An approach to measuring informal activity in California

Allegra Saggese

ECON221 Economic slack
Early results presentation

November 13, 2025

How large is the informal economy in CA?

- Consumption based informal economy measures (Pappada & Rogoff 2023).

- Informality during the business cycle (Bracha & Burke 2017, Restrepo-Echavarria 2018, Leyva & Urrutia 2020, Valleta et al. 2020, Schneider 2023).

- *Future contribution:* Quantifying labor market participation to improve measurement of labor market slack (Sahin & Hojin 2021, Barnichon & Shapiro).

Methodology

Core estimate: a synthetic taxable base. I use two distinct data sources within same model, applying effective tax parameters.

Transaction based approach

- Use NAICS-level taxable transactions, selecting all industries within coverage
- Direct mapping from observed data to taxable base
- Revenue estimate:

$$\widehat{T}_t^{trans} = \sum_i \tau_{t,avg} \left(\widehat{c}_t^{trans} \cdot \widehat{i}_{t,avg} \right)$$

PCE based (Household) approach

- Use county-level disposable income and local tax rates, but requires quintile-level parameterization on taxed spending.
- Revenue estimate:

$$\widehat{T}_t^{pce} = \sum_{i,q} \tau_{it} \left(\widehat{c}_{itq}^{pce} \cdot \widehat{i}_{qt} \right)$$

Data and model

Measure		Description	Source
GDP (nominal)	GDP_t	Output in current prices	FRED FED (1999-2023)
Collected tax	T_t	Collected tax revenue	FRED FED (1999-2023)
Compliance rate	ϕ_t	$Rev_t / \sum_i \tau_{ti}(c_{it} \cdot i_t)$	*calc* (2003-2024)
County quintiles	τ_{it}	y_t grouped	BLS PCE - CA (2000-2023)
GVA	GVA_t	GSP	BEA NAICS (1999-)
Export share	x_t	CA export val	CA TFA, (2005-2024
Incidence	i_t	Share y_t taxable	BLS PCE - CA

Parameter estimation:

$$\gamma_{t,trans} = \frac{T_t^{collected}}{\widehat{T}_t^{trans}} \quad \sim \quad \gamma_{t,pce} = \frac{T_t^{collected}}{\widehat{T}_t^{pce}} \quad \rightarrow \quad \text{compliance ratio}$$

$$EVADE_t = (1-\delta_t)(1-x_t)\frac{GSP_t}{GDP_t} + \left(\frac{1}{\gamma_t} - 1\right)\frac{T_t^{collected}}{GDP_t} \quad \rightarrow \quad \text{missing trans + tax rev}$$
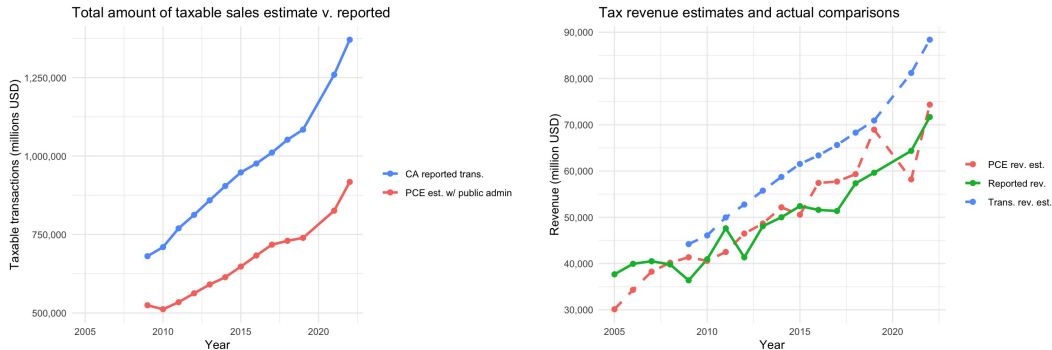
Figure 1: *Left plot shows estimated taxable receipts $\sum_i \tau_{ti}(c_{it} \cdot i_t)$ at county level, and industry level $\tau_t(c_t \cdot i_t)$. Right panel plots modeled revenue using both PCE and transaction-based incidence calculation, against actual tax receipts.*

formula 1

$$GDP_t^{full} = GSP_t + NTprod_t$$
$$+(1 - \delta_t)(1 - x_t)GVA_t$$
$$+(\phi_t - 1)T_t$$

formula 2

$$EVADE_t = \frac{GDP_t^{full} - GDP_t}{GDP_t}$$

$\Rightarrow$ Large swings in $EVADE$, volatility of GDP_t and $GDP_t^{full} - GDP_t$

EVADE measures vs. nominal GDP annual rate of change in CA

Legend:
- EVADE YoY (v1)
- EVADE YoY (v2)
- GDP YoY Growth

Next steps
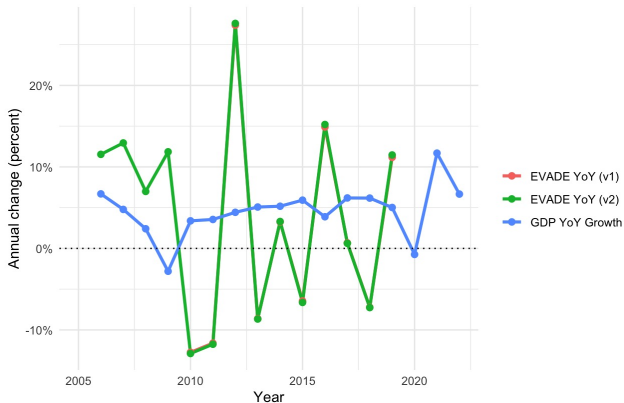
- Continue to **improve the EVADE measure**:
  - ▶ Distinguish measurement error from informality between *projected tax revenue figures* between the PCE and the transaction-based estimates
  - ▶ Disaggregate GDP, potentially imports to the county level for full alignment with PCE-approach
  - ▶ Include consumer-level subsidies on products for further refinement
- **Refine interpretation**: what does it mean for the EVADE measure to be volatile?
- **Provide more descriptives:** Plot against other raw data series, including collected income tax, and try to match model estimates with total tax incidence in CA