Measuring Income Tax Evasion Using Bank Credit: Evidence from Greece Get access Arrow by **?**

February 7, 2024

Some Odd Patterns in Bank Loans

Table I: Monthly Reported Income & Debt Servicing by Industry

The bank data are from a large Greek bank, with industry, income distribution, and geography weighted to be representative to the population of Greece using the tax authority distribution. Data are from 2003-2009. The sample consists of mortgage and non-homeowner consumer loan applicants. Monthly reported income is the verified income reported to the tax authority. Monthly debt payments is the servicing on the loans, with the interest rate equal to the applicable rate for new consumer loans and new mortgages. For term-loan, non-homeowner customers, we assume the prior debt has an outstanding term of 3 years, because much of this debt may be other term-loans. We assume a median rate of 11% for these obligations. For mortgage applicants, we impose a prior-debt maturity of 10 years and rate of 6% (conservative averages). The last column reports the mean of the payment-to-income ratio for applicants in each industry.

Industry	Mean Monthly Reported Income (€)	Mean Monthly Debt Payments (€)	Mean Ratio of Payment/Income
Accounting & Fin. Services	1,762	1,072	1.30
Business Services & Trade	1,359	916	0.75
Construction & Transport	1,190	549	0.51
Education	1,330	1,098	1.09
Engineering & Science	1,506	1,117	0.91
Farming	1,000	424	0.42
Law	1,792	1,559	1.17
Lodging & Tourism	1,440	1,183	1.16
Manufacturing	1,475	1,150	0.97
Media & Entertainment	1,486	1,098	0.93
Medicine	1,660	1,282	1.39
Personal Services	1,053	654	0.95
Retail	1,642	1,362	1.25
Overall	1,339	884	0.78

Context

- Use loan decisions data from a big Greek Bank.
- After algorithm assigns the scores, bank managers make decisions after considering soft information such as potential real income.
- Data is universe of credit applications including those rejected.
- Tax data includes total income reported and total no of households at postal code and occupation level.
- Occupations: self-employed, wage workers,merchants and agriculture

Methods

- Banks decide level of credit as a linear combination of their assessed true income, risk level and other area X individual soft skills.

$$credit = \beta_1 Y^{True} + \Phi Risk + \Psi SOFT$$

- In data, you observe reported income Y^R. They use *credit* to infer Y^{True}.
- Assume that

$$Y_{ij}^{True} = \begin{cases} \lambda_j Y_{ij}^R, & \text{if self-employed} \\ Y_{ij}^R, & \text{if wage worker} \end{cases}$$

- Use this to break credit equation to

$$credit_{ij} = \beta_{1j} Y^R ij(1-SE_i) + \beta_{2j} Y^R SE_i + f.e^{CreditGrade} + SOFT_{ij}\Psi$$

λ is ratio of the two betas. What are identifying assumptions of this equation?

Results

Figure: Debt Capacity

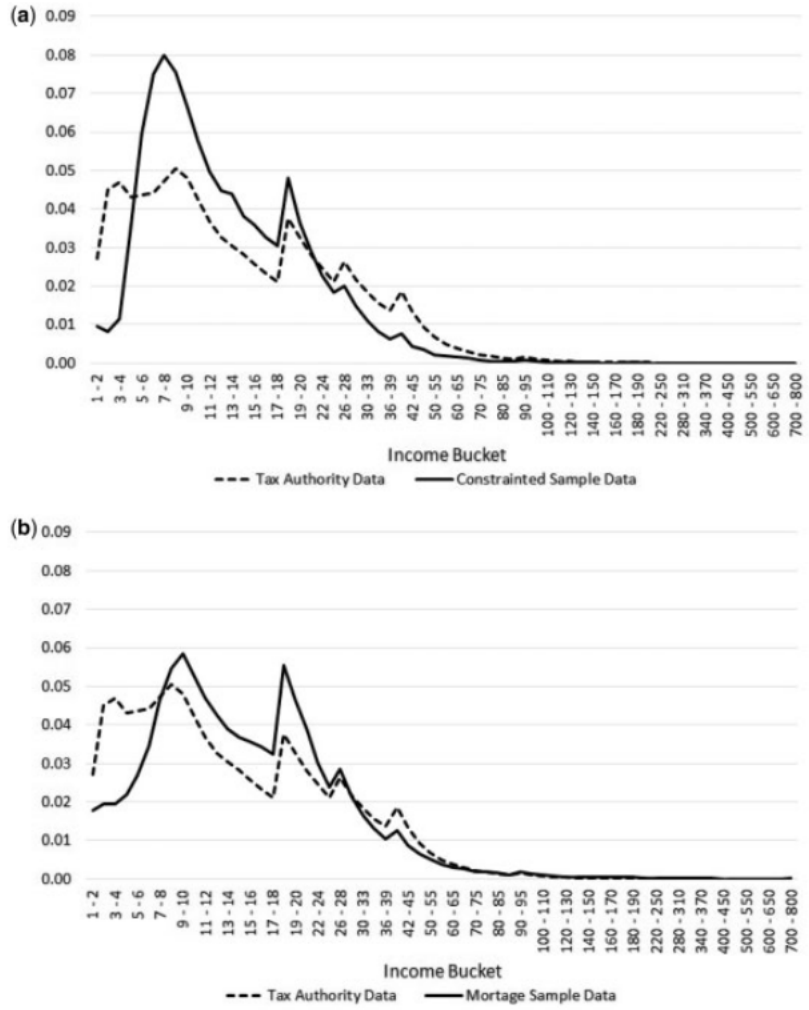

FIGURE I

National and Sample Reported Income Distributions

Results

Figure: Estimates

TABLE II

TAX-EVASION ESTIMATION OF SELF-EMPLOYED (CONSTRAINT SAMPLE)

	(1) OLS	(1a) λ	(2) OLS	(2a) λ	(3) OLS	(3a) λ	(4) Quantile	(4a) λ
Income*wage worker	0.3185***		0.3235***		0.3391***		0.3610***	
	[0.0467]		[0.0491]		[0.0544]		[0.0023]	
Income*SE	0.5575***	1.75***	0.5755***	1.78***	0.6257***	1.84***	0.6490***	1.79***
	[0.0602]		[0.0569]		[0.0514]		[0.0025]	
IncomeRisk					1,811		-663	
					[2,983]		[671]	
SE*IncomeRisk					1,564		878	
					[3,197]		[821]	
Lag(income growth)					1,729		-8,501***	
					[9,857]		[1,434]	
SE*lag(income growth)					8,336		9,814***	
					[10,517]		[1,810]	
Real estate wealth					0.9400***		0.1968***	
					[0.3193]		[0.0576]	
SE*Real estate wealth					-0.1045		0.0608	
					[0.5807]		[0.1255]	
Credit grade F.E.	Yes		Yes		Yes		Yes	
Industry*SE F.E.	Yes		Yes		Yes		Yes	
Branch F.E.	No		Yes		Yes		Yes	
Year F.E.	Yes		Yes		Yes		Yes	
Adj. R^2	0.100		0.118		0.120		0.140	
Tax evasion rate		42.85%		43.82%		45.65%		44.13%

Notes. Dependent variable: credit capacity = outstanding debt + approved loan. The table presents estimated coefficients (columns (1), (2), (3), and (4)) from equation (5). Columns (1a), (2a), (3a), and (4a) present lambdas (λ), defined as the ratio of the coefficient on income for self-employed divided by the coefficient for the wage worker. The sample consists of applicants, whose loan amount approved is lower than the amount requested (constrained sample). The dependent variable is credit capacity, defined as the sum of outstanding debt and the approved loan. *Income* refers to the after-tax reported income verified by the tax return. *Wage worker* and *SE* are dummy variables that indicate the employment type of the applicant. The vector of control variables includes the standard error of *log*(*income*) over the prior five years by tax cell (*IncomeRisk*), the per capita annual income growth by tax cell (*Lag*(*IncomeGrowth*)), and the median presumed real estate value of the postal code of the applicant's residence (*real estate wealth*). Models (1)–(3) are OLS, whereas model (4) presents results of a quantile regression. All specifications include 50 credit grade buckets as risk fixed effects, year fixed effects, and industry interacted with self-employment fixed effects. Models (2)–(4)include additionally branch fixed effects. The two last rows report the aggregate lambda and the tax evasion rate, defined as $\frac{Y^{True}-Y^{R}}{Y^{True}}$. Observations are weighted to be population-representative using the tax authority distribution. Heteroscedasticity-robust standard errors are presented in brackets. Significance at the 1%, 5%, and 10% levels is indicated by ***, **, and *, respectively. Significance of lambdas refers to statistical difference from unity.

Results

Figure: Mortgage Sample

TABLE IV

TAX EVASION ESTIMATION OF THE SELF-EMPLOYED (MORTGAGES) AND LARGE FIRM TEST

	(1)	(1a)	(2)	(2a)	(3)	(3a)	(4)	(4a)
	All Wage Workers				Wage Workers in Large Firms			
	OLS	λ	OLS	λ	OLS	λ	OLS	λ
Income∗Wage worker	0.0306∗∗∗		0.0286∗∗∗		0.0298∗∗∗		0.0281∗∗∗	
	[0.0106]		[0.0103]		[0.0091]		[0.0088]	
Income∗SE	0.0548∗∗∗	1.79∗∗	0.0540∗∗∗	1.89∗∗	0.0558∗∗∗	1.87∗∗	0.0549∗∗∗	1.95∗∗
	[0.0092]		[0.0092]		[0.0093]		[0.0093]	
IncomeRisk			46.95				132.6	
			[133.2]				[118.0]	
SE∗IncomeRisk			−16.21				−111.0	
			[144.6]				[131.2]	
Lag(income growth)			1,792∗∗∗				1,114∗∗∗	
			[432.1]				[282.6]	
SE∗lag(income growth)			−1,917∗∗∗				−1,229∗∗∗	
			[448.4]				[309.5]	
Real estate wealth			0.0951∗∗∗				0.0873∗∗∗	
			[0.0233]				[0.0166]	
SE∗Real estate wealth			0.0059				0.0177	
			[0.0330]				[0.0303]	
Credit grade F.E.	Yes		Yes		Yes		Yes	
Branch F.E.	Yes		Yes		Yes		Yes	
Year F.E.	Yes		Yes		Yes		Yes	
Adj. R^2	0.0903		0.0957		0.0931		0.0982	
Tax evasion rate		44.13%		47.09%		46.52%		48.71%

Notes. Dependent variable: monthly payment of approved mortgage. The table presents estimated coefficients (columns (1), (2), (3), and (4)) from equation (5). Columns (1a), (2a), (3a), and (4a) present lambdas (λ), defined as the ratio of the coefficient on income for the self-employed divided by the coefficient for the wage worker. The sample consists of mortgage applicants (mortgage sample). The dependent variable is monthly payments of the approved mortgage. *Income* refers to the after-tax reported income verified by the tax return. *Wage worker* and *SE* are dummy variables that indicate the employment type of the applicant. The vector of control variables includes the standard error of *Log*(*income*) over the prior five years by tax cell (*IncomeRisk*), the per capita annual income growth by tax cell (lag(income growth)), and the median presumed real estate value of the postal code of the applicant's residence (*Real estate wealth*). Models (3) and (4) repeat the analysis of models (1) and (2), using as benchmark only wage workers that work in large companies with more than €30 million in annual sales. All specifications include 50 credit grade buckets as risk fixed effects, branch fixed effects, and year fixed effects. The two last rows report the aggregate lambda and the tax evasion rate, defined as $\frac{Y^{True}-Y^{R}}{Y^{True}}$. Observations are weighted to be population-representative using the tax authority distribution. Heteroskedasticity-robust standard errors are presented in brackets. Significance at the 1%, 5%, and 10% levels is indicated by ∗∗∗, ∗∗, and ∗, respectively. Significance of lambdas refers to statistical difference from unity.

Results

- A λ of 1.75 translates into 43% evasion rate.

Applications

- Application in income taxation.
- Application in social insurance.

Behavioral Models

- Application to behavioral economics.

Broader Implications

- Potential applications in various economic fields.

Conclusion

- Main findings and contributions.
- Importance in policy analysis.

Bibliography