Measuring informality in California

Allegra Saggese

December 2025

Measurement of the informal economy is conducted using an adapted approach to the EVADE (VAT tax based) measure of informal activity, such that the difference between real tax revenue on consumption is compared to a synthetic tax revenue base. The synthetic base is constructed from household survey data for the period from 2005 to 2024, with the period from 2020-2024 being dropped due to the idiosyncratic nature of the data. This measurement technique, while novel, does not corroborate previous evidence that the informal economy in developed countries grows with volatility, and follows an inverse relationship to formal work. These results suggest significant failures of the methodology to capture the real components of informal work, or a significant measurement error. Further analysis and refinement of the approach would be necessary.

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 (Pappadà and Rogoff 2023; Bracha and Burke 2017).

Answer to the question. Using the case study of California, we explore a consumption-tax-based measurement for quantifying informal economic activity in the US 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). Overall, California serves as a good case study for this exercise to test the efficacy of using a tax-based approach to measure the informal activity in California.

Motivation. The size and nature of informal activity in the economy in the United States is of interest as it encompasses all activity that could be taxed, and is not (Marshall et al. 2023). This is a revenue deficit for the government, and is symptomatic of an economy where tax evasion is either not enforced properly or has integrated into the social norm of work within a country. While the US is a developed nation, its economic norms are unique: people in the US continue to carry cash, and there is a persistent share of individuals (about 15%) who have a preference for carrying cash tender (Bayeh et al. 2025). Next, the US prosecutes very few cases of criminal tax evasion, around 360 a year (US Sentencing Commission 2025). While tax evasion prosecution is low, and the tax avoidance behavior of the highest income earners is better understood (Slemrod 2007; Zucman, Alstadsæter, and Johannesen 2019), there is a strong motivation to understand if low levels of tax avoidance are occurring. Finally, recent evidence using newer approaches including the use of household survey data (Abraham and Houseman 2019) and VAT data (Pappadà and Rogoff 2023) find that workers engage in informal activities, in developed countries, particularly in response to recessionary events.

Positioning in the literature. On average, that 13% of economic activity in developed countries occurs in the informal sector(Restrepo-Echavarría 2018). Measuring the informal economy is not a new exercise, and there currently exist several methods to answer this question in a developed economy context including macroeconomic econometric methods, as well as models of the interaction between informal and formal economy (Schneider 2023). While (Schneider 2023) provides an overview of the different methodologies, the key methods are:

a. currency demand approach
b. MIMIC approach
c. VAT-measure approach
d. other micro-survey methods.

With currency demand, the size of the informal economy is an estimate based on how much cash is demanded in the economy, given that much informal work occurs with cash payments (Quiros-Romero, Alexander, and Ribarsky 2021). MIMIC uses macroeconomic variables to determine the size of informal activity based on statistically significant drivers (Restrepo-Echavarría 2018; Valletta, Bengali, and van der List 2020). The VAT approach, is more novel, and uses the difference in observed and actual collected VAT to measure economic activity, and other micro-survey methods may use self-reported measures of informal or secondary work at the household level over time. One significant takeaway from all the papers which measure informality, is that there is some relationship between both formal and informal employment levels (Leyva and Urrutia 2020) and that informal activity increases with labor force volatility (Restrepo-Echavarría 2018).

While each of these measures is able to find some share of informal activity occurring, each has drawbacks. This paper is not a comparative analysis, so it ignores the fundamental tradeoffs, and instead builds off the novel contribution to the literature by (Pappadà and Rogoff 2023). With a simplified approach, we will apply the VAT methodology to the US case, where sales tax is applied at different levels to different products of goods, and look solely at consumer spending. But, consumer spending alone is insufficient. Consumer spending, reported by households, should be compared with actual recorded transactions in the economy, to account for current levels of economic activity. Therefore, I propose a methodology to compare actual collected tax revenues on consumption goods with reported consumer spending to capture whether consumers are spending in cash-only or non-taxed transactions. The extension of this work would then be to assess if the measure varies with other economic indicators.

2. Methods

Data. Data is collected from both state level and federal sources. Both contain state level data, although, where further disaggregation at the county level is available, it is provide a more granular measurement. Given this approach using a consumption based measure of tax evasion, we ignore other measures that may indicate discrepancies between time spent working - such as a comparison between time use survey data and reported employee hours. This is because the hours-based measurement is outside the scope of this work, and instead we focus on identifying evasive transactions. This requires data on the current consumer tax base from the Bureau of Labor Statistics and the US Census (U.S. Bureau of Labor Statistics 2024; U.S. Census Bureau 2023). This data allows for an estimated of self-reported household level spending and income.

Then, we are able to compare household level spend against real tax collected. Real tax data is provided by the St. Louis Federal Reserve Bank, as well as The California Tax Revenue Authority (California Department of Tax and Fee Administration 2025a,b; Federal Reserve Bank of St. Louis 2024). Additional data including California state-level GDP, estimates for county level tax rates and for unemployment is provided by the Bureau of Economic Analysis, California Department of Finance, and the Internal Revenue Service (U.S. Bureau of Economic Analysis 2024; California Department of Finance 2024; Internal Revenue Service 2024).

Below is a descriptive depiction of both estimates for building the synthetic tax base in 1 in the left hand side panel. While on the right in 2 we see a comparison of total GDP and total personal consumption spending compared over time.

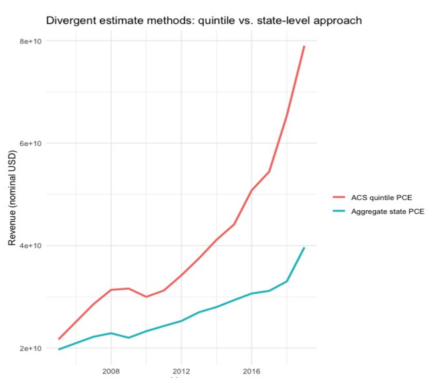

FIGURE 1. Both estimation methods (transaction-based, PCE-based) for tax base

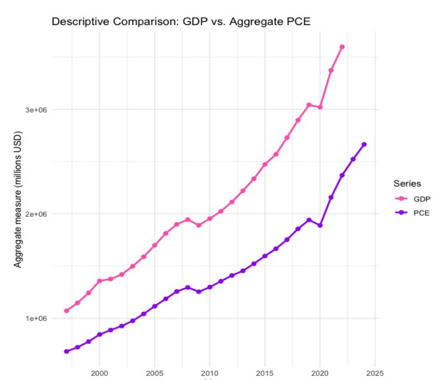

FIGURE 2. Total consumer spending (PCE) and GDP in California

Methods. The key equation we find is one which estimates the current tax compliance, or γ_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, τ_t^{real}. Note that γ 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:

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

Descriptive. Results are able to gather variation at the county-level, although our final results will be presented in aggregate at the state level. The below figure demonstrates the difference in personal consumption expenditure, median income, sales tax rates, and revenue estimates for 2023 across California counties (**?**Internal Revenue Service 2024; U.S. Census Bureau 2023).

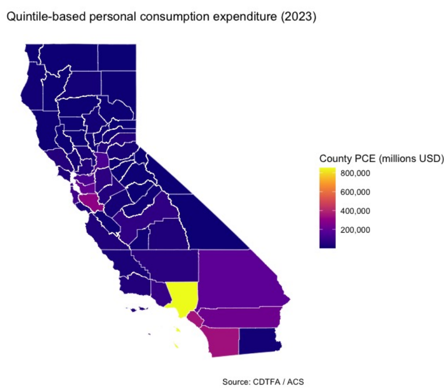

FIGURE 3. PCE (2023)

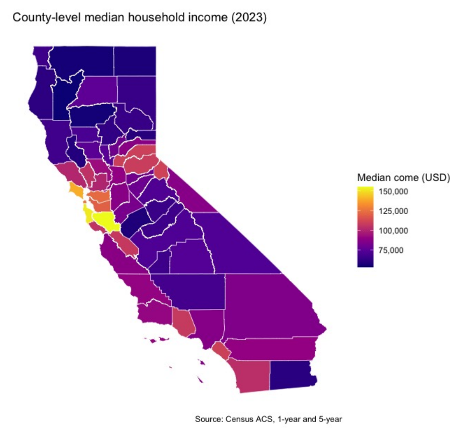

FIGURE 4. Median income (2023)

FIGURE 5. Revenue estimate (2023)

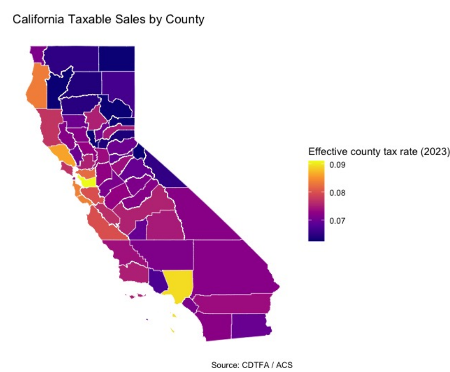

FIGURE 6. Local sales tax (2023)

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, 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 $county \cdot year$ 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 (ϕ) 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 (υ) were only available for the most recent year (2025), and backfilling was required. The median effective tax rate (Υ) 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 (ρ) 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 (Π_{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.

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 2025a). 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.

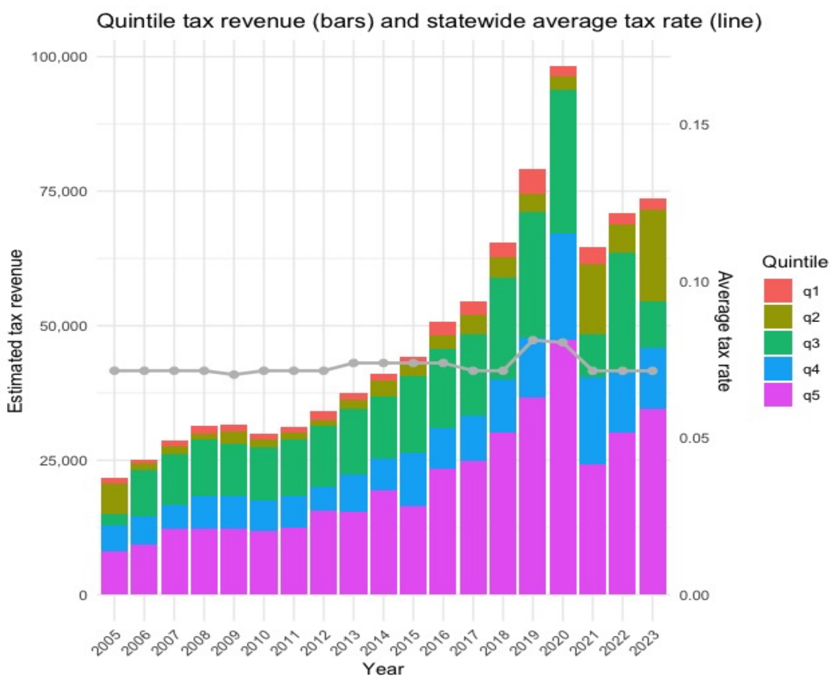

FIGURE 7. Quintile level estimates of taxable consumption by households (left axis), plotted against average effective consumption tax rate (right axis)

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. Overall, with the tax revenue approach, we find that in the pre-2011 period, tax revenue collected is actually *greater* than the estimated amount of tax revenue from consumption data provided in the household surveys. This indicates some measurement error in the analysis, which may require further probing into the survey methods themselves.

Next, we find that the range of the value of γ fluctuates significantly. In 8, we can see that more recent trends indicate that compliance is declining, such that the informal economy is expanding. Again, recent drops in the compliance correspond with 9, such that the gap between actual and projected revenue grew in the build up to 2020. Although, volatility in the measure of informal activity yields further questions on the temporal dimension of informality - it may be more short term, given the fluctuations. At this time, our measurement itself is uncertain, and the measurement itself does 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.

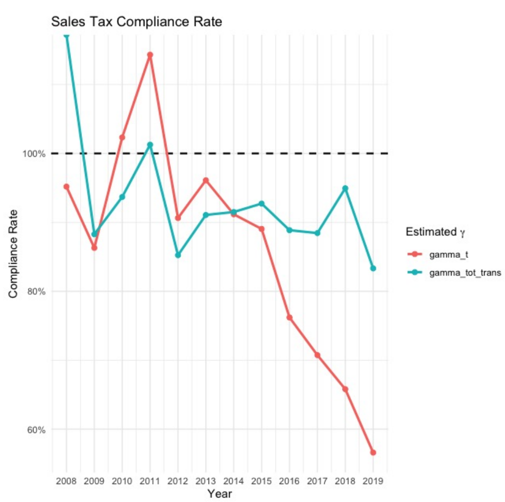

FIGURE 8. Share of realized revenue collected (γ)

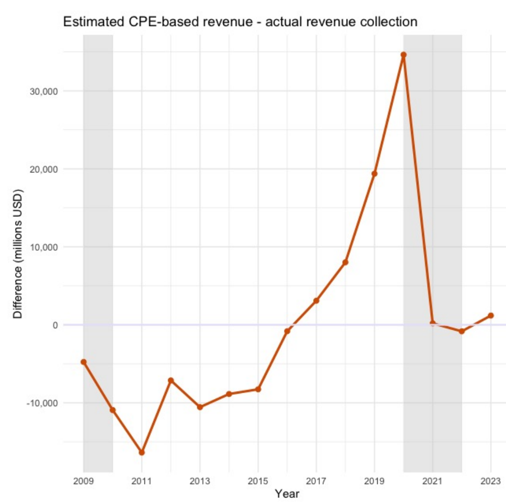

FIGURE 9. Difference in projected vs. annual revenue

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. In 10, a normalized presentation of the variables is given. This indicates there seem to be a trend between the unemployment rate and the compliance rate (γ). At the same time, PCE, GDP and measure of GVA in the economy to grow. This shows there may be a pattern between the level of compliance and unemployment, such that compliance decreases when unemployment decreases, but this robustness check then yields a hypothesis that counters the original claim of the paper.

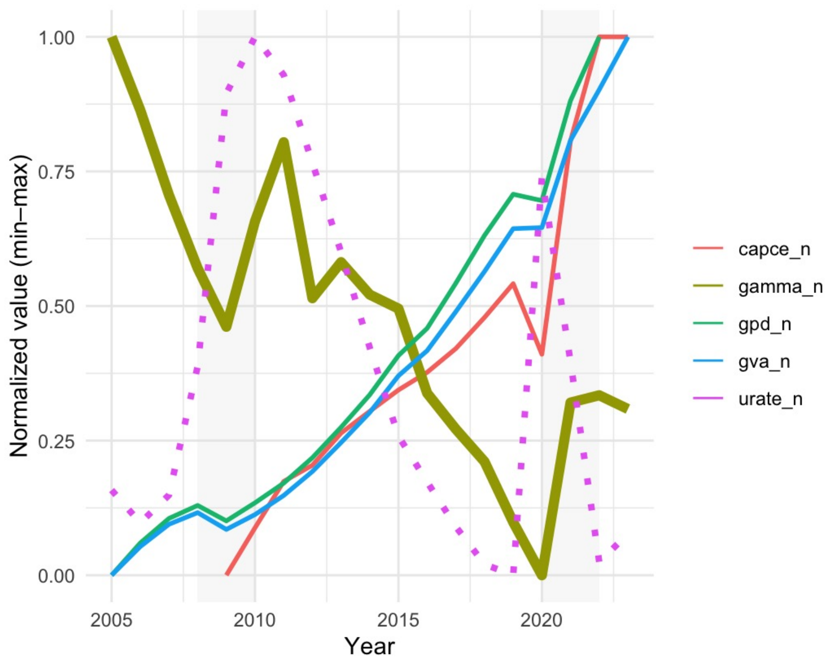

FIGURE 10. Normalized variables of interest over time

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

Abraham, Katharine G. and Susan N. Houseman. 2019. "Making Ends Meet: The Role of Informal Work in Supplementing Americans' Income." Upjohn Working Papers 19-315, W.E. Upjohn Institute for Employment Research. None.

Bayeh, Berhan, Isaiah Nardone, Shaun O'Brien, Hailey Phelps, Federal Reserve Banks of Atlanta, Federal Reserve Banks of Boston, and Federal Reserve Financial Services FedCash® Services. 2025. "Findings From the Diary of Consumer Payment Choice." Tech. rep. https://www.frbservices.org/binaries/content/assets/crsocms/news/research/2025-diary-of-consumer-payment-choice.pdf.

Bracha, Anat and Mary A. Burke. 2017. "Informal Work, Underemployment, and the Business Cycle." *Federal Reserve Bank of Boston Working Paper* .

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 Finance. 2024. "County Population Estimates and Components of Change."

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

California Department of Tax and Fee Administration. 2025b. "Taxable Sales, by Type of Business (Table 1)."

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

Internal Revenue Service. 2024. "Statistics of Income: California Adjusted Gross Income by ZIP Code (CAAGI)."

Leyva, Gerardo and Carlos Urrutia. 2020. "Informality, Labor Regulation, and the Business Cycle." *Journal of International Economics* 126: 103346.

Marshall, Emily C., James Saunoris, Mario Solis-Garcia, and Trang Do. 2023. "Measuring the size and dynamics of U.S. state-level shadow economies using a dynamic general equilibrium model with trends." *Journal of Macroeconomics* 75: 103491. https://doi.org/10.1016/j.jmacro.2022.103491.

Pappadà, Francesco and Kenneth Rogoff. 2023. "Tax Evasion and Capital Mobility." *Journal of Public Economics* 222: 104892.

Quiros-Romero, Gabriel, Thomas Alexander, and Jennifer Ribarsky. 2021. "Measuring the Informal Economy." Tech. rep.

Restrepo-Echavarría, Paulina. 2018. "Macroeconomic Volatility: The Role of the Informal Economy." *European Economic Review* 106: 454–469.

Schneider, Friedrich. 2023. "Do Different Estimation Methods Lead to Implausible Differences in the Size of Nonobserved or Shadow Economies? A Preliminary Answer." *Annual Review of Resource Economics* 15 (1): 257–277. 10.1146/annurev-resource-090822-114308.

Slemrod, Joel. 2007. "Cheating Ourselves: The Economics of Tax Evasion." *Journal of Economic Perspectives* 21 (1): 25–48.

U.S. Bureau of Economic Analysis. 2024. "Gross Domestic Product (GDP) by State and by Industry:

California (SAGDP2N)."
U.S. Bureau of Labor Statistics. 2024. "Consumer Expenditure Survey: Income Quintile Tables, California."
U.S. Census Bureau. 2023. "American Community Survey (ACS): Median Household Income (B19013_001), 1-Year and 5-Year Estimates."
US Sentencing Commission. 2025. "Tax Fraud." `https://www.ussc.gov/research/quick-facts/tax-fraud`.
Valletta, Robert G., Leila Bengali, and Carmen van der List. 2020. "Cyclical and Market Determinants of Involuntary Part-Time Employment." *Journal of Labor Economics* 38 (S1): S45–S92.
Zucman, Gabriel, Annette Alstadsæter, and Niels Johannesen. 2019. "Tax Evasion and Inequality." *American Economic Review* 109 (6): 2073–2103.