimate the revenue, and aggregate it the state level to compare to our *actual* observed tax revenue collection data.
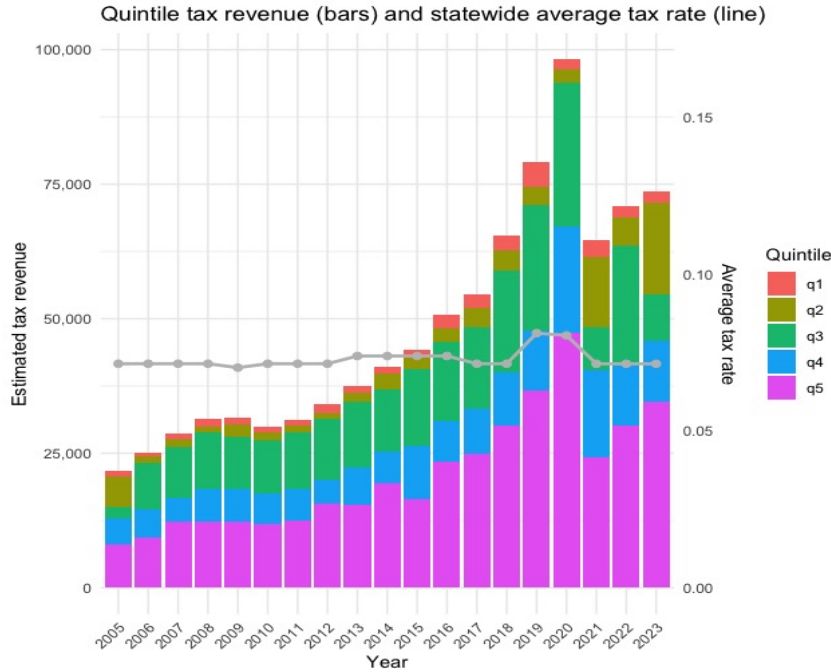


FIGURE 7. Quintile level estimates of taxable consumption by households (left axis), plotted against average effective consumption tax rate (right axis)

*Observed data.* The observed data is taken from the California Tax Revenue Authority (CATR) and the Federal Reserve state-level published statistics on total tax revenue collection. Using both forms, we estimate different measures of actual revenue collected

(Federal Reserve Bank of St. Louis 2024; California Department of Tax and Fee Administration 2025). The data is taken from 2008 until 2024. Description of the data for the full sample period is available in 9, while the results of the measure are presented for 2008 through 2019, to exclude the idiosyncratic nature of the COVID-19 shock. Data from the CATR is comprehensive across all recorded transactions in the state, with specific classifications across industrial categories. Categories to generate the base of taxable consumer income are SIC codes that correspond to industries of consumer goods, and exclude thaose with business-to-business, wholesale or non-taxable goods. Next, the FRED data provides an estimate of all tax revenue from consumer taxation, excluding other large categories of state revenue such as income tax and property tax.

### 3. Findings

*Descriptive findings.* Measurement of the informal economy yield a fairly low estimated of the size of the informal economy in California, reflective of wider trends in the United States. Although, volatility in the measure of informal activity yields further questions on the temporal dimension of informality - it may be more temporary or short term, given the fluctuations - although findings do not suggest strong alignment with the Hugo Effect.

*Results.* My findings suggest there is no clear pattern between economic recessionary periods and the level of informality activity, given the measurement techniques trialed.
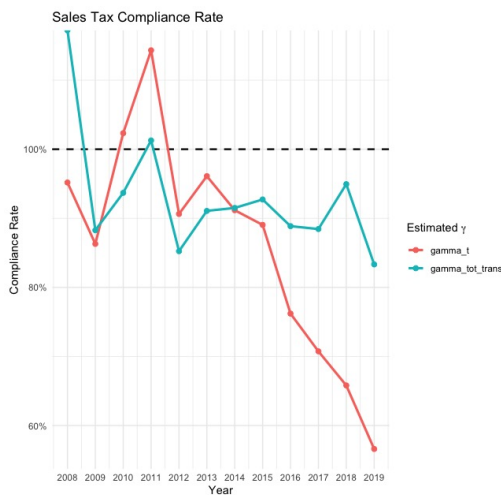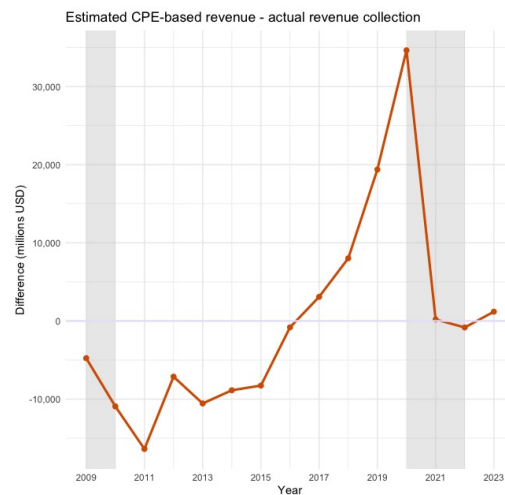


FIGURE 8. Share of realized revenue collected ($\gamma$)



FIGURE 9. Difference in projected vs. annual revenue

*Challenges.* One of the major challenges with measurement of unobserved data, is the potential interaction with measurement error, which was addressed in careful choices

within the data selection process,

*Robustness checks.* This paper presents two estimation methods for the levels of currently collected tax revenue. With both methods, there is notable differences in the net annual estimates for collectable tax revenue from consumption, which indicates that fluctuations in the share of collected tax is likely attributable to measurement error or measurement choice. Looking at
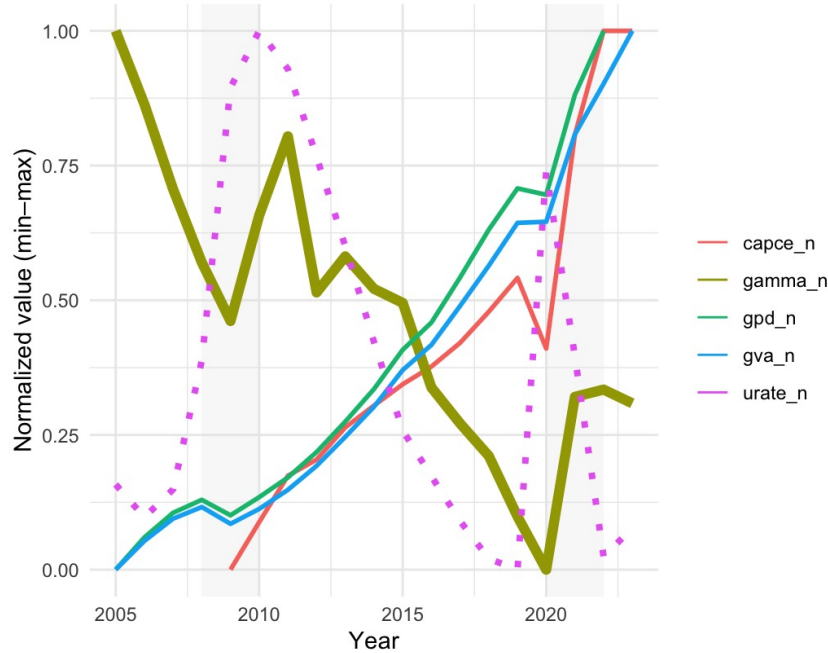


FIGURE 10. Normalized variables of interest over time

## 4. Conclusion

Overall findings suggest further refinement and study is needed to draw sincere conclusions from the analytical exercise. Future research should should create a strong classification of NAICS-level activity to consumer spending. This may involve classifying the share of consumer spending per NAICS-code based on aggregate data, such as provided by the Bureau of Economic Analysis (BEA). While this was outside the scope of this work, it is wroth considering the trade-off between the accuracy or sector of spending and by consumers with the disaggregation approach with emphasis on income distribution.

# References

California Budget & Policy Center. 2024. "California's Persistent Poverty Crisis: 2024 Rates Remain Alarmingly High." `https://calbudgetcenter.org/resources/californias-persistent-poverty-crisis-2024-rates-remain-alarmingly-high/`. Accessed December 2025.

California Department of Tax and Fee Administration. 2025. "State and Local Sales and Use Tax Collections (SUTSStateCollNoPermits)."

Federal Reserve Bank of St. Louis. 2024. "General Sales and Gross Receipts Taxes: California (CATSXGRTAX)."