Measuring Income Tax Evasion using Bank Credit: Evidence from Greece[*]

NIKOLAOS ARTAVANIS, ADAIR MORSE AND MARGARITA TSOUTSOURA

September 25, 2015

Abstract

We document that in semiformal economies, banks lend to tax-evading individuals based on the bank's assessment of the individual's true income. This observation leads to a novel approach to estimate tax evasion. We use microdata on household credit from a Greek bank, and replicate the bank underwriting model to infer the banks estimate of individuals' true income. We estimate that 43%-45% of self-employed income goes unreported and thus untaxed. For 2009, this implies 28.2 billion euros of unreported income, implying foregone tax revenues of over 11 billion euros or 30% of the deficit. Our method innovation allows for estimating the industry distribution of tax evasion in settings where uncovering the incidence of hidden cash transactions is difficult using other methods. Primary tax-evading industries are professional services – medicine, law, engineering, education, and media. We conclude with evidence that contemplates the importance of institutions, paper trail and political willpower for the persistence of tax evasion.

I Introduction

Tax evasion is notoriously hard to uncover, particularly when economic exchange occurs in cash transactions (Slemrod & Weber [2012]). Going back at least as far as Cagan (1958), economists have developed innovative methods to infer the aggregate size of cash transactions in countries with large black markets based on input observables, such as electricity usage or money supply (Tanzi [1983], Lackó [1999], Schneider & Enste [2000], Alexeev & Pyle [2003]). It is challenging, however, to use such methods to learn about the incidence of tax evasion across industries, given that users of the observable inputs (i.e. electricity) may not correlate spatially with hidden income across households. Moreover, input-based methods may not be well positioned to estimate tax evasion for high-income professionals and service industries that are characteristic of middle or high-income countries. Other studies, use tax auditing or consumption inference to estimate tax evasion.[1] Although these methods work well in the very developed economies, when economic exchange is pervasively in cash, as in semiformal or informal economies, tax audits struggle to uncover hidden income (Slemrod [2007])[2], and survey data can be fraught with underreporting of expenses (Hurst, Li & Pugsley [2014]), as households hide or distort consumption.[3]

In this paper, we develop a new method to estimate tax evasion and its incidence across industries. Our method is based on the new observation that the formal financial sector adapts to income being semiformal rather than relegating semiformal finance to the black market. In particular, in an effort to stay profitable and competitive, banks adapt to semiformality by providing credit to individuals based on their inference of individuals' true income. We refer to this extending of credit off the soft information of unobservable income as *soft credit*. Our paper uses detailed individual-level application and performance data from household credit products – term loans, mortgages, and overdraft facilities – from a large Greek bank to estimate the magnitude of tax evasion in Greece and to speak the industry distribution across the population. Knowing the heterogeneity

---

[1]Pissarides & Weber (1989), Klepper & Nagin (1989), Clotfelter (1983), Christian (1994), Feinstein (1999), Lyssiotou, Pashardes & Stengos (2004), Feldman & Slemrod (2007), Gorodnichenko, Martinez-Vazquez & Peter (2009), Kleven et al. (2011), US Department of the Treasury, Internal Revenue Service (1988, 1996).

[2]Semiformal economies are those in which registered businesses (as opposed to black market businesses) transact in cash, thus evading taxes. In appendix Table 1, we compile a list of countries with large semiformal sectors, which cover most of southern and eastern Europe, Latin America, China, India, and many countries in southeast Asia and the Middle East.

[3]Braguinsky, Mityakov & Liscovich (2014), for example, offers an administrative data solution based on car registrations in Russia, an interesting consumption-based approach.

of hidden income across industries is essential for policy prescriptions. In the process, our paper also sheds more light on the phenomenon of semiformality, using tax evasion measurement to help us better understand the microeconomics of semiformal adaptations.

Our method adds to the literature in a few key respects. Using bank loan information and the adaptation of the financial sector, our approach can overcome some of the difficulties of audit and consumption methods in uncovering tax evasion in semiformal economies where hidden cash transactions permeate. Furthermore, our approach improves upon input observable and macroeconomy relationship methods by better capturing the huge extent of tax evasion in high level services.

We motivate our study with Table I, which illustrates soft credit at work. The data, described more fully later, are from a large Greek bank, nationally representative and covering tens of thousands of applications by individuals for credit products. Column 1 reports average monthly reported income for the self-employed by industry. Column 2 reports the monthly payments that individuals must be making to service debt on existing and approved credit. Column 3 presents the payments-to-income ratio. On average, the data suggest, implausibly, that self-employed Greeks spend 78% of their monthly reported income servicing debt.[4]

The point of Table I is to establish that soft-credit adaptation is happening and to motivate how we use bank data to examine tax evasion. We confirmed with other banks in southern Europe that it is common practice to use adaptation formulas varying by industry to adjust clients' reported income to the bank's best estimate of true income. Take the examples of manufacturing, engineering, medicine, education, media, accounting, law, and proprietors of restaurants, hotels, and retail in Table I. In all of these industries, the self-employed individuals appear to be paying nearly-or-over 100% of their reported income flows for debt servicing. Clearly, this percentage is infeasible.

Our approach recreates the bank's underwriting decision model that maps true income to credit, while controlling for credit risk and soft information for individual wealth, local conditions, and income risk. We observe credit and ask what level of true income a self-employed individual must make to support the credit extended by financial institutions. This method of inverting the mapping function from an outcome to income is parallel to Pissarides & Weber (1989) (see also Feldman &

---

[4]To put this number in perspective, a typical budget share for spending on just food and housing ranges from 39% to 62% across countries according to World Bank 2005 International Comparison Program (ICP) data.

Slemrod [2007]), who study the relationship between food consumption and income to uncover tax evasion.

Our estimating procedure makes a few key assumptions. First, similar to Pissarides & Weber (1989), we assume wage workers do not evade taxes. We show our results are robust to this assumption in alternative tests using only wage workers from large private companies, who do not underreport income as a result of third-party reporting and lack of employee-employer collusion (Kleven et. al [2011], Braguinsky, Mityakov & Liscovich [2014]). Second, our procedure assumes the bank determines the level of credit. We therefore focus our analysis on constrained borrowers, for whom the bank's supply decision will determine credit capacity. Third, we assume that after absorbing credit risk, the credit sensitivity to true income is the same for wage workers and self-employed. We use information on default rates to show that our results are not driven by unobserved heterogeneity in credit decisions between self-employed and wage workers that might be correlated with income.

We estimate that the true income of self-employed individuals is 1.75 to 1.84 times their reported income.[5] Across years 2006-2009, we document annual tax evaded income ranging from 22.8 to 28.2 billion euros for self-employed individuals, or 9.1 to 11.2 billion euros in foregone tax revenues. In 2009 the self-employed reported income was 35.7 billion euros, which according to our lambda multiple estimates corresponds to tax evaded income between 26.8 and 29.9 billion euros, with a median of 28.2. These foregone revenues would account for at least 30% of the budget deficit for 2009, and much higher proportion of the deficit in the earlier years, before the deficit reached its peak.

Our main incidence result is that we consistently find a high tax evasion multiple for doctors, engineers, educators, media, and lawyers. These professions are highly educated, usually with strong industry guilds and powerful government representation. Furthermore, we find that tax evasion is pervasive both in urban and rural areas and across the wealth distribution.

We use the estimated distribution to better understand which factors might allow tax evasion to persist. Much in the spirit of Kleven et al. (2011) and Morse, Karlinsky & Bankman (2009),

---

[5]To put some perspective on the magnitudes, Pissarides & Weber (1989) find that on average the true income of self-employed individuals in Great Britain is 1.55 times their reported income. Feldman & Slemrod (2007) use the relationship between reported charitable contributions and reported income, and find that in US tax evasion among self-employed, non-farm small-business, and farm income are 1.54, 4.54, and 3.87 times the reported income, respectively.

tax evasion is high in industries that have low paper trail, mainly professional services with cash income. Using tax-enforcement data from Greece, we present evidence consistent with a story that enforcement officials are aware of the incidence of tax evasion, but lack the ability or the willpower to collect evaded revenues, presumably because of a host of institutional frictions including state capacity weaknesses and the lack of paper trail to prove violations. Additionally, the distribution suggests a political economy story. Even excluding lawyers, sixty percent of parliamentarians are in the top four tax-evading industries, and over eighty percent come from the top six (out of thirteen) tax-evading industries. The alignment of the occupational backgrounds of Greek parliamentarians to our top tax-evading industries is only an association, but may suggest one possible reason behind the lack of willpower to enact tax reform

Our paper connects several strands of literature. First, the paper relates to the formal economy tax-evasion literature. We argue that the consumption (Pissarides & Weber [1989], Lyssiotou, Pashardes & Stengos [2004], Feldman & Slemrod [2007]) and audit (e.g., Klepper & Nagin [1989], Christian [1994], Feinstein [1999], Kleven et al. [2011]) methods may not be always well suited to estimate tax evasion in cash economy settings. In this literature, our paper is closest to Braguinsky, Mityakov & Liscovich (2014) and Gorodnichenko, Martinez-Vazquez & Peter (2009), who offer estimates of tax evasion in Russia, a semiformal economy, using consumption methods.

Second, we build off the vast literature on informal economies. Important work has been done not just on understanding the size of informal economies (Cagan [1958], Tanzi [1983], Schneider & Enste [2000], Alexeev & Pyle [2003], Gorodnichenko, Martinez-Vazquez & Peter [2009], Braguinsky, Mityakov & Liscovich [2014]), but also on understanding informality's implications (La Porta & Shleifer [2008], Hasker & Okten [2008], Çule & Fulton [2009], Paula & Scheinkman [2010] , Goel & Saunoris [2014], Callen & James [2015]). We offer a different approach to understanding informality that enlightens the incidence of tax evasion across industries, especially high level ones.

Recent advances in the literature tax evasion in both formal and informal economies have focused on compliance, understanding both the behavior of households in complying and the actions enforcers can do to induce more compliance (Blumenthal et al. [2001], Kleven & Waseem [2011], Del Caprio [2013], Hallsworth et al. [2014], Fellner, Sausgruber & Traxler [2013], Dwenger et al. [2014], Pomeranz [2015]). Certainly understanding how evasion responds to tax policy, audits, third-party information, and social incentives in a setting of semiformality and how persistence of

tax evasion in developed semiformal economies relates to compliance incentives would be welcome complements to our different goal of setting out a method for estimating the extent to semiformality and documenting semiformality's incidence.

Our paper also relates to the literature on the use of soft information in lending decisions (Petersen & Rajan [2002], Petersen [2004], Agarwal & Hauwald [2010]). Our results show that in countries with pervasive hidden income, tax evasion enters as soft information in bank's credit decisions.

Finally, the study adds to the discussion regarding the roots of the recent financial crisis in Europe, and particularly in Greece. The literature suggests a number of factors leading to the recent European crisis, including the deterioration of competitiveness (Shambaugh [2012]), the fragility of the financial sector and inefficient policies to address the problem (Brunnermeier et al. [2011], Acharya, Drechsler & Schnabl [2014]), and contagion effects (Chudik & Fratzscher [2011], Ang & Longstaff [2013]). Lane (2012) emphasizes on the failure of states to tighten their fiscal policies during the preceding period. In this context, tax evasion obviously plays an important role, as forgone tax revenues weaken the fiscal position of a country. It is no coincidence that Adjustment Programs for Greece, Portugal and Cyprus included strict provisions for limiting tax evasion, as a mean to improve their fiscal position.

The remainder of the paper is as follows. Section II describes the lending decision process, our data from the bank and the tax authority, and our main samples. Section III lays out our methodology. Section IV reports our empirical results. Section V uses the estimated industry distribution to examine some of the factors and norms that allow tax evasion to persist in semiformal economies. Section VI compares our soft-credit method to other approaches, and section VII concludes the study.

II   Data and Bank Underwriting Process

II.A   Lending & Underwriting Process

We begin by providing an overview of the bank lending process. When a prospective borrower applies for a loan, a branch officer starts the process by collecting information on the borrower's demographics, occupation, employment status, reported income, and outstanding debt. Reported

income is the after-tax income from the applicant's tax form, which is verified according to the mandatory procedures of the bank. The branch officer, after checking whether the borrower is on the national "black list" of bad accounts, transcribes the collected information into electronic form. The application is then sent to a central location, where loan decisions are made. Based on the information collected, the central system uses a scoring algorithm to compute a credit grade for the application. The credit grade determines the loan's acceptability.

If the borrower qualifies, debt-burden rules and interest-rate schedules guide the specifics of the loan offering, in particular, credit capacity and interest rates. These credit decisions are the essence of our study and also reside with bank-lending managers at a central location. At this point, the adaption to hidden income happens. The central lending managers adjust up reported income to the bank's best estimate of true income, according to internal guidelines. Banks operating in semiformal environments observe years of data of repayment and performance behavior of the borrowers, which they use to calibrate estimates of true income. With the estimate of true income in hand, the central lending managers decide how much debt the applicant's income can support.

Note the incentives of loan officers in this process. Compensation of loan officers is almost entirely based on seniority and promotion. Promotions are thus the incentive system for Greek bankers. The bank promotes bankers based on the volume of performing loans. Thus, Greek bankers have the incentive to use soft-credit information to issue more loans, but only as long as doing so is profitable.

II.B   Bank Data

Our main data are proprietary files from one of the eight large Greek banks, which together account for 80% of the domestic banking market share. The bank has tens of thousands of customers, with branches across the country. We are unable throughout to reveal exact observation counts and other details that might reveal the bank's identity. On average, one of these eight banks has 33.4 billion euros in loans outstanding and 373 branches across Greece. Our dataset is the universe of applications for consumer credit products and mortgages, both approved and rejected over the period 2003-2009. Consumer credit products include term loans, credit lines, credit cards, overdraft

facilities, and home loans.[6]

The dataset includes all information recorded by the bank during the application procedure, including loan characteristics, applicant demographics, and applicant credit-worthiness data. The loan characteristics data include the date of the application, branch office code, purpose of the loan, requested and approved loan amounts and durations, debt outstanding at this bank, total debt outstanding elsewhere, and interest rate charged. Our primary dependent variable will be credit capacity, defined as the approved sum of the current application and all prior debt outstanding from all sources. Individual characteristics data include reported after-tax income (as reported in the applicant's tax return and verified by the bank), the primary applicant's industry, employment type (wage worker or self-employed), age, co-applicant's or spouse's income, marital status, and number of children. The additional credit-worthiness variables include years in current job, years at current address, deposits with the bank, homeownership, and the length of the relationship with the bank. In addition to application data, we have monthly performance files from the bank. This dataset includes monthly loan-payment flows, days late on payments, interest rates, and loan amortization on all the loan products. Appendix Table A.2 defines our variables in greater detail and describes the information collection for mortgage and consumer credit product applications.

We apply a couple of sample restrictions and filters to the data. First, we drop applications after October 2009, the beginning of the crisis period in Greece. Our objective is to focus on the working population of Greece; therefore, we remove applicants under the age of 19 and those over 75 as well as applications by students, pensioners, homemakers, and the unemployed. We then drop applicants whose job code is bank clerk, to avoid picking up any special treatment resulting from bank employment. Beyond these sample restrictions, we trim the sample in three ways. First, we drop applications whose debt balance from other accounts is missing. Second, we filter out individuals with income under 1,000 euros because it is likely that they are receiving the loan based on an income base that is not their wages (e.g., pension, guarantor, etc.). Third, we remove applicants with income over 1.5 million euros or debt over 1 million euros to remove extreme outliers. Winsorizing these large income or large debt observations does not change any result. All three filters only remove 0.9% of the sample.

---

[6]Appendix Tables A.4 and A.5 present summary statistics for the consumer products used in the analysis. The missing consumer product is vehicle loans. We have these files, but household identifiers are not linked to the other information in the bank.

The bank uses the information collected in the application (loan characteristics, applicant characteristics, and credit-worthiness variables) to construct a credit grade for each application. As we previously mentioned, the credit grade determines whether the central system will accept or reject an application. This grade is the main application risk-profile variable that we use throughout the analysis.

The credit grade is a scoring whereby the bank assigns points non-parametrically for certain characteristics of the applicant or loan. We do not have access to the bank's proprietary scoring algorithm; however, we do have the credit grade for each application and the underlying variables, except histories of any prior accounts. We re-engineer this risk profiling in appendix Table A.3. The dependent variable is the credit grade for each application in our dataset, including those rejected. We are not permitted to show coefficients, however, in that it might reveal the bank's scoring algorithm, which lenders hold as intellectual property. We instead show the p-values for each coefficient to demonstrate that our variables are significant and partial r-squares to demonstrate which variables are incrementally important. Column 1 omits occupation industry variables. The model explains 56.5% of the overall variation. The most important variables are the relationship variables - years in job, age, homeownership, years of relationship with the bank - and a few industries. Inclusion of industries, column 2, increases the r-square to 66.2%. A second objective of including appendix Table A.3 is to investigate the change in variation explained when we add in the interactions of industry with self-employment. The r-square does not improve, implying that the bank's credit grade algorithm is unlikely to include these variables. The largest partial r-square in self-employed-industry variables is 0.0017, and most of the partial r-squares are 0.0001 or less. The analysis renders confidence in our understanding from conversations with the bank that adaptation to self-employed hidden income is orthogonal to the credit grading, which is important for our empirical design.

II.C Tax Authority Data

The revenue arm of the tax authority provided us with total annual reported income and total number of households represented within a zip-code. These data are further broken down into national income deciles and four occupation classes. These occupation categories are merchants, self-employed individuals, wage workers, and agriculture. The first two categories comprise by

definition the entirety of the self-employed in Greece, according to the tax authority's classification. The income deciles are national income deciles that are static during our sample. Greece has more than 1,450 zip-codes, which means we have more than 58,000 cells, for which we know the total cell income and number of households. The total set of tax-reporting households is just over 5 million; thus, our level of aggregation is 86 households.

At this level of aggregation, we first construct a local economic conditions variable and proxy for income growth. We measure income growth variable as the lagged annual per-capita income growth over the prior year in the cell. Second, we construct a measure of income risk in the cell. To do so, we have to take into account the difference in the number of people in the cell year to year. Thus, we proxy for income risk with the standard error (equal to the standard deviation divided by the square root of the observation count) of log income over the prior 5 years in the cell.

We further use the tax authority income data to construct weights for making our estimates representative of the population. For the purpose of constructing weights, the level of disaggregation of the tax authority data is too fine when we apply it to our bank samples. We apply a weighting scheme that uses quintiles of national income buckets (rather than deciles), the four occupation categories, and 14 meta-prefecture geographies, which amounts to 280 cells.

Finally, the tax authority also provided us with presumed real estate values by building block. We use these values as a proxy for the wealth of individuals in the zip-code-occupation. We calculate the proxy using the median of the real estate values at the zip-code level.

III   Methodology

Our empirical approach uses the bank's credit capacity decisions to estimate tax evasion. In the bank lending process, once an application has been deemed acceptable based on its credit grade, the bank makes credit capacity decisions based on true income (Y^{True}), credit grade ($Risk$), and any local soft information variables ($SOFT$):

$$\text{credit decision} = f(Y^{True}, Risk, SOFT).$$

In practice, the natural functional form of a lender's credit capacity decision is usually linear. The bank applies debt-to-income ratio rules concerning the amount of income needed to support debt which implies a linear relationship of credit scaling up with income. Banks then adjust the capacity according to risk-shifting guidelines and soft information, which enter credit capacity non-parametrically:

$$(1) \qquad credit_{ij} = \beta_{1j} Y_{ij}^{True} + Risk_i \Phi + SOFT_{ij} \Psi,$$

where $credit_{ij}$, is the credit capacity decision for loan applicant i in industry j.[7]

Equation (1) includes three determinants of credit capacity. The $Risk$ variable comprises verifiable ("hard") information the bank collects at the time of the application and is reflected in the application credit grade. $SOFT$ is any local soft information that enters the bank's credit capacity decisions.[8]

The most important determinant of credit capacity is true income, Y^{True}, which is not observable. The bank instead observes reported income, Y^R. Like Pissarides & Weber (1989) (PW), our methodology infers true income from a well-measured dependent variable – in their case, food consumption, and in our case, extended credit. The PW seminal innovation to invert an outcome to estimate true income relies on a few key assumptions, which are also natural for us. In particular, PW assume wage workers do not evade taxes, and true current income is on average a multiple of reported income for the self-employed. These assumptions are given by

$$(2) \qquad Y_{ij}^{True} = \begin{cases} \lambda_j Y_{ij}^R, & \text{if } i \text{ is self-employed} \\ Y_{ij}^R, & \text{if } i \text{ is wage worker} \end{cases}$$

We know the bank (and other banks) uses multipliers applied to reported income, as in equation

---

[7] The linearity of our credit model differs from Pissarides & Weber (1989) (PW), who write a similar form for income supporting consumption in a log form. We deviate from PW in this aspect because bank underwriting does not scale naturally to log form.

[8] By soft information, we have in mind Petersen (2004)'s definition: "[Soft information] includes opinions, ideas, rumors, economic projections, statement of [a household's] future plans, and [local] commentary." In our specification, we proxy for household wealth, local economic conditions, and income risk. These may be codifiable hard information variables, but these variables proxy for soft extractions of information. For example, a loan officer might learn whether an area is doing well economically or whether a household is wealthy is based on bank conversations with households or bank experience with the area.

(2) to adjust up self-employed income to their best estimate of true income. Banks calibrate these multipliers based on their long-term experience with different industries. Thus, this assumption of PW is the same as bank practice.

Denoting a self-employed individual i by the indicator SE_i and combining equations (1) and (2), the credit-capacity decision is

$$(3) \qquad credit_{ij} = \beta_{1j}Y_{ij}^R(1-SE_i) + (\beta_{1j}\lambda_j)Y_{ij}^R SE_i + Risk_i\Phi + SOFT_{ij}\Psi,$$

or in estimating form:

$$(4) \qquad credit_{ij} = \beta_{1j}Y_{ij}^R(1-SE_i) + \beta_{2j}Y_{ij}^R SE_i + f.e.^{CreditGrade} + SOFT_{ij}\Psi + \varepsilon_{ij}.$$

Empirically, our objective is to absorb application *Risk* and *SOFT* information so that we can focus on the income-to-credit relationship. Thus, we include fixed effects for 50 application credit-grade ($f.e.^{CreditGrade}$) levels. We include *SOFT* information local variables of wealth (real estate values), local conditions (lagged income growth), and income risk (the standard error of log income over the last five years). We write equations (3) and (4) in levels (euros) rather than logs, because debt rules operate as linear scaling, not as percentage changes. Because we are not working in logs, the multiple of the tax-evasion term interacts with income.[9] Our estimate of λ_j is the coefficient on self-employed individuals' income divided by the coefficient on wage workers' income, $\lambda_j = \frac{\beta_{2j}}{\beta_{1j}}$.[10]

III.A   Discussing Estimation Identifying Assumptions

In addition to making the PW assumption that true income is on average a multiple of reported income for the self-employed, we rely on the following assumptions in order to be able to interpret the coefficients as identifying λ_j, the multiple of tax evasion:

---

[9]In PW, the terms separate because of the log form.

[10]Note the parallel between our specification and that of Pissarides & Weber (1989). PW write an estimating equation as $\log food_{ij} = \alpha_{0j} + \alpha_{1j}\log Y_{ij}^R + \alpha_{2j}SE_{ij} + \varepsilon_{ij}^{pw}$, where SE is a self-employment dummy. The multiple of tax evasion λ_j^{pw} can be backed out of the estimates as $\lambda_j^{pw} = \exp\left(\frac{\alpha_{2j}}{\alpha_{1j}}\right).$

1. Wage workers do not evade taxes.

2. After absorbing credit risk, the credit sensitivity to true income is the same for self-employed as for wage workers.

3. Credit sensitivity to reported and hidden parts of true income are the same.

4. Any unobserved heterogeneity affecting credit decisions is orthogonal to reported taxable income.

5. We can measure credit capacity and not just demand-driven debt outstanding.

III.A.1   Tax Evasion of Wage Earners

We make the standard assumption that wage workers do not evade taxes, and thus that their reported income is accurate. Greece has third-party income reporting for wage workers' income; the employer reports the worker wages to the tax authority directly. Kleven et al. (2011) show that for income subject to third-party reporting, the tax-evasion rate in Denmark is close to zero. However, in countries with pervasive semiformality, like Greece, even wage workers likely do evade taxes in a variety of ways. Firms and employees may collude with respect to third-party reported income, especially in small companies (Kleven, Kreiner & Saez [2009]). Furthermore, employees may evade taxes by taking bribes or holding side-jobs. To the extent that wage workers do evade taxes, our estimates of β_{1j} will be overstated. A smaller income will appear to support more credit. Thus, our estimate of tax evasion will be conservative.

A potential concern, however, is that wage workers might evade taxes differentially by industry, implying conservatism might vary by industry. To address this issue, we provide additional tests in which we exclude all wage workers except those who work in large companies with more than 30 million euros in sales. In these large companies, the probability of employee-employer collusion with respect to third-party reported income is small, and these employees are less likely to have self-reported income on the side. We show that the results we obtain using only employees of large firms as the benchmark are similar to the results we get when we use all wage workers as the benchmark.[11]

---

[11]A lingering possibility is that the selection of individuals in larger-versus-small companies varies across industries in a way that correlates with credit risk. However, estimating the wage worker sensitivity of credit to income, β_{1j} by

III.A.2 Use of Credit, Selection to Self-Employment, and Hidden Risk

We assume the use of credit to support one's lifestyle is similar for the self-employed and wage workers. A parallel comes from Feldman & Slemrod (2007), who use charitable contributions to infer true income. They assume that the selection of who is self-employed is unrelated to one's charitable inclinations. In our case, if debt were to serve a different role for the self-employed (supporting income growth) than for wage workers (facilitating consumption out of permanent income), then a systematic difference between the self-employed and wage workers could be reflective of risk differentials.

Two other stories reflect a similar concern for hidden risk differentials between the self-employed and wage workers. First, being constrained may imply something different for a self-employed individual than for a wage worker, which in turn implies a selection on hidden risk. Second, the behavioral or cognitive reasons to select into self-employment may imply a different risk for the bank.

We take a couple of steps to address hidden risk. We make the estimating equation econometrically stringent by saturating the model with fixed effects for time, bank branch, and industry crossed with self-employment. If we lack pertinent soft information characteristic of localities or if a particular branch caters to (or appeals to) different types of customers, we now absorb these heterogeneities with bank branch fixed effects. The year fixed effects capture the impact of the business cycle on underwriting depth. In addition, we allow the loadings on soft information variables (the variables themselves vary by geography interacted with four occupation classes and income) to vary by self-employment:

$$credit_{ij} = \beta_{1j}Y_{ij}^{R}(1 - SE_i) + \beta_{2j}Y_{ij}^{R}SE_i + f.e.^{CreditGrade} + \beta_3 Wealth_{ij} + \beta_4 Wealth_{ij}SE_i$$

$$(5) \qquad + \beta_5 IncomeGrowth_{ij} + \beta_6 IncomeGrowth_{ij}SE_i + \beta_7 IncomeRisk_{ij}$$

$$+ \beta_8 IncomeRisk_{ij}SE_i + f.e.^{Branch} + f.e.^{Industry*SE} + f.e.^{Year} + \varepsilon_{ij}.$$

In making the model more stringent with fixed effects, however, we no longer are able to identify the full industry distribution in a model with the β_{1j} differentiating by industry. The collinearity

---

industry j allows us to isolate these differences. In the next section, we take up industry-specific selection of who is self-employed.

makes the coefficients implausibly large and small, and the variance inflation factors are huge. Thus, when we estimate the distribution of tax evasion at our most granular level of industry, we keep the $f.e.^{Industry*SE}$ and assume $\beta_{1j} = \beta_1$. This assumption should not affect the overall level of tax evasion. However, a potential concern is whether it affects the ranking. We show robustness to this assumption using less granular industry definition, where we can include $f.e.^{Industry*SE}$ and β_{1j}.

A related concern is the use of business accounts to absorb some of the personal consumption. What if, in certain industries, proprietors expense certain items as business use? In particular, we can consider cars. If a self-employed individual pays for her car through the business and uses the expense to lower taxes, she might have more cash flow available to service debt for a given level of income. Including industry fixed effects interacted with self-employment should address this, unless the absorption of personal consumption is correlated with income. Although coming up with a few items that proprietors can expense through the business (e.g., lunches, office supplies, etc.) is easy, coming up with substantial items that are tax expensible and correlated with income is difficult. Fixed-sum items, such as cars, will be absorbed in the industry-self employment fixed effects. We cannot totally rule out this omitted variable, but making the case that it is large enough to drive our results is, in our opinion, somewhat of a challenge.

III.A.3 Credit Sensitivity to the Reported and Hidden Parts of True Income

A related story might be that the hidden income portions of a self-employed individual's true income is more risky. One might argue that hidden income lacks observability and tangibility, implying collections are less effective. A couple of reasons suggest that hidden income is no less of a future payments source than reported income. First, given that our main analysis is on uncollateralized loans, all debt-servicing payments are already made without a forcing mechanism other than access to future credit. Second, in Greece during this period, laws did not allow garnishment of wages for credit obligations; thus, neither formal nor informal income would serve that purpose. Finally, and perhaps most importantly, Greece operates in the norm of semiformality. A banker would not assume that a proprietor had no income to make debt payments because the proprietor's income statement showed no operating profits.

Even if exceptions existed to these arguments that formal income is no better "collateral"

than informal, violations to this assumption should bias our estimates of self-employed tax evasion downwards.

III.A.4 Unobserved Heterogeneity Affecting Credit Decisions

Although we control for industry interacted by self-employed fixed effects, any unobserved heterogeneity affecting credit decisions that is correlated with reported income might bias our results or confound interpretation. One of the advantages of our framework is that we have information on interest rates and default rates, enabling us to test whether our model fully adjusts for risk that could potentially vary along the income distribution.

Default rates are the metric on which underwriting models are gauged and can inform us the extent to which our model, $credit\ decision = f(Y^{True}, Risk, SOFT)$, captures unobserved heterogeneity among borrowers. If unobserved heterogeneities were to remain, we should see ex-post differences in default rates that correlate with tax evasion. We discuss bank profitability in greater detail and test this point in section IV.E. We are able to show that there are not significant ex-post differences in default rates or ex-ante prices of loans that correlate with the extent of tax evaded income.

III.A.5 Issues of Demand and Application Data: Samples & Selection

Outstanding debt is an artifact of demand decisions as well as supply, but our model is one of a bank's decisions regarding the credit capacity of a borrower. We employ two samples which approximate situations in which debt outstanding is at its capacity.

Our main sample is a *constrained sample*, containing all applicants, whose requested amount is greater than the approved amount for the period 2003-2009. For such individuals, credit outstanding is likely at capacity.[12] However, this sample seemingly introduces two layers of selection that we are now imposing on the data. First, we are selecting on applicants. Second, we are selecting on applicants who are constrained. But these people are the selection that the bank uses over time

---

[12]Similar to US, truncation of a loan size request is largely due to income-to-payments mechanical constraints. Banks operate with debt-to-income rules; thus approved amount can be lower than requested if applicants do not have sufficient cash flow from income to cover payments. A lingering possibility is that approving a lower loan amount than requested reflects the bank's perception of hidden risk that comes from the signal of a loan size request. To the extent that this hidden risk does not correlate with income, this hidden risk should be absorbed by the fixed effects. If it does correlate with income, we should pick up differences in metrics of risk as we describe in the previous subsection III.A.4.

to calibrate the relative informativeness of reported income across professions in loan-performance statistics; it informs the bank's decision at the margin. Thus, the selection of constrained applicants is the selection we need to uncover the bank's estimate of true income.

The constrained sample may, however, not generalize to the population. Hence, we implement a second sampling: a *mortgage sample*. Individuals who take out a mortgage often choose to buy as much house as their economic situation supports. In the United States, whereas lower and middle income households check with bankers or online tools to uncover how much mortgage their incomes can support before buying, higher income U.S. households may decide on housing consumption based on a host of heterogeneous factors. In Greece, the home ownership decision is much more deliberately considered for all households, and the income distribution is much less skewed (see Figure I). Eighty percent of Greek households eventually end up owning homes. Families buy a house once in their lives, once they are established[13], and it is not common to move. Households will choose mortgage payments (our dependent variable in this sample) to be very near to the capacity that income flow can support, with a precautionary buffer.

A caveat is that the home-buying selection may be biased toward individuals who have positioned themselves favorably in economic accounts to facilitate buying a house. This should be consistent across individuals in the sample, but affects the generalizability of the results. However, home buyers are from all spectrums of the workforce in Greece, which makes the generalizability concern somewhat weaker. In addition, note that this potential bias is likely to be the opposite of the constrained sample, which consists of individuals who may be positioned unfavorably in their economic accounts. We cannot prove that either sample is without biases, but our goal is to provide enough evidence from different angles of bank underwriting, in this case, samples, to test the robustness of the results.

Two final advantages of the mortgage sample come from using monthly payments as an alternative dependent variable. By using payments, we first test robustness to our main dependent variable of credit capacity, and we second have a dependent variable reflecting risk adjustments from the pricing of a loan. The downside of this sample is that we only have data for 2006-2009, and the sample is only one-fifth of the observation count of the constrained sample in a given year.

---

[13]Young people usually wait a decade or two after school to establish their income trajectory and location before buying.

Thus, we are limited as to the detail of analysis we can perform with this sample.

Figure I presents the reported income distribution of our two samples in conjunction with the national income distribution, as provided by the tax authorities. Our samples closely follow the national distribution. As expected, the mortgage sample is slightly more representative at the upper income levels.

IV   Results

IV.A   Aggregate Results

In this section, we show the baseline results based on our main sample of constrained individuals, where we estimate a single coefficient β_2 for all self-employed individuals. The estimating equation is:

$$credit_{ij} = \beta_1 Y_i^R (1 - SE_i) + \beta_2 Y_i^R SE_i + f.e.^{CreditGrade} + \beta_3 Wealth_{ij} + \beta_4 Wealth_{ij} SE_i$$

(6)
$$+ \beta_5 IncomeGrowth_{ij} + \beta_6 IncomeGrowth_{ij} SE_i + \beta_7 IncomeRisk_{ij}$$

$$+ \beta_8 IncomeRisk_{ij} SE_i + f.e.^{Branch} + f.e.^{Industry*SE} + f.e.^{time} + \varepsilon_i.$$

The dependent variable is credit capacity, defined as the total debt outstanding plus the loan amount approved. The specification includes 50 credit-grade fixed effects, year fixed effects, and 13 industry dummies interacted with self-employment status. We report population-weighted estimates, using the tax authority distribution and robust standard errors.

Table II reports the results. Column 1 is the simplest version of the model. The first row of Table II reports the coefficient $(\widehat{\beta}_1)$ on reported income for wage workers. A coefficient of 0.3185 implies that for every additional euro of income, wage workers have 32 cents more credit capacity. The third row of Table II reports the coefficient $(\widehat{\beta}_2)$ on reported income for self-employed individuals. (Note that the level effect of self-employment is absorbed in the fixed effects.) A coefficient of 0.5575 implies that for every additional euro of income, the self-employed have 56 cents more credit capacity. Column 1(a) shows that the income multiplier λ (defined as $\widehat{\lambda} = \frac{\widehat{\beta}_2}{\widehat{\beta}_1}$) is 1.75, indicating that the total true income for self-employed individuals is 1.75 times the reported income, which

implies a tax-evasion rate of 43%. (The tax-evasion rate is defined as $\frac{Y^{True} - Y^R}{Y^{True}} = \frac{\lambda - 1}{\lambda}$.) Given that the average income of self-employed individuals is 19,478 euros, the average tax-evaded income is 14,608 euros per self-employed individual in Greece.

Column 2 adds branch fixed effects, which negligibly change the results. The fact that our estimates shift around so little when we include branch fixed effects provides at least some confidence that unobservables might be of lesser concern, as argued by Altonji et al. (2005). Column 3 adds proxies for income growth, variability of income, and wealth, both in levels and interacted with self-employment status. The goal of this specification is to mitigate concerns that differential income growth or wealth between self-employed and wage workers might drive the results. The coefficients on income growth and variability of income are not significant in the presence of credit-grade fixed effects. The coefficient on real estate wealth is positive and significant. The estimated multiple λ changes negligibly compared to column 2.

One possible story that would be consistent with the coefficient on income being higher for the self-employed than for wage workers is that the self-employed may have a higher preference for debt as their businesses grow. This would imply that the preference for debt would increase with the income distribution. If true, a mechanical selection might be the source of our λ. The fact that wealth has a positive, significant coefficient in column 3 is consistent with this, but the fact that the coefficient on self-employed income does not decline from column 2 to column 3 is not consistent with this theory.[14] Finally, in column 4 we consider whether our mean result is driven by outliers, in particular estimating a medians quintile regression.[15] The results are very much in line with the OLS results.

Related to the concern about a wealth bias, we can also consider robustness to the assumption that the multiple of tax evasion by industry is linear in reported income. Figure II plots estimated credit capacity as a function of reported income, comparing the linear specification with a spline specification. The spline regression estimates the sensitivity of credit capacity to reported income

---

[14]In Section IV.F.2 we go one step further to consider whether the adaptation multiplier varies by wealth, and we find that λ is stable across the wealth distribution. This addresses not just this point, but also whether there is a different credit-capacity-to-income relationship for those with more income or wealth. We should emphasize that income and wealth are highly persistent across generations in Greece and mobility is low, making these concepts very related.

[15]We have to estimate the quantiles regression without branch fixed effects to get convergence. We checked if the OLS results were sensitive to excluding branch fixed effects; they were not, and we decided the results were informative despite this simplification.

for 8 regions of income, separately for wage workers and the self-employed.[16] Also included in the figure is the histogram of reported income in our sample. The spline model is almost identical to the linear model for wage workers. For the self-employed, the spline model predicts a higher credit capacity than the linear model for those with upper income, starting at 25,000 euros, re-converging to the linear model for extreme high income individuals. The wage worker spline model also exhibits a similar pattern for upper-reported income individuals, but to a smaller extent. This effect concerns those in the 83rd to 98th percentiles of reported income, and implies that our estimate of tax evasion may be somewhat conservative for this range of high reported income individuals.

Finally we do additional robustness tests of our results. In 6% of our loans there is a guarantor. Our results do not change when we drop these loans. Also our results are similar when we only keep uncollateralized loans to account for any differential ability to post collateral (99.8% of consumer loans is uncollateralized).

IV.B   Aggregate Economy Economic Significance

Overall, Table II shows income multipliers (λ) that range from 1.75 to 1.84, which correspond to a tight range of tax-evasion rates from 43% percent to 45% for the self-employed. Turning to the aggregate inference for the economy, Table III shows the self-employed reported income from 2006 to 2009 and the implied tax evaded income based on our low, high and median estimates of λ from Table II. Across all years, we estimate that tax evasion ranged from 22.8 to 28.2 billion euros.

If we look at specific years, we can calibrate this magnitude relative to the deficit. In 2009 the self-employed reported income was 35.7 billion euros, which according to our lambda multiple estimates corresponds to tax evaded income between 26.8 and 29.9 billion euros. This amount, under a 40% tax rate, implies foregone tax revenues between 10.7 and 12 billion euros, or 30-34% of the deficit of 2009 (44-49% of the primary deficit). Using 2009 de-emphasizes the importance of forgone tax revenues since during this year the deficit reached its peak. Foregone tax revenues account for well over 100% of the deficit in 2006 (113%), trending down as a percent because of the rising deficit in the years immediately prior to the crisis.

How does our 43%-45% estimated tax-evasion rate compare with previous studies? Pissarides

---

[16]The linear spline nodes are 1,000, 5,000, 10,000, 15,000, 20,000, and 30,000 euros.

& Weber (1989) find that, on average, the true income of self-employed individuals in Great Britain is 1.55 times their reported income, which corresponds to a tax-evasion rate of 35%. Kleven et al. (2011) show underreporting of 41.6% in Denmark for self-employed income, while research building off the Tax Compliance Measurement Program (TCMP [2001, 2006], Slemrod [2007], Black et al. [2012]) estimate an underreporting for the U.S. of approximately 50% for the self-employed. Galbiati & Zanella (2012) report an evasion rate of 46.6% for Italy, whereas Braguinsky, Mityakov & Liscovich (2014) find 80% tax-evasion rates in Russia.

IV.C   Robustness: Mortgage Sample & Wage-Worker Evading

In Table IV, we repeat the analysis in the sample of mortgage-approved applicants. The dependent variable is the approved monthly payment implied by the mortgage amount, duration, and interest rate. Column 1 presents the simplest version of the model, whereas in column 2, we add branch fixed effects as well as soft information proxies for income and wealth. The results are similar to the estimates in Table II.

Throughout the analysis, we assume wage workers do not evade taxes. One potential concern is the robustness of our results if wage workers also evade taxes, which would imply that our estimates would understate tax evasion for the self-employed. To address this issue, in columns 3 and 4 of Table IV, we repeat the analysis using as a benchmark only wage workers that work in large companies. Braguinsky, Mityakov & Liscovich (2014) show that large firms are more transparent than smaller firms. Wage workers in large companies are more likely to accurately report their earnings, because in addition to third-party reporting, for very large firms the probability of employee-employer collusion with respect to third-party reported income is very small, and these employees are unlikely to have self-reported income on the side.

To identify employees of large companies, the bank internally matched the names of employers from the mortgage-application file to balance-sheet information on the revenues of the firms. Based on the matching, they provided us with a dummy variable that identifies whether the wage worker works in a firm with more than 30 million euros in annual sales.[17] The estimated λ in column 3(a) is slightly higher, but not statistically significantly different from the estimated λ in column 1(a).

---

[17]The employer name was available only in the mortgage applications; thus, similar analysis can not be performed in the other samples.

Similarly, the estimated λ in column 4(a) is similar to that of column 2(a). Thus, our results in Table IV do not change when we use as a benchmark the group of employees who work in large companies and for whom the assumption that they do not evade taxes likely holds.

IV.D   New Checking Account Sample

One potential bias with the constrained sample is that if the self-employed are more likely than wage workers to be constrained by the bank, we are getting different sample profiles across the two groups. We add an additional robustness sample of applicants opening an overdraft line of credit who have no prior relationship with the bank. Although at first blush, this group sounds like debt-laden individuals, the opposite is true. These individuals are opening new checking accounts; overdraft protection is a standard product attached to new accounts. Table A4 (columns 7-9) reports the summary statistics for this sample. Reported income in the new customer sample (€21,780) is higher than that of the term-loan applicants (€13,209) and that of the mortgage applicants (€17,600).[18]

The limitation to using this sample is that a different form of selection may be at play. Individuals opening new bank accounts, especially the self-employed, may belong to an industry distribution selection of higher-income professions. The overdraft sample is much smaller and may not have the full distribution of lower-income occupations. Our point is not to claim this sample is without bias, but to look for robustness to the bank's decision on constrained borrowers being endogenous. The estimated multiplier (λ) is 2.96 for the baseline specification and 2.87 for the specification with income and wealth controls (Table A.6). These results provide additional reassurance that our main estimates based on the constrained sample are conservative.

IV.E   Defaults, Interest Rate, and Tax Evasion

The bank provided us with information on interest rates and defaults, allowing us to test whether our estimates of tax evasion are picking up an omitted risk characteristic. The thought experiment is to take the estimates of tax evasion (in euros) for each self-employed individual in the main, constrained sample. We then regress the loan pricing (interest rate assigned to the loan) or loan

---

[18]The difference in the caliber of these applicants is not reflected in the credit grade, because these applicants do not have a prior relationship with the bank (having a prior relationship with the bank adds points to the internal credit grading).

performance (whether or not the loan ever goes 90 days delinquent) on the estimated tax evasion. If our tax evasion estimates are positively or negatively related to the default or profitability of loans, then one might argue that our estimates are picking up an omitted risk heterogeneity related to self-employed income. By contrast, if the bank is appropriately adjusting for the creditworthiness of the self-employed by using income multipliers, then our estimated tax evasion should not imply any difference either in defaults (realized risk) or interest rate charged (ex ante risk).

In Table V we test this idea in a setup of estimating one of these outcomes on our estimated tax evasion from Column 3 of Table II. The dependent variable in columns 1-4 is the dummy variable *Defaults*, which takes the value 1 if the account is ever more than 90 days delinquent, and 0 otherwise. Thus we have collapsed to one observation per applicant. We also control for contract terms following Agarwal & Hauwald (2010). All specifications include fixed effects for the year of the application. [19]

We find that tax evasion is statistically unrelated to default. The tax evasion coefficient in column 1 is statistically and economically zero. Even if we compare within credit grade (column 2), borrowers who evade taxes more do not have different default rates compared to those with lower tax evasion. Although, in columns 1 and 2 we control for interest rate, a potential concern is that simple controlling for interest rate linearly might not be sufficient to absorb the risk pricing-default rate relationship. In columns 3 and 4 of Table V, we allow for a nonparametric relationship between interest rate and defaults. We make 20 categories of interest rates, and enter them into the equation as fixed effects buckets of interest rate. Our results do not change.

In Panels B and C, we explore the possibility that our tax evasion estimated may be picking up a risk heterogeneity related to default that is only relevant at certain levels of ex ante risk. We divide individuals into 5 interest rate buckets (panel B) and 5 credit grade buckets (panel C). We repeat the analysis regressing default on tax evasion within each quintile to explore whether the relationship between tax evasion and defaults is different within interest rate buckets or credit grade buckets. We find no statistical or economic relationship between default and tax evasion in any of the buckets.

---

[19]Because tax evasion euros are generated regressors, we bootstrap the standard errors in the first estimating equation (estimating the tax evasion $\widehat{\lambda} = \frac{\widehat{\beta_2}}{\widehat{\beta_1}}$ from Eq.(5)) and then take 500 iterations of the estimate of tax evasion to calculate the additional variance around the estimate of effect of tax evasion on default or interest rates. We then add in this additional variance to the estimated standard error of the direct OLS equation of default or interest rates on tax evasion following the procedure in Petrin & Train (2003).

Overall the results in Panels A (column 1-4), and Panels B and C suggest that sound business decision making drives taking hidden income into account in the credit-supply decision. Furthermore the results inform the plausibility of our identification assumption that any unobserved heterogeneity affecting credit decisions is orthogonal to reported income.

In Panel A columns 5 and 6, we test the flip side of profitability of the bank loan, whether our estimate of tax evasion lines up with bank pricing of risk via the application interest rate. The dependent variable is the loan interest rate in percentage points (i.e., a 1.00 is 1%). Consistent with the default results, we do not find any relation between tax evasion and interest rates.

Together these results reaffirm our assertions that our tax evasion estimating methodology is not picking up expected or realized hidden risk.

IV.F   Tax Evasion Incidence

In this section we analyze two aspects of the incidence of tax evasion. First, we investigate whether tax evasion differs between urban and rural areas of the country. Second we analyze whether there is a difference in tax evasion among high and low wealth people. Exploring the heterogeneity of tax evasion along a couple of dimensions may facilitate more precision in our estimation and allows us to offer some further evidence toward considering the distributional effects of tax evasion.

IV.F.1   Tax Evasion and Geography

Table VI repeats the main analysis separately for urban and rural areas of the country. Urban areas include Athens, Thessaloniki and cities with population higher than 100,000 people. The rest of areas, including the islands, are rural. Columns 1-3 provide the simple estimates without the branch fixed effects and other local soft information proxies. We focus on Columns 4-6, which include all controls. Column 4 repeats the estimates for all of Greece, while Columns 5 and 6 provide estimates for individuals living in urban and rural areas respectively. We find an income multiplier λ of 1.78 for urban areas and 1.99 for rural areas. The difference in these two estimates, although meaningful in magnitude, is not statistically significant. Tax evasion is evident in both rural and urban areas, and may be even stronger in rural areas, perhaps because of less formal contracting with large corporations or more social contracting in smaller communities. Table VI seemingly confirms the widely accepted "truth" that tax evasion is pervasive across Greece.

IV.F.2   Tax Evasion and Wealth

An equally important question when thinking about the distributional effects of tax evasion is whether tax evasion resides only with a particular income or wealth profile. We unfortunately cannot do an analysis by true income, because of the endogeneity of tax evasion. However, we can use the neighborhood real estate value variable as a proxy for wealth to consider the wealth incidence of tax evasion.[20] Table VII investigates whether the adaptation multiplier varies by the wealth quintiles based on the real estate measure of the tax authority. Quintile 1 (Column 1) is the lowest wealth quintile and quintile 5 (Column 5) is the highest wealth quintile. We find that in lower wealth areas, the self-employed evade taxes at least as much in percent of income (statistically the same) as they do in the high wealth quintiles.

IV.G   Industry Distribution Results

In this section, we report estimates of tax evasion by industry. Table VIII presents results from estimating equation (5). The dependent variable is credit capacity. We limit our across-industry analysis to the constrained sample because the mortgage sample is only a fifth of the observation size of the constrained sample, limiting our ability to do industry disaggregation. Throughout Table VIII, we assume all wage workers have an equal credit-capacity sensitivity to income. We weight observations in estimation to be population-representative using the tax authority distribution. A final note is that although we include farming as a sector, our data in this sector are not representative, because most farmers bank with a dedicated agricultural bank.

Column 1 is the simplest version of the model, including only year, branch and industry fixed effects. The first row of Table VIII reports the coefficient $(\widehat{\beta}_1)$ on reported income for wage workers. A coefficient of 0.308 implies that for every additional euro of income, wage workers have 31 cents more credit capacity. The next thirteen rows present the coefficients on the self-employed reported income by industry (the $\widehat{\beta}_2$'s). Recall that we identify the income multiplier as $\lambda_j = \frac{\widehat{\beta_2}}{\widehat{\beta_1}} = \frac{\widehat{\beta_1 \lambda_j}}{\widehat{\beta_1}}$. Column (1.a.) represents the multipliers. To give an example of interpretation, the first industry is accounting and finance. Column 1 reports a multiple of $1.71 = 0.526/0.308$ for accounting and finance, implying an evasion rate of 42%. Tax-evaded income in this case is 14, 477 euros, reported

---

[20]In the analysis, we drop the observations whose real estate values appear too frequently in the data to be representative of the continuous formula the tax authority uses to calculate the scale.

in column (1.b.).

In column 2, we include self-employment crossed with industry fixed effects, as well as year and branch fixed effects. We also include the proxies for income growth, income risk and wealth, in levels and interacted with self-employment status. The coefficients on the income-profiling variables are not significant in the presence of credit grade fixed effects. The coefficient on the real estate wealth proxy is positive and significant; a one standard deviation increase in wealth is associated with an additional 487 euros of credit capacity, an increase of 4% to total capacity. Comparing across columns 1 and 2 shows that the tax evasion multiple estimates are robust: some of the industries have slightly smaller multiples, but most of the multiples for the highly-evading industries increase when we control for wealth and industry interacted with self-employment fixed effects.

We find that the highest tax evaders in euros are in law, medicine, education, and engineering. In these industries, the self-employed underreport their incomes by 24,000-29,000 euros, on average, according to our first OLS specification. The inclusion of additional wealth and self-employment controls raises our estimates of underreporting to 36,000-42,000 euros. Other noteworthy tax-evading industries in euros are the media, lodging and restaurants, and business services.

In the final column, we investigate whether the results hold in medians, rather than means, to ensure extreme values do not drive our results. The results are very much in line with the OLS results.[21] The last two rows, show the aggregate multiplier across all industries for each specification. The aggregate multipliers range from 1.72 to 1.95, which translates to tax-evasion rates between 41.80% to 48.71%.

IV.H   Robustness of Industry Distribution to Assumptions

Before looking more deeply at the industry distribution of tax evasion, we consider robustness to the assumptions that wage workers do not evade taxes and that different industries have the same wage-worker sensitivity of credit to income, $\beta_{1j} = \beta_1$. If wage workers evade taxes, we underestimate the overall tax evasion for self-employed individuals, but if tax evasion of wage workers differs in the various industries, the estimated industry distribution of tax evasion can potentially be affected. The industry distribution may also differ if wage workers in industries have different β_{1j}'s.

---

[21]As before, we have to estimate the quantiles regression without branch fixed effects to get convergence. We checked if the OLS results were sensitive to excluding branch fixed effects; they were not.

The mortgage sample is one-fifth the number of observations in the constrained sample, and we do not have power to estimate coefficients for all thirteen industries. Instead, we group industries into three groups for these robustness tests. In the first group, we gather professional services. In Greece, professional services command a distinction of class/respect for their traditional high education and prominent role in the economy and social life. The first group includes Accounting & Financial Services, Education, Medicine, Media & Entertainment, Personal Services, and Law. In the other two groups, we have divided the economy along its two primary functions. Nearly one-fifth of the Greek GDP comes from tourism and its associated industries: lodging, retail, business services and trade. These industries form our second group. We include business services in this category, because professionals in this industry mainly cater to professionals of the group. Finally, construction, transport, engineering and manufacturing constitute the third group, that accounts for the core of the secondary sector of the economy and are characterized by high use of intermediate goods and paper-trail intensity.

In Table IX, we allow the β_{1j}'s (the wage worker income variable coefficient) to be estimated for each sector. In columns 1 and 2, we report results from the constrained sample; in columns 3 and 4, from the mortgage sample; and in column 5, from a mortgage sample in which we limit wage workers to be only from large corporations. The emphasis, thus, is on column 5, because wage workers in column 5 are unlikely to be evading taxes. The first result to note is that for the constrained sample (columns 1 and 2), the coefficients on wage-worker income (β_{1j}'s) range from 0.336 to 0.386, close to the magnitude of β_1 in Table VIII ($\beta_1 = 0.366$). This result suggests that imposing the assumption that $\beta_{1j} = \beta_1$ should not significantly bias the industry rankings in Table VIII. Not surprisingly, the estimated *lambdas* in columns 1a and 2a show a pattern similar to Table VIII, where the highest tax-evading occupations are those that are mainly service-oriented.[22]

Our main focus is columns 3 to 5. In columns 3 and 4, we repeat the analysis of columns 1 and 2, but now the sample is the mortgage sample. The estimated *lambdas* for services are the highest, whereas the estimates *lambdas* for construction are the lowest, which is consistent with the evidence in Table VIII. Most importantly, in column 5, we repeat the analysis excluding all

---

[22]In this setup also we repeat the analysis separately for urban and rural areas. We report these estimates in appendix Table A.7. We find that the prior ranking of professional services as an evasive industry is very strong in both urban and rural areas. In rural areas, which include the islands, we also find a strong effect for the tourism and related sectors.

wage workers, except from those employed in large companies, because the large corporations' wage workers are unlikely to evade taxes. Comparing columns 4 and 5, we find the estimates are robust when we benchmark against only wage workers who are employed in large Greek companies. Overall, column 5 suggests our industry ranking of tax evasion is robust to allowing the sensitivity of credit to true income to vary across industries not only for the self-employed, but also for wage workers, and to considering the potential tax evasion of wage workers.

IV.I Economic Significance by Industry & Outside Validation

Table X summarizes the industry distribution results. Column 1 ranks the tax-evading industries in euros from the estimates in Table VIII (the average tax evasion is calculated using the average of the coefficients in columns 2 and 3). This distribution is not what one would expect when thinking about tax evasion in very formal economies or very informal economies. The highest tax evaders are services requiring advanced degrees and certification, and whose revenue depends on reputation (e.g., doctors, lawyers, engineers, journalists, tutors, accountants, and financial agents).[23] Columns 4 and 5 of Table X list the degree requirement and issuing agency for certification. The distribution suggests semiformal economies are not just informal economies developing into formal, but have a unique character, with the highly educated being very much part of the system, perhaps even allowing the semiformal nature to persist.

We offer an additional piece of evidence to validate this professional-like distribution. In 2010, a bill was proposed to the Greek Parliament that targeted a group of 11 occupations. The bill targeted certain professions as most likely to evade, and included provisions for audits of taxpayers in these professions that reported lower income than a pre-specified threshold of 20,000 euros. The occupations that the bill targeted were doctors, dentists, veterinarians, lawyers, architects, engineers, topographer engineers, economists, business consultants, tax auditors, and accountants. Our estimates of the big tax offenders coincide very well in comparison (Table X, column 3). The Greek Parliament rejected the bill, a point to which we return to later.

---

[23]Education emerges as a big tax-evading industry. To a non-Greek, this emergence may seem odd. However, the system in Greece is such that anyone with a little excess disposable income hires private tutors for their children. Not surprisingly, the private sector of tutoring is lucrative and unrecorded.

V Institutions in Semiformal Economies

In this section, we use the estimated industry distribution to better understand some of the factors and norms that allow tax evasion to persist in semiformal economies.

V.A Paper Trail

Prior research suggests that tax evasion is higher in cash sectors (Klepper & Nagin [1989], Morse, Karlinsky & Bankman [2009]) and in situations when individuals think underreporting is less likely to be uncovered (Dubin, Graetz & Wilde [1990], Blumenthal et al. [2001], Alm, Jackson & McKee [2009], Kleven et al. [2011], Pomeranz [2015]).The implication is that in the cross-section of industries, compliance should be higher in industries with traceable information. To explore this idea, we develop a measure of traceability of income based on the positioning of the business in the upstream-downstream spectrum, in particular, how much the business produces or uses intermediate input goods. Rather than face the selection and biases of constructing such a measure in accounting data, we apply a survey instrument. We survey a class of executives in an executive business program in Greece.[24] We asked the participants to score each industry on a scale of 1-5 on (i) the use of intermediate goods as inputs and (ii) the extent to which the output is an intermediate good.

Table XI (Panel A) presents the mean scores of paper-trail input and output by industry, where we have de-meaned individuals' responses to capture any level biases by individuals. In Panel B, we form tertiles (High, Medium, and Low) based on paper trail intensities and report the average tax-evasion of the respective industries. The correlations suggest that industries with low paper trail align closely with our top tax-evading industries. Figure III plots this intuition. Industries with the lowest paper trail include some of our biggest tax evading professional services: law, education, and medicine. Conversely, an industry such as transport, which involves bills of laden and other tracking documents displays a lower tax evasion rate. We do not try to assert causation in this simple correlation exercise, but we find this evidence consistent with the story of Kleven, Kreiner & Saez (2009), Kleven et al. (2011), and Pomeranz (2015) that paper trail is a factor that contributes

---

[24]We conducted the survey at the Executive M.Sc. Program of Banking and Finance at the University of Piraeus. We chose this program based on the experience of the participants in inflows and outflows of industries and the accounting thereof.

to higher compliance.

V.B   Enforcement

We gather data from the tax authority on enforcement cases closed and total assessment values for each of the 235 Greek tax districts from January 2011 to July 2012. We are interested in testing (i) whether tax-enforcement offices know about the prevalence of tax evasion, and (ii) whether they are able to enforce actions against tax offenders. Weak enforcement may result from the aforementioned story where enforcement is hampered by the lack of paper trail that makes it harder for tax authorities to prove evasion (Klepper & Nagin [1989]). However, weak enforcement also results from a host of other stories, including corruption (Azariadis & Ioannides [2014], Shleifer & Vishny [1993]) and weak state capacity (Besley & Persson [2010]). Weak state capacity refers to both the general ability of the state to raise revenues (fiscal capacity), which could relate to paper trail, and the power to enforce contracts and support markets (legal capacity). Greece suffers from weak state capacity on both counts (IMF [2014], Council of Europe [2014]). Papaioannou & Karatza (2015) in an excellent overview of the deficiencies of the Greek judicial system note that over 400,000 cases are currently pending in administrative courts, each taking from 5 to 10 years to complete, while the low clearance ratio (below 0.80) implies that the situation is expected to aggravate in the future. Our goal in this section is simply to present correlations of enforcement effectiveness and tax evasion which would be consistent with a number of the causes above, not to pin down the exact bundle of frictions causing weak tax collection in Greece.

Ideally, we would overlay an industry distribution of households onto the 235 tax-office districts so that we can compare enforcement statistics with tax evasion implied by industries. However, our sample is not sufficiently large to be representative at that level. Thus, we instead study how tax enforcement relates to the density of self-employment and wealth in each district, since our largest tax-evading industries map to wealthy occupations on average.

Panel B of Table XII reports evidence on the closure of tax cases using an OLS elasticity model (columns 1-3) and a Poisson model (columns 4-6). Columns 1 and 4 show that the elasticity and the Poisson-rate sensitivity of cases closed are strongly and significantly positively related to self-employment percentage. In the other specifications of Panel B, we add the interaction of wealth and self-employment. In columns (3) and (6) we orthogonalize the variables using the modified

Gram-Schmidt procedure of Golub & Van Loan (1996).[25] We find that local tax officials close more cases especially in places with wealthy self-employed individuals. Panel C repeats the exercise for the assessment amount. Across specifications, we find the percentage of self-employed individuals explains the location of assessments, but that assessment amounts do not increase where there is higher wealth or self-employed wealth. The results collectively suggest that tax enforcers know where tax evasion exists, but lack the ability or the willpower to collect evaded revenues, because of a host of institutional frictions and the lack of paper trail to prove violations.

V.C Semiformality and Political Economy

In 2009, the Greek Parliament introduced a bill, mandating tax audits for declared income below 20,000 euros on certain self-employed professionals, including doctors, lawyers, and engineers. The quick dismissal of this bill triggers the question of whether politicians look out for their own profession or act as functionaries in powerful industry associations and guilds in a way the fails to act on needed tax reform to address the problem.[26]

We collect data on the occupational backgrounds of the Greek parliamentarians for the year 2009 from their official websites to ask whether the distribution of tax-evading industries correlates to the politician occupational backgrounds. Table XIII presents these descriptive results, ordered by our tax evasion estimates. Column 5 reports the cumulative distribution of parliamentary members, excluding lawyers, which is the default profession for politicians. Out of 13 industries, the four most tax-evading industries (medicine, engineering, education, and media) account for 61.7% of the non-lawyer Parliamentary votes. If we add in the next two categories (tourism and accounting/finance), the percentage rises to over 80%, excluding lawyers, which would also be high on tax evasion. The alignment of the occupational backgrounds of Greek parliamentarians to our top tax-evading industries is only an association, but may suggest one possible reason behind the lack of willpower to enact tax reform.[27]

---

[25]This technique gives the heaviest weight to the level variables and lets the interaction try to explain the variance of what is left over (i.e. ensuring robustness of the interaction coefficient to multicollinearity concerns). We also report results without orthogonalization in columns (2) and (5); however these results should be interpreted with caution, because the variance inflation factors (VIFs) are over 200 in our small sample.

[26]To our knowledge, little academic research explores political economy stories related to politician occupations. An exception is Faccio (2006), who offers evidence that politicians support their family industries.

[27]The *New York Times*, February 2, 2012, reports: "In Greece, the real power is the power of resistance, the power of inertia", said Giorgos Floridis, a former member of Parliament from the Socialist Party, who recently founded a reform-minded civic movement. "Today, the main power centers in Greece - political parties, business leaders,

VI Discussion: Method Contribution

We claim to offer a methodological contribution to the tax evasion estimation literature, especially relevant in places where cash transactions are pervasive. Our innovation is the notion that, in such settings, one can use observable private sector adaptions to cash-based economies as indicators of hidden income. As Table A.1 shows, most of the world's population lives in semiformal settings. In the paper, we use the adaptation by the banking sector to develop a new method to estimate tax evasion based on bank loan information. Nevertheless, our methodological contribution is reproducible with more generally-available data, as we illustrate in the appendix by applying simple debt-to-income rules in survey data from the United States in which income may be underreported. In this section, we briefly re-visit the arguments that the tax-audit method and the consumption method may lack power to identify the full extent of tax evasion in semiformal economies and tie these assertions to what we have shown in our study on factors allowing persistence of tax evasion.

Prior literature, has identified the difficulties of tax audits to uncover undetectable tax evasion. Slemrod & Weber (2012) states that tax audits struggle to uncover tax evasion primarily because cash transactions are not only hidden, but unrecoverable. Kleven et al. (2011) explain that "despite the large amount of resources spent on these audits, they are unlikely to uncover all tax evasion for all taxpayers." Tax audits would naturally be even less effective in many semiformal countries where tax authorities are resource constrained. Kleven et al. (2011) continue, "The same issue arises in the TCMP [tax compliance] studies, which blow up detected tax evasion by a multiplier of 3.28 to arrive at the official U.S. tax evasion estimates. Unfortunately, this multiplier is large and has a very large measurement error, so that total evasion rates are at best rough approximations." In the prior section, we reported a negative correlation between paper trail and our estimated tax evasion across industries. This correlation would be consistent with this literature-based understanding of cash transactions being not just hidden, but unrecoverable in many instances, rendering tax audits without amble power to prove hidden income. We also show correlations suggesting that tax-enforcement officials are unwilling or unable to issue more assessments and fines for evasion cases in places where they seem to know that tax evasion exists. Such correlations are consistent with the arguments of Slemrod & Weber (2012) that enforcement officers cannot prove tax evasion

---

professional guilds, public sector unions and the media - are fighting to preserve their privileges, blocking structural changes that could make the economy more functional."

magnitudes, although many other reasons may be at work why tax collections are low.

A final note on tax audits concerns the reality that, in countries with pervasive tax evasion, tax officials often resort to consumption audits as indicators of income. However, when tax officials audit consumption in their procedures to uncover hidden income, a norm emerges of populations hiding and distorting consumption. This is a trademark of semiformal economies.[28]

Measuring tax evasion off consumption is the other primary tax evasion method. In all but the most formal economies, consumption surveys are rare due to the expense and the difficulties involved. Even if costs were not a constraint, these consumption surveys would likely be fraught with underreporting of expenses, because of the hiding norm in semiformal economies. [29]

VII   Conclusion

We start with the insight that in semiformal economies, the private sector adapts to the culture of semiformality. Using this observation, we develop a new methodology to estimate tax evasion. We observe the credit extended by the banks, and estimate the true income that these credit capacities imply.

Using individual-level household-lending data across different credit products and samples, we conservatively estimate that at least 26.8 billion euros of income went untaxed in Greece for 2009. At the tax rate of 40%, the foregone tax revenues account for 30% of the county's deficit. Ranked by tax-evaded euros, the largest offending industries are medicine, engineering, education, accounting and financial services, and law. We provide a host of tests and delve in detail into the underwriting

---

[28]For example, in 2010, the Greek tax authorities attempted to uncover part of the unreported income by imposing taxes on presumed income, based on wealth criteria. One of these criteria was the ownership of a swimming pool. Only 324 taxpayers reported on their tax returns that they owned a swimming pool; however, the use of Google maps revealed almost 17,000 swimming pools across the country ("*Greek Wealth Is Everywhere but Tax Forms,*" New York Times, May 1st, 2010).

[29]We conduct an illustrative exercise of estimating the PW consumption equation using car loan application data from Greek bank, with the dependent variable $\log car_i$ being the car collateral value measured one time per individual i. We estimate:

$$\log car\ consumption_i = \underset{[.016]}{.337} \log Y_i^R + \underset{[.017]}{.118} SE_i + \underset{[.0356]}{.146} \log age + \underset{[.011]}{.110} male + \underset{[.012]}{.131} homeowner - \underset{[.150]}{.146} \log dependents$$

$$+ \underset{[.022]}{.035} \log realestate + \underset{[.006]}{.003} \log sd + \underset{[.233]}{.101} \log inc.growth + f.e.(year * geography).$$

with a vector of control variables Z that includes $LogAge$, $Gender$, $Homeownership$, $LogDependents$, the income variance local proxy ($LogStDevLogIncome$), the wealth proxy ($LogRealEstate$), and local conditions ($LogIncomeGrowth$). Using the PW derivation that $\lambda^{pw} = \exp\left(\frac{\alpha_2}{\alpha_1}\right) = \exp\left(\frac{.118}{.337}\right)$, we find that true income is 1.42 times reported income, a smaller tax evasion rate of 29.5% relative to our bank method estimate of 45%. We repeat the caveat that this exercise is only illustrative due to our data limitations.

and lending processes of the bank to ensure our estimates are robust to adjusting for risk, wealth, or other differences between the self-employed and wage workers or differences among the industries.

The estimated industry distribution offers valuable insights. The highest tax evaders are the highly educated and powerful professions in Greece. We show that these professions together hold the majority of the Greek Parliament, and that they are governed by strong professional guilds. We document this association as suggestive of one possible reason tax evasion may persist, but also to highlight some of the differences in thinking about semiformal settings.

Another thought emerging from the distribution of tax-evading industries concerns potential measures that policymakers could take. Slemrod (2007) notes that "good tax policy should be designed with the realities of evasion in mind". Understanding the realities of tax evasion can foster good tax policy and efficient enforcement. Presumed income-tax initiatives, under which the government mandates a minimum tax reporting by occupation, are politically unsavory. However, policy professionals might consider whether occupation licenses for tax-evading industries might line up with our rankings. City authorities worldwide use business licenses to collect revenues, and in Greece, the list of professions evading taxes is strongly related to the existence of mandatory industry associations, providing a potential forum for collections.

A final thought concerns the welfare implications of banks issuing soft credit. In a Coasian or De Soto view of the world, the fact that banks give an entitlement to informal income provides a property right that allows individuals to use borrowing to optimally smooth lifetime consumption or overcome shocks. A negative side is that access to soft credit reduces the costs to informality, but we would be surprised if the negative outweighs the positive in the second-best world of a norm of semiformality.

References

Acharya, V., Drechsler, I. & Schnabl, P. (2014), 'A pyrrhic victory? bank bailouts and sovereign credit risk', *The Journal of Finance* **69**(6), 2689–2739.

Agarwal, S. & Hauwald, R. (2010), 'Distance and private information in lending', *Review of Financial Studies* **23**(7), 2757–2788.

Alexeev, M. & Pyle, W. (2003), 'A note on measuring the unofficial economy in the former soviet republics1', *Economics of Transition* **11**(1), 153–175.

Alm, J., Jackson, B. R. & McKee, M. (2009), 'Getting the word out: Enforcement information dissemination and compliance behavior', *Journal of Public Economics* **93**(3), 392–402.

Altonji, J. G., Elder, T. E. & Taber, C. R. (2005), 'Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools', *Journal of Political Economy*.

Ang, A. & Longstaff, F. A. (2013), 'Systemic sovereign credit risk: Lessons from the us and europe', *Journal of Monetary Economics* **60**(5), 493–510.

Azariadis, C. & Ioannides, Y. M. (2014), Thinking about corruption in Greece, Technical report, Department of Economics, Tufts University.

Besley, T. & Persson, T. (2010), 'State capacity, conflict, and development', *Econometrica* **78**(1), 1–34.

Black, T., Bloomquist, K., Emblom, E., Johns, A., Plumley, A. & Stuk, E. (2012), 'Federal tax compliance research: tax year 2006 tax gap estimation', *IRS Research, Analysis & Statistics Working Paper*.

Blumenthal, M., Christian, C., Slemrod, J. & Smith, M. G. (2001), 'Do normative appeals affect tax compliance? evidence from a controlled experiment in minnesota', *National Tax Journal* pp. 125–138.

Braguinsky, S., Mityakov, S. & Liscovich, A. (2014), 'Direct estimation of hidden earnings: Evidence from russian administrative data', *Journal of Law and Economics* **57**(2), pp. 281–319.

Brunnermeier, M., Garicano, L., Lane, P., Pagano, M., Reis, R., Santos, T., Van Nieuwerburgh, S. & Vayanos, D. (2011), 'European safe bonds (ESBies)'.

Cagan, P. (1958), The demand for currency relative to total money supply, *in* 'The Demand for Currency Relative to Total Money Supply', National Bureau of Economic Research, Inc, pp. 1–37.

Callen, M. & James, D. (2015), 'Institutional corruption and election fraud: Evidence from a field experiment in Afghanistan', *American Economic Review* **105**(1), 354–81.

Christian, C. W. (1994), 'Voluntary compliance with the individual income tax: Results from the 1988 TCMP study', *The IRS Research Bulletin* **1500**, 35–42.

Chudik, A. & Fratzscher, M. (2011), 'Identifying the global transmission of the 2007–2009 financial crisis in a gvar model', *European Economic Review* **55**(3), 325–339.

Clotfelter, C. T. (1983), 'Tax evasion and tax rates: An analysis of individual returns', *The review of economics and statistics* pp. 363–373.

Çule, M. & Fulton, M. (2009), 'Business culture and tax evasion: Why corruption and the unofficial economy can persist', *Journal of Economic Behavior & Organization* **72**(3), 811–822.

Del Carpio, L. (2013), 'Are the neighbors cheating? evidence from a social norm experiment on property taxes in peru', *Working paper*.

Djankov, S., Ganser, T., McLiesh, C., Ramalho, R. & Shleifer, A. (2010), 'The effect of corporate taxes on investment and entrepreneurship', *American Economic Journal: Macroeconomics* **2**, 31–64.

Dubin, J. A., Graetz, M. J. & Wilde, L. L. (1990), 'The effect of audit rates on the federal individual income tax, 1977-1986', *National Tax Journal* pp. 395–409.

Dwenger, N., Kleven, H. J., Rasul, I. & Rincke, J. (2014), 'Extrinsic and intrinsic motivations for tax compliance: Evidence from a field experiment in germany', *Working paper*.

Faccio, M. (2006), 'Politically connected firms', *The American Economic Review* **96**(1), 369–386.

Feinstein, J. S. (1999), 'Approaches for estimating noncompliance: Examples from federal taxation in the united states', *The Economic Journal* **109**(456), 360–369.

Feldman, N. E. & Slemrod, J. (2007), 'Estimating tax noncompliance with evidence from unaudited tax returns', *The Economic Journal* **117**(518), 327–352.

Fellner, G., Sausgruber, R. & Traxler, C. (2013), 'Testing enforcement strategies in the field: Threat, moral appeal and social information', *Journal of the European Economic Association* **11**(3), 634–660.

Galbiati, R. & Zanella, G. (2012), 'The tax evasion social multiplier: Evidence from Italy', *Journal of Public Economics* **96**(5), 485–494.

Goel, R. K. & Saunoris, J. W. (2014), 'Global corruption and the shadow economy: spatial aspects', *Public Choice* **161**(1-2), 119–139.

Golub, G. H. & Van Loan, C. F. (1996), 'Matrix computations', *Johns Hopkins studies in mathematical sciences*.

Gorodnichenko, Y., Martinez-Vazquez, J. & Peter, K. S. (2009), 'Myth and reality of flat tax reform: Micro estimates of tax evasion response and welfare effects in Russia', *Journal of Political Economy* **117**(3), pp. 504–554.

Hallsworth, M., List, J., Metcalfe, R. & Vlaev, I. (2014), The behavioralist as tax collector: Using natural field experiments to enhance tax compliance, Technical report, National Bureau of Economic Research.

Hasker, K. & Okten, C. (2008), 'Intermediaries and corruption', *Journal of Economic Behavior & Organization* **67**(1), 103–115.

Hurst, E., Li, G. & Pugsley, B. (2014), 'Are household surveys like tax forms? Evidence from income underreporting of the self-employed', *Review of economics and statistics* **96**(1), 19–33.

Klepper, S. & Nagin, D. (1989), 'The anatomy of tax evasion', *Journal of Law, Economics and Organization* **5**, 1.

Kleven, H. J. & Waseem, M. (2011), 'Tax notches in pakistan: Tax evasion, real responses, and income shifting', *Quarterly Journal of Economics, Forthcoming*.

Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S. & Saez, E. (2011), 'Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark', *Econometrica* **79**(3), 651–692.

Kleven, H. J., Kreiner, C. T. & Saez, E. (2009), Why can modern governments tax so much? An agency model of firms as fiscal intermediaries, Technical report, National Bureau of Economic Research.

La Porta, R. & Shleifer, A. (2008), 'The unofficial economy and economic development', *Brookings Papers on Economic Activity* **2008**(2), 275–363.

Lackó, M. (1999), *Do power consumption data tell the story?: Electricity intensity and hidden economy in post-socialist countries*, Inststitute of Economics, Hungarian Academy of Sciences.

Lane, P. R. (2012), 'The european sovereign debt crisis', *The Journal of Economic Perspectives* **26**(3), 49–67.

Lyssiotou, P., Pashardes, P. & Stengos, T. (2004), 'Estimates of the black economy based on consumer demand approaches*', *The Economic Journal* **114**(497), 622–640.

Morse, S. C., Karlinsky, S. & Bankman, J. (2009), 'Cash business and tax evasion', *Stanford Law and Policy Review* **20.1**, 37–68.

Papaioannou, E. & Karatza, S. (2015), The Greek justice system: Collapse and reform, *in* C. Meghir, P. Christopher, V. Nikos & V. Dimitri, eds, 'Reforming the Greek Economy', MIT Press.

Paula, . d. & Scheinkman, J. A. (2010), 'Value-added taxes, chain effects, and informality', *American Economic Journal: Macroeconomics*.

Petersen, M. A. (2004), Information: Hard and soft, mimeo, Northwestern University.

Petersen, M. A. & Rajan, R. G. (2002), 'Does distance still matter? The information revolution in small business lending', *The Journal of Finance* **57**(6), 2533–2570.

Petrin, A. & Train, K. (2003), Omitted product attributes in discrete choice models, Technical report, National Bureau of Economic Research.

Pissarides, C. A. & Weber, G. (1989), 'An expenditure-based estimate of Britain's black economy', *Journal of Public Economics* **39**(1), 17–32.

Pomeranz, D. (2015), 'No taxation without information: Deterrence and self-enforcement in the value added tax', *American Economic Review* **105**(8), 2539–69.

Schneider, F. & Enste, D. (2000), *Shadow Economies Around The World-Size, Causes, and Consequences*, Vol. 196, International Monetary Fund.

Shambaugh, J. C. (2012), 'The euro's three crises', *Brookings Papers on Economic Activity* pp. 157–207.

Shleifer, A. & Vishny, R. W. (1993), 'Corruption', *The Quarterly Journal of Economics* **108**(3), 599–617.

Slemrod, J. (2007), 'Cheating ourselves: the economics of tax evasion', *The Journal of Economic Perspectives* pp. 25–48.

Slemrod, J. & Weber, C. (2012), 'Evidence of the invisible: toward a credibility revolution in the empirical analysis of tax evasion and the informal economy', *International Tax and Public Finance* **19**(1), 25–53.

Tanzi, V. (1983), 'The underground economy in the united states: Annual estimates, 1930-80', *IMF Staff Papers* **30**(2), 283–305.

Figure I: National & Sample Reported Income Distributions
The graph presents the distribution of national reported income based on all taxpayers in Greece, along with the distribution of reported income for the (a) constrained and the (b) mortgage sample. Tax-authority data refer to the personal income reported to the tax authority by all taxpayers for the year 2009 and is from the the annual statistical bulletin of the Ministry of Finance. The reported income of the constrained and the mortgage sample is from the loan applications of the bank. Applicants with income over €1.5 million or under €1,000 are removed from the estimation samples. Income buckets become wider as income increases following the reporting standards of the annual statistical bulletin.
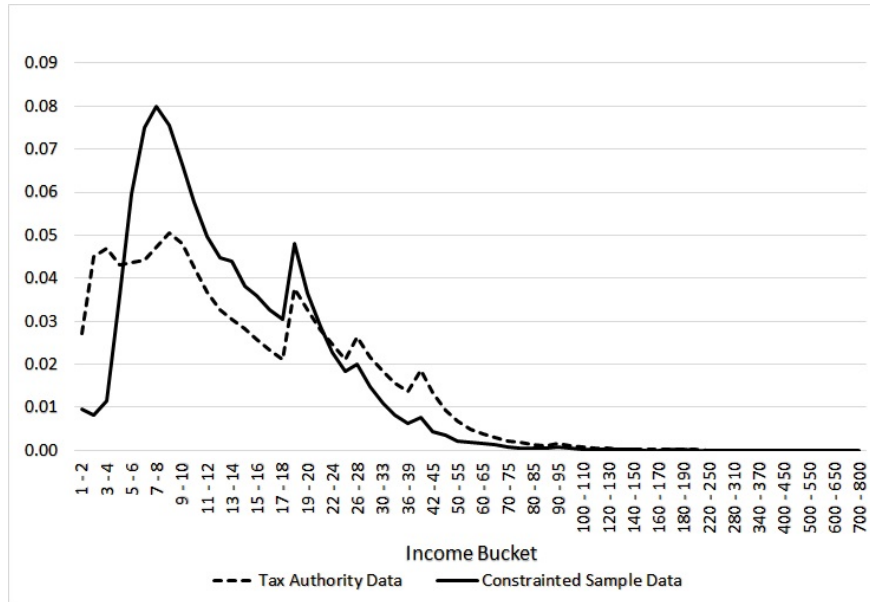
(a) National & Constrained Sample Reported Income Distributions
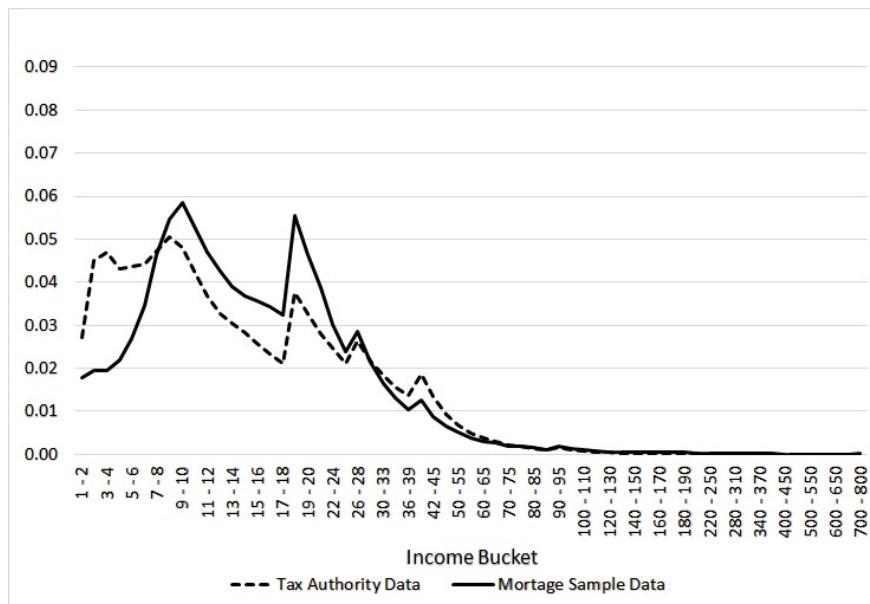
(b) National & Mortgage Sample Reported Income Distributions

Figure II: Linear and Spline Specification of Credit Capacity

The figure plots estimated credit capacity as a function of reported income, comparing the linear specification to a spline specification, implementing equation (5). The dependent variable is credit extended and the independent variables include reported income for wage workers, reported income for the self-employed, self-employment indicator, credit grade buckets, and the proxies for local conditions and income risk, interacted with self-employment. The sample consists of applicants whose approved loan amount is lower than the amount requested (constrained sample). The vertical axis presents the prediction from this model for a linear regression, and a regression that takes 8 linear spline regions with nodes of 1,000, 5,000, 10,000, 15,000, 20,000, and 30,000 euros for wage worker and self-employed reported income. Spline estimates are depicted by dashed lines according to the legend. Also included in the figure is a histogram of reported income for the constraint sample from the loan applications of the bank.
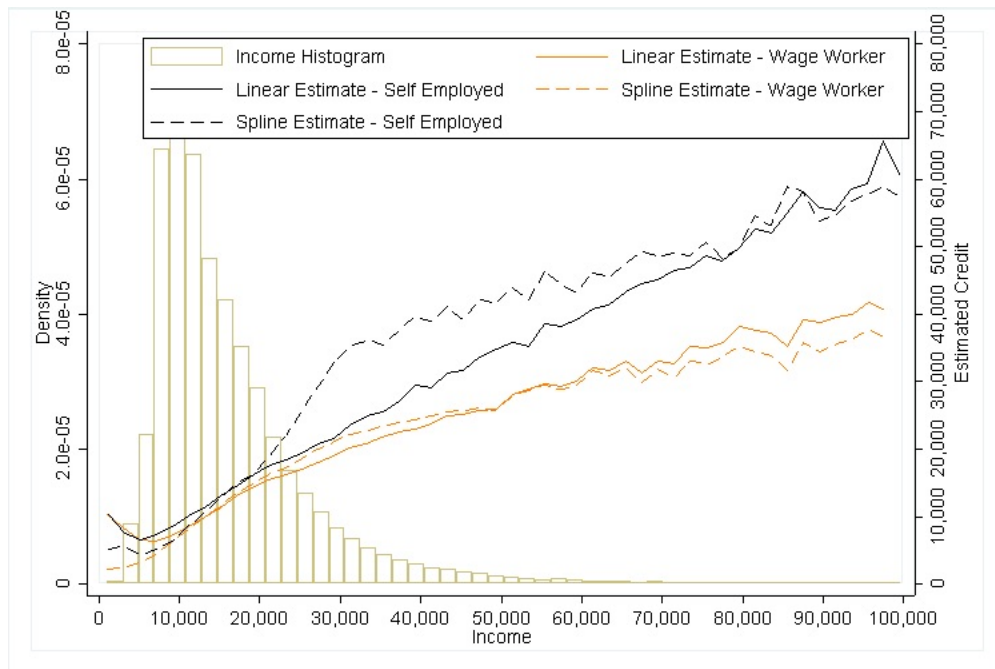
Figure III: Paper-Trail Intensity and Industry Tax Evasion

The figure presents the paper-trail intensity with respect to the estimated tax-evaded income in euros across industries. Paper-trail intensities (x-axis) are defined as the sum of input and output intensities from Table XI. The intensities come from a survey of executives in a managerial educational program in Greece. The scale is from 1 to 5 (low to high intensity) and answers are demeaned at the respondent level to remove any individual biases. The estimates of tax-evaded income (y-axis) are from Table X.
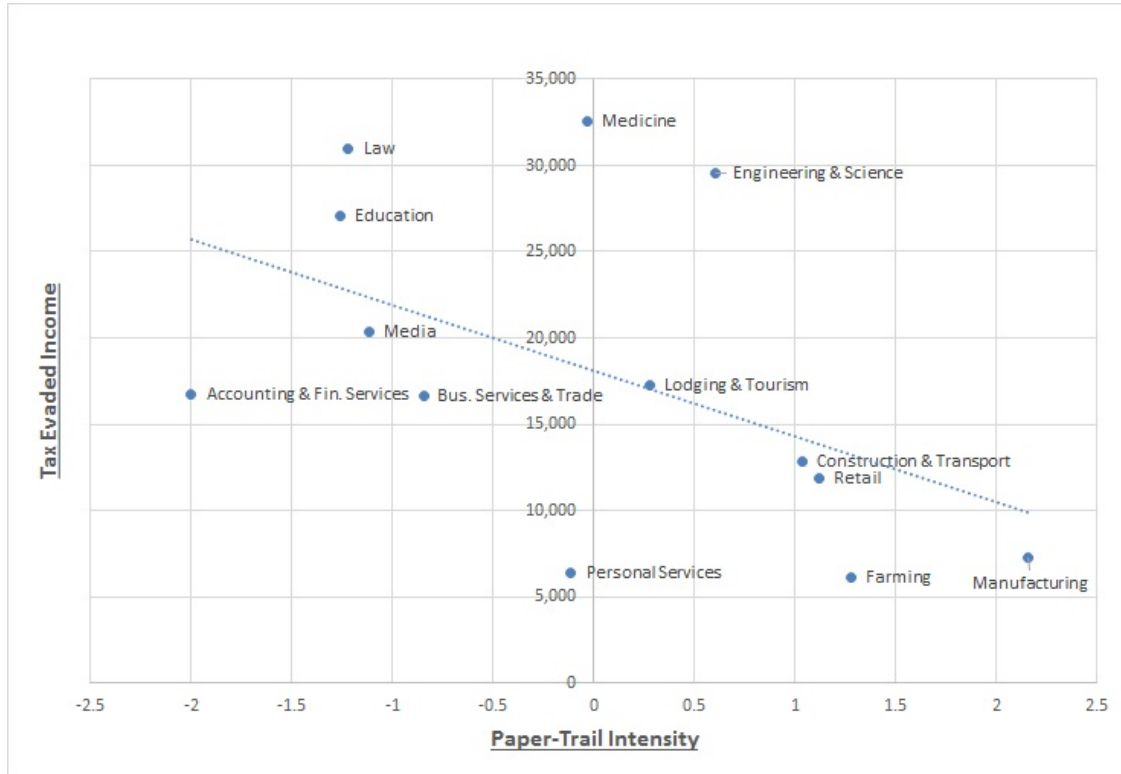
Tables

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

Table II: Tax-Evasion Estimation of Self-employed (Constraint Sample)

The table presents estimated coefficients [col. (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. *WageWorker* 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 5 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 zip code of the applicant's residence (*RealEstateWealth*). 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.

	Dependent Variable: Credit Capacity = Outstanding Debt + Approved Loan							
	(1)	(1a)	(2)	(2a)	(3)	(3a)	(4)	(4a)
	OLS	λ	OLS	λ	OLS	λ	Quantile	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%

44

Table III: Aggregates at the Economy Level

The table presents aggregate estimates of tax evasion at the economy level. The lambda multiples (λ) are from Table II. The aggregate reported income for the self-employed is from the annual bulletin of the Ministry of Finance. A range of economy-wide tax evaded income for the self-employed individuals and foregone taxes is calculated based on low and high estimated lambda multiples from Table II. Estimations based on the median lambda multiple are provided in parentheses. Foregone taxes are estimated under a 40% tax rate. Total and primary (in parentheses) deficits are from Eurostat. Amounts are in billions of euros.

	Lambda Estimate:	Low 1.75	Median 1.79	High 1.84
	Reported Income (Self-Employed)	Tax Evaded Income Low-High (Median)	Foregone Taxes Low-High (Median)	Deficit (Primary)
Year 2006	28.8	21.6 - 24.2 (22.8)	8.64 - 9.68 (9.10)	8.0 (3.6)
Year 2007	30.4	22.8 - 25.5 (24.0)	9.12 - 10.21 (9.60)	14.8 (5.0)
Year 2008	32.4	24.6 - 27.6 (25.9)	9.84 - 11.02 (10.36)	23.3 (12.1)
Year 2009	35.7	26.8 - 29.9 (28.2)	10.71 - 12.00 (11.28)	35.4 (24.3)

Table IV: Tax-Evasion Estimation of the Self-employed (Mortgages) & Large Firm Test

The table presents estimated coefficients [col. (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. *WageWorker* 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 5 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 zip code of the applicant's residence (*RealEstateWealth*). 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. 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.

	Dependent Variable: Monthly Payment of Approved Mortgage							
	All Wage Workers				Wage Workers in Large Firms			
	(1) OLS	(1a) λ	(2) OLS	(2a) λ	(3) OLS	(3a) λ	(4) OLS	(4a) λ
Income*WageWorker	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(IncomeGrowth)			-1,917***				-1,229***	
			[448.4]				[309.5]	
Real EstateWealth			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%

46

Table V: Defaults, Interest Rates, and Tax Evasion

The sample consists of self-employed applicants, whose loan amount approved is lower than the amount requested (constrained sample). The sample collapses to one observation per applicant. The dependent variable in columns (1)-(4) of Panel A and Panel B is the dummy variable Defaults, which takes the value 1 if the account is ever more than 90 days delinquent and 0 otherwise. The dependent variable in columns (5)-(6) of Panel A and Panel C is the loan interest rate. Interest rates are in percentage points (i.e., a 1.00 is 1%). Tax evasion for each applicant is calculated based on the estimates from equation 5 and is expressed in 10,000's euros. Because tax evasion euros are generated regressors, we bootstrap the standard errors in the first estimating equation (estimating the tax evasion $\widehat{\lambda} = \widehat{\beta_2}/\widehat{\beta_1}$ from Eq.(5)) and then take 500 iterations of tax evasion estimate to calculate the additional variance around the estimate of effect of tax evasion on interest rates or default. We then add in this additional variance to the estimated standard error of the direct OLS equation of interest rates or default on tax evasion following the procedure in Petrin & Train (2003). Loan amount is the approved loan amount and duration is the approved duration of the loan (in months). In Panels B and C the sample is separate in interest rate and credit grade quintiles respectively. All specifications include year fixed effects. Specifications in columns (2), (4) and (6) of Panel A and Panel B include 50 credit-grade buckets as risk fixed effects. Specifications in columns (3) and (4)f Panel A and Panel C also include 20 interest-grade buckets as risk fixed effects. Heteroscedasticity robust standard errors are presented in brackets. Significance at the 1%, 5%, and 10% levels is indicated by ***, **, and *, respectively.

Panel A

Dep.Variable:	Defaults	Defaults	Defaults	Defaults	Interest Rate	Interest Rate
	(1)	(2)	(3)	(4)	(5)	(6)
Tax Evasion	-0.00451	0.00226	-0.00380	0.00162	-0.0267	-0.0135
	[0.00479]	[0.00485]	[0.00461]	[0.00466]	[0.0375]	[0.0355]
Interest Rate	0.0193***	0.0172***				
	[0.00350]	[0.00349]				
Loan Amount	-0.0169	0.00715	-0.0188	0.00178	-0.502***	-0.473***
	[0.0140]	[0.0145]	[0.0131]	[0.0130]	[0.170]	[0.170]
Duration	0.00307***	0.00299***	0.00185***	0.00180***	0.0133***	0.0128***
	[0.00055]	[0.00055]	[0.000568]	[0.000561]	[0.00259]	[0.00260]
Credit Grade F.E.	No	Yes	No	Yes	No	Yes
Interest Rate F.E.	No	No	Yes	Yes	No	No
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.0187	0.0494	0.168	0.188	0.160	0.1674

Panel B

Dep.Variable:	Defaults	Defaults	Defaults	Defaults	Defaults
Interest Rate Quintiles:	(Q1)	(Q2)	(Q3)	(Q4)	(Q5)
Tax Evasion	-0.00105	0.0132	-0.00187	-0.00449	-0.00679
	[0.00515]	[0.0107]	[0.0228]	[0.0125]	[0.0148]
Credit Grade F.E.	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.067	0.041	0.17	0.072	0.028

Panel C

Dep.Variable:	Defaults	Defaults	Defaults	Defaults	Defaults
Credit Grade Quintiles:	(Q1)	(Q2)	(Q3)	(Q4)	(Q5)
Tax Evasion	0.00381	-0.00627	-0.00529	0.0177	0.00155
	[0.0216]	[0.0114]	[0.0177]	[0.0147]	[0.00643]
Interest Rate F.E	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.174	0.022	0.00668	0.015	0.011

Table VI: Tax-Evasion Estimation of the Self-employed (Constrained) & Urban vs Rural Test

The table presents estimated coefficients [col. (1), (2), (3), (4), (5) and (6)] 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 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. *WageWorker* 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 5 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 zip code of the applicant's residence (*RealEstateWealth*). The "All" columns refer to the entire sample, the "Urban" columns include applicants inhabiting in Athens, Thessaloniki and cities with population higher than 100,000 people and the "Rural" columns all other areas. All specifications include 50 credit-grade buckets as risk fixed effects, industry interacted with self-employment fixed effects and year fixed effects. Specifications (4)-(6) include branch fixed effects. Observations are weighted to be population-representative using the tax authority distribution. The last row reports the tax-evasion rate, defined as $\frac{Y^{True}-Y^R}{Y^{True}}$. 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.

	All		Urban		Rural		All		Urban		Rural	
	(1)	(1a)	(2)	(2a)	(3)	(3a)	(4)	(4a)	(5)	(5a)	(6)	(6a)
	OLS	λ	OLS	λ	OLS	λ	OLS	λ	OLS	λ	OLS	λ
Income*WageWorker	0.3204***		0.3325***		0.2803***		0.3416***		0.3486***		0.3148***	
	[0.0465]		[0.0699]		[0.0536]		[0.0543]		[0.0793]		[0.0631]	
Income*SE	0.5561***	1.74***	0.5304***	1.60**	0.5673***	2.02***	0.6239***	1.83***	0.6201***	1.78***	0.6277***	1.99***
	[0.0596]		[0.0923]		[0.0783]		[0.0511]		[0.0773]		[0.0676]	
IncomeRisk							1,824		1,218		1,187	
							[2,980]		[4,099]		[3,320]	
SE*IncomeRisk							2,073		11,412		1,700	
							[3,179]		[8,789]		[3,454]	
Lag(Income Growth)							1,841		28,110		-13,030	
							[9,877]		[21,291]		[8,904]	
SE*Lag(Income Growth)							8,060		-13,114		22,170**	
							[10,549]		[23,312]		[9,706]	
Real Estate Wealth							0.9500***		0.7244*		0.5811	
							[0.3200]		[0.4009]		[0.6587]	
SE*Real Estate Wealth							-0.2321		1.2281		-1.2751*	
							[0.5858]		[1.1145]		[0.7489]	
Credit-Grade F.E.	Yes		Yes		Yes		Yes		Yes		Yes	
Industry*SE F.E.	Yes		Yes		Yes		Yes		Yes		Yes	
Branch F.E.	No		No		No		Yes		Yes		Yes	
Year F.E.	Yes		Yes		Yes		Yes		Yes		Yes	
Adj. R^2	0.1001		0.0847		0.1287		0.11200		0.1859		0.1475	
Tax-Evasion Rate	42.53%		37.50%		50.49%		45.36%		43.82%		49.75%	

48

Table VII: Tax Evasion Estimation of the Self-employed (Constrained) by Wealth Quintile

The table presents estimated coefficients [col. (1), (2), (3), (4) and (5)] from equation (5), and lambdas (λ), defined as the ratio of the coefficient on income for the self-employed divided by the coefficient for the wage worker [col. (1a), (2a), (3a), (4a) and (5a)] for wealth quintiles based on the real estate measure of the tax authority. The sample consists of applicants, whose approved loan amount is lower than the amount requested (constrained sample). We drop observations whose real estate values appear too frequently in the data to be representative of the continuous formula the tax authority uses to calculate the scale. 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. *WageWorker* and *SE* are dummy variables that indicate the applicant's employment type. The vector of control variables includes the standard error of *Log(income)* over the prior 5 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 zip code of the applicant's residence (*RealEstateWealth*). All specifications include 50 credit-grade buckets as risk fixed effects, industry interacted with self-employment fixed effects, branch fixed effects, and year fixed effects. Model (2) includes additionally wealth and income controls. The last row reports 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.

	Quintile 1		Quintile 2		Quintile 3		Quintile 4		Quintile 5	
	(1) Estimate	(1a) λ	(2) Estimate	(2a) λ	(3) Estimate	(3a) λ	(4) Estimate	(4a) λ	(5) Estimate	(5a) λ
Inc.*Wage Worker	0.3484***		0.3418***		0.3742***		0.4321***		0.4431***	
	[0.0488]		[0.0380]		[0.0491]		[0.0638]		[0.0908]	
Income*SE	0.6850***	1.96***	0.7170***	2.09***	0.7168***	1.92***	0.7616***	1.76***	0.7448***	1.68***
	[0.0838]		[0.0862]		[0.1652]		[0.1057]		[0.1505]	
IncomeRisk	-5,914		3,621		-3,578		-1,121		-2,403	
	[7,958]		[5,185]		[5,434]		[4,976]		[4,578]	
SE*IncomeRisk	3,853		-1,671		6,831		7,918		21,844	
	[8,080]		[5,530]		[6,039]		[7,988]		[17,885]	
Lag(IncomeGrowth)	7,560		-30,398***		4,831		-8,486		17,570	
	[8,804]		[8,658]		[10,557]		[13,505]		[20,539]	
SE*Lag(IncomeGrowth)	-781		28,446***		25,293*		10,865		-1,393	
	[9,729]		[10,632]		[15,265]		[20,733]		[29,985]	
Real Estate Wealth	-8.1618		-5.4639**		-17.1447***		-3.1358		2.4906	
	[5.1054]		[2.7156]		[4.5174]		[2.1118]		[1.5323]	
SE*Real Estate Wealth	4.0382		9.0504*		10.0573		2.4578		-1.9669	
	[10.3917]		[4.7931]		[9.6845]		[4.2574]		[3.9890]	
Credit Grade F.E.	Yes		Yes		Yes		Yes		Yes	
Industry*SE F.E.	Yes		Yes		Yes		Yes		Yes	
Branch F.E.	Yes		Yes		Yes		Yes		Yes	
Year F.E.	Yes		Yes		Yes		Yes		Yes	
Adj. R^2	0.175		0.164		0.108		0.142		0.100	
Tax-Evasion Rate	48.98%		52.15%		47.92%		43.18%		40.48%	

49

Table VIII: Tax-Evasion Estimation by Industry (Constrained Sample)

The table presents estimated coefficients [col. (1), (2), and (3)] from equation (5). Columns (1a), (2a), and (3a) present lambdas (λ), defined as the ratio of the coefficient on income for the self-employed divided by the coefficient for the wage worker. Columns (1b), (2b), and (3b) report the estimated tax-evaded income. The sample consists of applicants whose approved loan amount 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. *WageWorker* and *SE* are dummy variables that indicate the employment type of the applicant. The vector of control variables includes 13 industry dummies, the standard error of $Log(income)$ over the prior 5 years by tax cell (*IncomeRisk*), the per-capita annual income growth by tax cell (*Lag(IncomeGrowth)*), the median presumed real estate value of the zip code of the applicant's residence (*RealEstateWealth*). Models (1) and (2) are OLS, whereas model (3) presents results of a quantile regression. All specifications include 50 credit-grade buckets as risk fixed effects, branch fixed effects, and year fixed effects. Model (1) includes industry fixed effects and models (2) and (3) include industry fixed effects crossed with self-employment status. The two last rows report the aggregate lambda and the tax-evasion rate. 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.

	Dependent Variable: Credit Capacity = Outstanding Debt + Approved Loan								
	OLS			OLS			Quintile		
	(1)	(1.a)	(1.b) Tax-Evaded Income	(2)	(2.a)	(2.b) Tax-Evaded Income	(3)	(3.a)	(3.b) Tax-Evaded Income
	Estimate	λ	Income	Estimate	λ	Income	Estimate	λ	Income
Income*WageWorker	0.308*** [0.0487]			0.313*** [0.0529]			0.366*** [0.00399]		
*Accounting & Finance	0.526*** [0.138]	1.71*	14,477	0.600*** [0.233]	1.92	18,732	0.670*** [0.0279]	1.83*	16,971
*Bus.Services & Trade	0.627*** [0.0656]	2.04***	17,654	0.690*** [0.0821]	2.20***	20,508	0.620*** [0.00940]	1.69***	11,833
*Constr. & Transport	0.571*** [0.0600]	1.85***	11,555	0.649*** [0.0959]	2.07***	14,508	0.705*** [0.0202]	1.93***	12,522
*Education	0.781*** [0.105]	2.54***	24,689	1.028*** [0.228]	3.29***	36,687	0.821*** [0.0608]	2.24***	19,971
*Engineering & Science	0.756*** [0.129]	2.46***	28,306	0.996*** [0.212]	3.18***	42,477	0.703*** [0.0198]	1.92***	17,913
*Farming	0.182 [0.125]	not sig	–	0.0910 [0.0799]	not sig	–	0.552*** [0.0103]	1.51*	6,121
*Law	0.716*** [0.116]	2.33***	29,415	0.871*** [0.119]	2.79***	39,557	0.762*** [0.0193]	2.08***	23,965
*Lodging & Tourism	0.496*** [0.186]	1.61	9,952	0.866*** [0.178]	2.77***	28,700	0.660*** [0.0240]	1.80**	13,059
*Manufacturing	0.403*** [0.0884]	1.31	5,382	0.393*** [0.112]	1.26	4,457	0.617*** [0.0123]	1.69	11,975
*Media & Entert.	0.587*** [0.158]	1.91*	15,039	0.904*** [0.155]	2.89***	31,290	0.691*** [0.0566]	1.89**	14,718
*Medicine	0.683*** [0.129]	2.22***	29,346	0.811*** [0.233]	2.59**	38,275	0.826*** [0.0142]	2.25***	30,021
*Personal Services	0.343*** [0.111]	1.11	1,661	0.470*** [0.176]	1.50	7,329	0.620*** [0.0714]	1.69	10,131
*Retail	0.468*** [0.0451]	1.52**	10,342	0.480*** [0.0630]	1.53**	10,635	0.634*** [0.0103]	1.73**	14,588
Self-Employed (SE)	-1,213 [1,061]			absorbed			absorbed		
IncomeRisk				764.3 [1,910]			288.0 [791.7]		
SE*IncomeRisk				3,274 [2,258]			-143.7 [1,149]		
Lag(Income Growth)				5,840 [9,004]			-5,526** [2,265]		
SE*Lag(Income Growth)				3,418 [9,583]			6,656** [2,970]		
Real Estate Wealth				0.987*** [0.310]			0.174* [0.104]		
SE*Real Estate Wealth				-0.183 [0.544]			0.0831 [0.225]		
Industry F.E.	Yes			No			No		
Industry*SE F.E.	No			Yes			Yes		
Branch F.E.	Yes			Yes			No		
Year F.E.	Yes			Yes			Yes		
R^2	0.118			0.122			0.159		
Aggregate multiple (λ)		1.76***			1.95***			1.72***	
Tax-Evasion Rate		43.18%			48.71%			41.80%	

(Income*SE label appears on the left spanning the industry rows.)

Table IX: Tax-Evasion Estimates with Flexible (by Industry group) Sensitivities of Credit to Income

The table presents estimated coefficients [col. (1), (2), (3), (4), and (5)] from equation (5). Columns (1a), (2a), (3a), (4a), and (5a) present lambdas (λ), defined as the ratio of the coefficient on income for the self-employed divided by the coefficient for the wage worker. Models (1) and (2) refer to the constrained sample that consists of applicants, whose approved loan amount is lower than the amount requested, and use as a dependent variable the credit capacity. Models (3)-(5) examine the mortgage sample with the monthly mortgage payments as the dependent variable. *Income* refers to the after-tax reported income verified by the tax return. *WageWorker* 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 5 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 zip code of the applicant's residence (*RealEstateWealth*). Models (1) and (3) present the baseline specification. Models (2) and (4) additionally include wealth and income controls. Model (5) repeats the analysis of model (4), using as a 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, industry interacted with self-employment 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. Heteroscedasticity robust standard errors are presented in brackets. Significance of lambdas refers to statistical difference from unity. Significance at the 1%, 5%, and 10% levels is indicated by ***, **, and *, respectively.

	Constrained by Industry				Mortgages by Industry				Mortgages by Industry (Large Firms)	
	(1)	(1a)	(2)	(2a)	(3)	(3a)	(4)	(4a)	(5)	(5a)
	Estimate	λ	Estimate	λ	Estimate	λ	Estimate	λ	Estimate	λ
Income*Wage Worker Interacted with:										
*Construction, Engineering & Manufacturing	0.386***		0.385***		0.0691***		0.0661**		0.0624**	
	[0.133]		[0.136]		[0.0262]		[0.0260]		[0.0301]	
*Lodging, Tourism and Retail	0.336***		0.339***		0.0272***		0.0256***		0.0256***	
	[0.0578]		[0.0607]		[0.00947]		[0.00913]		[0.00924]	
*Professional Services	0.365***		0.363***		0.0354**		0.0332**		0.0319**	
	[0.0562]		[0.0590]		[0.0171]		[0.0164]		[0.0162]	
Income*Self-Employed Interacted with:										
*Construction, Engineering & Manufacturing	0.508***	1.32	0.554***	1.44	0.0872***	1.26*	0.0867***	1.31*	0.0867***	1.39*
	[0.0966]		[0.107]		[0.0322]		[0.0322]		[0.0323]	
*Lodging, Tourism and Retail	0.480***	1.43*	0.662***	1.95***	0.0551***	2.02*	0.0548***	2.13*	0.0547***	2.14*
	[0.130]		[0.0681]		[0.0123]		[0.0127]		[0.0127]	
*Professional Services	0.641***	1.75***	0.698***	1.92***	0.0790***	2.23*	0.0753***	2.26*	0.0755***	2.36**
	[0.0808]		[0.0826]		[0.0148]		[0.0149]		[0.0150]	
Self-Employed (SE)	533.2		-103.3		182.8***		167.6***		163.4***	
	[1,786]		[1,686]		[25.73]		[44.49]		[45.65]	
IncomeRisk			2,287				116.5		126.1	
			[3,775]				[116.3]		[132.1]	
SE*IncomeRisk			-164.1				-52.29		-61.67	
			[4,332]				[145.0]		[157.6]	
Lag(Inc.Growth)			1,024				1,096***		1,115***	
			[13,556]				[280.9]		[290.7]	
SE*Lag(Inc.Growth)			9,374				-1,080***		-1,089***	
			[14,768]				[329.8]		[338.4]	
Real Est.Wealth			0.796*				0.0897***		0.0899***	
			[0.425]				[0.0175]		[0.0187]	
SE*Real Est.Wealth			-1.059				0.0178		0.0188	
			[0.870]				[0.0340]		[0.0342]	
Industry*SE F.E.	Yes		Yes		Yes		Yes		Yes	
Branch F.E.	Yes		Yes		Yes		Yes		Yes	
Year F.E.	Yes		Yes		Yes		Yes		Yes	
Adj. R^2	0.112		0.114		0.123		0.129		0.126	
Aggregate multiple (λ)		1.51***		1.74***		1.93**		1.96**		2.05**
Tax-Evasion Rate		33.95%		42.53%		48.18%		49.08%		51.21%

51

Table X: Tax-Evasion Estimation Summarized with Education and Legislation

The table presents the industries sorted on tax evasion in euros (in descending order) and information on anti-tax-evasion legislation targeting and the source of professional rights. The mean tax evasion is calculated as the average of the estimates from specifications (1)-(3) of Table VIII. "*Anti-Tax Evasion Legislation*" indicates whether the industry was targeted by the anti-tax evasion legislative bill proposed in 2010. "*Degree Requirement*" indicates whether a degree is required to follow the occupation. The last column shows whether the job permit is issued by the government or a union.

	(1) Mean Tax Evasion (€)	(2) λ	(3) Anti-Tax Evasion Legislation	(4) Degree Requirement	(5) Job Permit Issued by
Medicine	32,548	2.35	Yes	Yes	Union
Law	30,979	2.40	Yes	Yes	Union
Engineering & Science	29,565	2.52	Yes	Yes	Union
Education	27,116	2.69		Yes	Government
Media & Entertainment	20,349	2.23			
Lodging & Tourism	17,237	2.06			
Accounting & Finance	16,727	1.82	Yes	Yes	
Business Services & Trade	16,665	1.98	Mixed	Mixed	Government
Construction & Transport	12,862	1.95		Mixed	
Retail	11,855	1.60			Government
Manufacturing	7,271	1.42			
Personal Services	6,374	1.44		Mixed	Union/Gov.
Farming	6,121	1.51			

Table XI: Paper-Trail Intensity

The table presents the results of the survey on the paper-trail intensity in various industries. Industries are sorted on average tax evasion in euros (in descending order), calculated as the average of the estimates from specifications (2) and (3) of Table VIII. Columns (3) and (4) report the results from Question 1: *"What percentage of expenses goes towards buying intermediate goods or services as inputs?"*. Columns (5) and (6) report the results from Question 2: *"How often the output is used by clients as intermediate good?"*. Columns (3) and (5) report the demeaned paper-trail score, along with significance at the 1%, 5%, and 10% levels, of the difference of the responses from the median (3), indicated by ***, **, and *, respectively. The higher the scale the higher the paper-trail intensity. Panel B classifies paper-trail scale into three levels: Low (L), Medium (M), and High (H). The correlation of the paper-trail scale with the tax-evasion amount and lambda is provided.

	Panel A: Paper Trail Statistics					
			High Question Scale = High Paper Trail			
			Question 1: Use of Intermediate Goods as Inputs		**Question 2: Output is Intermediate Good**	
	(1)	(2)	(3)	(4)	(5)	(6)
	Tax Evasion	λ	Scale	Intensity	Scale	Intensity
Medicine	32,548	2.35	0.382*	M	-0.412**	L
Law	30,979	2.40	-1.049***	L	-0.169**	M
Engineering & Science	29,565	2.52	0.048	M	0.555*	H
Education	27,116	2.69	-0.781***	L	-0.474**	L
Media & Entertainment	20,349	2.23	-0.837***	L	-0.278**	L
Lodging & Tourism	17,237	2.06	0.540**	H	-0.263**	M
Accounting & Fin. Services	16,727	1.82	-0.893***	L	-1.105***	L
Business Services & Trade	16,665	1.98	-0.948***	L	0.105	M
Construction & Transport	12,862	1.95	0.698***	H	0.340**	H
Retail	11,855	1.60	0.804***	H	0.316	M
Manufacturing	7,271	1.42	1.382***	H	0.778**	H
Personal Services	6,374	1.44	0.257	M	-0.368***	L
Farming	6,121	1.51	0.593***	H	0.684*	H
Corr. with Tax Evasion			-0.521		-0.345	
Corr. with λ			-0.568		-0.311	

Panel B: Average of Tax Evasion by Input Ranking				
	Input-Paper-Trail Ranking		**Output-Paper-Trail Ranking**	
Tax Evasion In:	Euros	Multiple	Euros	Multiple
Low Paper Trail	22,367	2.22	20,623	2.11
Medium Paper Trail	22,829	2.10	19,184	2.01
High Paper Trail	11,069	1.70	16,183	1.99

Table XII: Enforcements

The unit of observation in the table is the tax office district. Tax-enforcement data are from the year 2011 and observed at a daily frequency. Cases are assigned by the central tax office, based on an automatic formula. Panel A presents summary statistics. Closed cases are the sum of cases closed by the tax office during the year. Amount assessed is the total value (in euros) of assessments for the year. One tax office has an assessment sum 10 times larger than the other. This observation is winsorized to the second-highest value. Percent self-employed is the percent of tax filers who are either merchants or self-employed in the tax files. The wealth proxy is a variable constructed based on the presumed value of real estate, in the jurisdiction of the tax office; it provides a relative ranking and has no cardinal meaning. In Panel B, the dependent variable in columns (1)-(6) is the number of closed cases. In Panel C, in columns (7)-(9) the dependent variable is the average assessment per closed case and in columns (10)-(12) is total assessments by tax office. All columns present OLS results of the elasticities (logs) of the the dependent variables, except columns (4)-(6) which report results from a Poisson model. In columns (3), (6), (9), and (12) we use the modified Gram-Schmidt procedure of Golub & Van Loan (1996) to orthogonalize the wealth and self-employment variables, due to high correlation. Robust standard errors are in brackets. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Summary Statistics

	Mean	St.Deviation	Skewness	Min.	Median	Max.	Obs.
Closed Cases	933.0	859.3	2.16	1	742	6,267	235
Assessments	2,124,272	4,880,819	11.5	40	1,268,537	68,700,000	226
Assessed/Close	3,098	20,604	14.9	12.0	1,517	311,044	226
Taxfilers	23,851	18,305	1.03	676	19,083	90,532	235
% Self Employed	0.192	0.060	3.73	0.11	0.181	0.651	235
Wealth Proxy	10.68	4.64	1.08	4.5	8.94	27.48	235

Panel B: Analysis of Enforcements, Cases Closed

Dep. Variable:	Ln Closes			Closes - Possion Model		
	(1)	(2)	(3)	(4)	(5)	(6)
Taxfilers	0.990***	1.015***	1.015**	Taxfilers is absorbed		
	[0.0645]	[0.0684]	[0.0684]	as exposure variable.		
(Ortho-)%Self-Employed	1.049***	-6.021*	0.299***	4.384***	-0.0366	0.236***
	[0.233]	[3.370]	[0.0411]	[0.832]	[1.840]	[0.0319]
(Ortho-)Wealth	0.227*	-1.586*	0.0729*	0.00911	-0.0361**	0.0373
	[0.118]	[0.877]	[0.0432]	[0.00680]	[0.0177]	[0.0285]
(Ortho-)%SE*Wealth		0.609**	0.0717**		0.231**	0.0831**
		[0.292]	[0.0344]		[0.0903]	[0.0325]
Constant	-8.935***	11.84	-3.504***	-4.202***	-3.360***	-3.253***
	[0.959]	[9.781]	[0.687]	[0.164]	[0.329]	[0.0384]
R^2/Pseudo R^2	0.741	0.744	0.744	0.271	0.293	0.293

Panel C: Analysis of Enforcements, Assessments

Dep. Variable:	Ln (Assessment / Close)			Ln Assessments		
	(7)	(8)	(9)	(10)	(11)	(12)
Taxfilers	0.290***	0.261***	0.261**	0.224*	0.176	0.176
	[0.0647]	[0.0678]	[0.0678]	[0.134]	[0.135]	[0.135]
Closes				1.065***	1.082***	1.082***
				[0.154]	[0.151]	[0.151]
(Ortho-)%Self-Employed	0.996***	8.462	0.256***	0.928***	9.000	0.232***
	[0.262]	[5.988]	[0.0560]	[0.265]	[6.097]	[0.0628]
(Ortho)-Wealth	0.0451	1.958	0.0255	0.0298	2.098	0.0193
	[0.103]	[1.507]	[0.0387]	[0.118]	[1.529]	[0.0430]
(Ortho-)%SE*Wealth		-0.642	-0.0757		-0.696	-0.082
		[0.505]	[0.0595]		[0.515]	[0.0607]
Constant	1.024	-20.9	4.728***	1.621	-21.97	5.031***
	[1.194]	[17.64]	[0.692]	[1.684]	[17.89]	[0.628]
R^2/Pseudo R^2	0.203	0.211	0.211	0.801	0.803	0.803

Table XIII: Tax-Evasion and Parliamentarians' Occupational Backgrounds

The table reports the distribution of the occupational backgrounds of Parliament members in Greece (2009). The occupational background information was hand collected from the members' official websites. Industries are sorted on average tax evasion in euros (in descending order), calculated as the average of the estimates from specifications (1)-(3) of Table VIII. Column 4 presents the percentage of parliament members in each occupation, and column 5 presents the cumulative density. We exclude lawyers as the natural default profession for politicians.

	(1)	(2)	(3)	(4)	(5)
	Tax Evasion	λ	N	Density	Cum. Density
Medicine	32,548	2.35	40	0.174	0.174
Engineering & Science	29,565	2.52	43	0.187	0.361
Education	27,116	2.69	28	0.122	0.483
Media & Entertainment	20,349	2.23	31	0.135	0.617
Lodging & Tourism	17,237	2.06	3	0.013	0.630
Accounting & Fin. Services	16,727	1.82	40	0.174	0.804
Business Services	16,665	1.98	20	0.087	0.891
Construction	12,862	1.95	7	0.030	0.922
Retail	11,855	1.60	3	0.013	0.935
Manufacturing	7,271	1.42	1	0.004	0.939
Personal Services	6,374	1.44	5	0.022	0.961
Farming	6,121	1.51	8	0.035	0.996
Other	-	-	1	0.004	1.000
			230	1.000	
Law (excluded)	30,979	2.40	70		

Appendix: Reproducibility Exercise in US Household Debt Data

In this appendix, we implement an illustrative exercise using data from the U.S. Survey of Consumer Finance (SCF) to demonstrate that our method is reproducible. In the United States, formal lenders do not provide credit on hidden income; however, people underreport income in surveys. Using consumption-to-income relationships, Hurst, Li & Pugsley (2014) find that self-employed survey participants underreport income by 25%. We use Hurst, Li & Pugsley (2014) as a benchmark exercise to examine whether we can reproduce hidden-income using a different dataset, the Survey of Consumer Finances (SCF), by applying our debt method.

We apply a simple version of our method, a debt-burden rule whereby lenders will not offer credit to an applicant whose monthly debt payments exceed a threshold percent of monthly income. The threshold that lenders apply varies according to whether the household is a homeowner or renter. Thus, we differentiate households by home ownership and apply two different ratios. For homeowners, we employ the so-called *Ability-to-Pay Rule* of the US Consumer Finance Protection Bureau, which states that for a mortgage to be deemed a "qualified mortgage", post-loan debt-servicing obligations cannot be larger than 43% of income.[30] For renters, a typical threshold rule is that non-housing debt servicing obligations should not be more than 20% of income.[31] We apply these rules to uncover hidden income as a simple inversion; we divide debt payments reported in the SCF by the appropriate rule (43% or 20%) to generate a debt-burden estimate of true income. By definition, if households are not maxed out on debt, the rule will give a conservative estimate of true income, and, thus, will understate hidden income.[32]

Panel A of appendix Table A.8 presents debt payment and income summary statistics (weighted using the SCF survey weights) for households in the United States in the 2001, 2004, and 2007 SCFs. Monthly payments comprise 26.1% of monthly income for the self-employed. Panel B reports our reproducibility exercise findings. The main finding of this exercise is reported in Panel B, that

---

[30]See http://files.consumerfinance.gov/f/201312_cfpb_mortgagerules.pdf.
This is often called the "back-end ratio" in mortgage parlance. The standard used to be 36%; however, we use the new, higher ratio to be conservative.

[31]In the US, this ratio can be called the "gap ratio", measuring the amount of that debt-servicing gap that is not mortgage related for potential home-buyers. The gap ratio standard is 15% (*Los Angeles Times*, January 27, 2002). *Yahoo Finance* ("7 Signs Your Credit Card Debt Is About to Implode"; January 11, 2013) advocates having revolving credit payments less than 15%. We use a conservative 20% (Credit Sesame).

[32]If the income resulting from this estimation procedure is less than reported income, we assume the household is not underreporting income. This also may be conservative.

the underreporting rate in the SCF is 19.1% for the self-employed, with a confidence band around this estimate ranging from 15% to 23%. This magnitude is quite similar to the 25% tax evasion rate of Hurst, Li & Pugsley (2014).

Appendix Tables

Table A.1: Semi-Formality and Credit Depth Info

The table presents the percentage of firms underreporting sales and the average of reported sales are from the 2002-2005 Enterprise Surveys and measures of semi-formality across 33 countries ranked on the size of their informal sector according to Djankov et al. (2010). Credit info refers to an index of credit depth of information index (World Bank (2005)) that measures rules affecting the scope, accessibility, and quality of credit information available through public or private credit registries (0=low to 6=high).

	Firms Underreporting	Average sales reported	Size of Inf. Sector	Credit Info
Brazil	73.58%	67.35%	35.83%	5
Dom. Republic	70.33%	50.73%	34.17%	5
Turkey	62.72%	63.66%	36.25%	5
Panama	50.36%	62.88%	27.50%	6
Mexico	48.77%	76.26%	36.67%	6
Georgia	47.90%	76.52%	32.50%	0
India	47.87%	73.07%	29.17%	2
Argentina	42.56%	82.53%	34.58%	6
Russia	37.35%	83.38%	37.50%	0
Uruguay	36.96%	85.34%	28.75%	5
Ecuador	35.48%	75.95%	31.67%	4
Greece	**34.86%**	**88.95%**	**27.92%**	**4**
Bulgaria	31.82%	86.51%	31.25%	4
Colombia	30.43%	82.93%	28.33%	5
Czech Republic	29.73%	88.24%	19.17%	5
Lithuania	29.60%	87.50%	16.67%	6
Hungary	28.91%	88.65%	27.50%	5
Poland	28.67%	90.21%	25.83%	4
Latvia	25.57%	90.45%	24.17%	3
Slovenia	25.56%	88.22%	17.08%	3
Portugal	25.10%	91.83%	21.67%	5
Croatia	24.86%	90.37%	30.42%	0
Ukraine	24.60%	87.67%	37.08%	0
Romania	22.99%	91.49%	25.42%	5
South Africa	21.96%	90.40%	29.58%	5
Kazakhstan	21.20%	90.46%	33.33%	0
Slovak Republic	20.68%	92%	20.83%	3
Jamaica	20.55%	88.20%	37.50%	0
Armenia	15.98%	94.03%	31.25%	3
Germany	15.86%	94.31%	17.50%	6
Chile	15.12%	91.81%	15.00%	5
Spain	12.67%	96.33%	21.25%	5
Ireland	11.02%	96.16%	16.25%	5

Table A.2:Variables, Definitions

The table presents a comprehensive list of variables used in analysis with detailed definitions. The list includes variables from loan applications and variables constructed from the tax authority dataset.

A.Income & Other Debt (in euros)

Reported Income: After-tax (disposable)income of the applicant as shown in the tax-return form in the year of application.
Credit Capacity: Total debt outstanding immediately following the loan-application decision.
Debt Outstanding: Total debt outstanding at the time of application.
Mean Monthly Payments: Monthly payments of approved mortgage.

B. Occupation Information

Type of Employment (WW or SE): Employment classification (self-employed, wage-worker). We exclude retirees, students, unemployed and other.
Occupation: Description of occupation according to internal detailed classification.
Employer: Name of the company that employs the applicant (available only for mortgages).
Years in Job: Number of years that the applicant is working under the same employer (wage workers) or is employed in the same occupation (self-employed).

C. Credit History

Agree to Credit Check: Takes the value 1 if the applicant agrees to credit check, and 0 otherwise.
Years in Address: Number of years that the primary applicant resides at the same address.
Years of Coop. with Bank: Number of years of the existing relationship of the applicant with the bank.
Existing Bank Customer: Takes the value 1 if the applicant is a current customer of the bank, and 0 otherwise.
Delinquent Existing Bank Customer: Takes the value 1 if the applicant is a customer of the bank but has a delayed payment in the last 6 months, and 0 otherwise.
Other Banks: Indicator of the existing relationship of the applicant with other banks.
Previous Customer: Takes the value 1 if the applicant was a customer of the bank in the past, and 0 otherwise.
Previous Homeowner: Takes the value 1 if the mortgage applicant is a homeowner, and 0 otherwise.
Deposits: Total amount of deposits of the applicant in the bank. It includes amounts in any accounts the applicant holds with the bank (checking, savings, time-deposits, investment accounts).
Home-ownership: Takes the value 1 if the applicant is a homeowner, and 0 otherwise.

D.Demographics

Marital Status: Takes the value 1 if the applicant is married, and 0 otherwise.
Number of Dependants: The number of the applicant's dependents at the time of the application.
Age: Age of applicant at the time of the application.
Address: The address of the applicant's primary residence.

E.Loan Characteristics

Type of Application: Indicator of new loan application or restructuring.
Product Description/Code: The description and the code of the product.
Purpose of Loan: The purpose of the loan according to internal classification (MA only).
Requested Amount: The requested amount of the loan.
Interest Rate: The interest rate of the loan and its type (fixed or floating), spread.
Maturity: The maturity of the loan in months.
Duration: Duration of loan (in months).
Has Collateral with Bank: The applicant has posted collateral with the bank.
Loan To Value: Ratio of the approved mortgage relative to the value of the real estate property. Available only for mortgages and calculated only for approved applications.
Solo Application: Indicator that takes the value 1 if the application has no co-signers, and 0 otherwise.

F.Soft Information

Real Estate Value, Zip Level: The median "presumed real estate value of the region where the primary applicant resides (specified by zip code). "Presumed real estate values are periodically published by tax authorities and used to determine real estate taxes. The values here come from the most recent release published in 2007.
Car Value, Mean Zip Level: The mean value of automobiles by zip code, as calculated from the car-loan dataset of the bank.
Car Loan-to-Value by Zip: Average loan-to-values of new cars by zip-code.
Tax Cell: Each tax cell is defined at the zip-code level (1,450 zip-codes) broken down into national income deciles and into the four occupations classes of the tax authority classification.
Lag Income Growth by Tax Cell: Per-capital annual income growth of the prior year at the level of the zip code crossed with the four occupation-levels and the 10 income deciles.
Income Risk by Tax Cell: Measure of the variability of income growth, defined as the standard deviation of income divided by the square root of the number of people in the zip-income decile-occupation cell.

Table A.3: Recreation of Credit Grade

Three OLS regressions are reported. The dependent variable is the credit grade of the bank. We do not report coefficients for proprietary reasons, but instead we report the partial-R^2 and p-values. The point is to show the overall variation explained by the different demographics. The sample is all consumer loan-product applications (not including homes and vehicles), both accepted and rejected, from 2003-2009.

	(1)		(2)		(3)	
	Partial-R^2	P-value	Partial-R^2	P-value	Partial-R^2	P-value
Income	0.0044	0.00	0.0028	0.00	0.0028	0.00
Debt Outstanding Other	0.0005	0.00	0.0001	0.00	0.0001	0.00
Debt Outstanding at Bank	0.0019	0.00	0.0008	0.00	0.0008	0.00
Solo Application	0.0006	0.00	0.0002	0.00	0.0002	0.00
Grace period Requested	0.0007	0.00	0.0001	0.00	0.0001	0.00
Years in Job	0.0196	0.00	0.0096	0.00	0.0089	0.00
Requested Amount	0.0009	0.00	0.0003	0.00	0.0002	0.00
Requested Duration Level 1	0.0000	0.34	0.0000	0.44	0.0000	0.35
Requested Duration Level 2	0.0039	0.00	0.0030	0.00	0.0030	0.00
Requested Duration Level 3	0.0008	0.00	0.0006	0.00	0.0006	0.00
Requested Duration Level 4	0.0047	0.00	0.0036	0.00	0.0036	0.00
Requested Duration Level 5	0.0010	0.00	0.0007	0.00	0.0007	0.00
Requested Duration Level 6	0.0059	0.00	0.0047	0.00	0.0047	0.00
Requested Duration Level 7	0.0013	0.00	0.0010	0.00	0.0010	0.00
Requested Duration Level 8	0.0063	0.00	0.0056	0.00	0.0056	0.00
Requested Duration Level 9	0.0045	0.00	0.0039	0.00	0.0039	0.00
Requested Duration Level 10	0.0082	0.00	0.0071	0.00	0.0072	0.00
Requested Duration Level 11	0.0013	0.00	0.0007	0.00	0.0007	0.00
Bank Relationship Years	0.0117	0.00	0.0081	0.00	0.0082	0.00
Relationship with Bank	0.0218	0.00	0.0252	0.00	0.0254	0.00
Age	0.0157	0.00	0.0219	0.00	0.0224	0.00
Gender	0.0009	0.00	0.0004	0.00	0.0003	0.00
Single	0.0020	0.00	0.0018	0.00	0.0018	0.00
Married	0.0067	0.00	0.0072	0.00	0.0073	0.00
Divorced	0.0000	0.00	0.0000	0.06	0.0000	0.11
Widowed	0.0002	0.00	0.0003	0.00	0.0003	0.00
Renter	0.0419	0.00	0.0384	0.00	0.0377	0.00
Guest	0.0125	0.00	0.0120	0.00	0.0122	0.00
Other Housing	0.0158	0.00	0.0137	0.00	0.0137	0.00
Num. of dependents 1	0.0000	0.00	0.0000	0.64	0.0000	0.80
Num. of dependents 2	0.0000	0.62	0.0000	0.01	0.0000	0.00
Num. of dependents 3	0.0000	0.06	0.0000	0.30	0.0000	0.32
Num. of dependents 4	0.0000	0.00	0.0000	0.00	0.0000	0.00
Num. of dependents 5	0.0000	0.00	0.0000	0.00	0.0000	0.00
Year 2004	0.0000	0.00	0.0002	0.00	0.0001	0.00
Year 2005	0.0002	0.00	0.0005	0.00	0.0005	0.00
Year 2006	0.0002	0.00	0.0011	0.00	0.0011	0.00
Year 2007	0.0006	0.00	0.0018	0.00	0.0018	0.00
Year 2008	0.0044	0.00	0.0074	0.00	0.0074	0.00
Year 2009	0.0439	0.00	0.0481	0.00	0.0477	0.00
App. Month Feb	0.0001	0.00	0.0000	0.00	0.0000	0.00
App. Month Mar	0.0000	0.00	0.0000	0.05	0.0000	0.04
App. Month Apr	0.0000	0.01	0.0000	0.45	0.0000	0.56
App. Month May	0.0000	0.00	0.0000	0.45	0.0000	0.62
App. Month Jun	0.0000	0.02	0.0000	0.16	0.0000	0.23
App. Month Jul	0.0000	0.14	0.0000	0.35	0.0000	0.44
App. Month Aug	0.0000	0.16	0.0000	0.12	0.0000	0.19
App. Month Sep	0.0000	0.10	0.0000	0.00	0.0000	0.00
App. Month Oct	0.0004	0.00	0.0004	0.00	0.0004	0.00
App. Month Nov	0.0013	0.00	0.0012	0.00	0.0012	0.00
App. Month Dec	0.0020	0.00	0.0020	0.00	0.0020	0.00

	(1)		(2)		(3)	
	Partial-R^2	P-value	Partial-R^2	P-value	Partial-R^2	P-value
Accounting & Fin. Services	0.0123	0.00	0.0052	0.00	0.0036	0.00
Business Services	0.0000	0.00	0.0000	0.92	0.0000	0.00
Construction			0.0024	0.00	0.0008	0.00
Education		0.565202	0.0256	0.00	0.0242	0.00
Engineering & Science			0.0112	0.00	0.0067	0.00
Farming			0.0019	0.00	0.0014	0.00
Law			0.0050	0.00	0.0025	0.00
Lodging & Tourism			0.0195	0.00	0.0198	0.00
Manufacturing			0.0050	0.00	0.0041	0.00
Media & Entertainment			0.0009	0.00	0.0007	0.00
Medicine			0.0116	0.00	0.0096	0.00
Personal Services			0.0046	0.00	0.0042	0.00
Retail			0.0394	0.00	0.0392	0.00
Self-employed (SE)			0.0007	0.00	0.0000	0.00
SE*Accounting & Fin. Services					0.0001	0.00
SE*Business Services					0.0001	0.00
SE*Construction					0.0009	0.00
SE*Education					0.0000	0.30
SE*Engineering & Science					0.0000	0.01
SE*Farming					0.0000	0.02
SE*Law					0.0000	0.00
SE*Logding & Tourism					0.0017	0.00
SE*Manufacturing					0.0000	0.03
SE*Media & Entertainment					0.0000	0.01
SE*Medicine					0.0001	0.00
SE*Personal Services					0.0001	0.00
SE*Retail					0.0012	0.00
Total R^2	0.5652		0.6623		0.6629	

Table A.4: Summary Statistics of Samples

The table reports statistics (means, medians, and standard deviations) for variables of the three samples used in our analysis. Columns 1-3 report summary statistics for term loans, columns 4-6 for mortgages term loans, and columns 7-9 for overdraft term loans.

Sample:	Term Loans (Non-homeowners)			Mortgages (Homeowners)			Overdraft Facilities (New Checking Customers)		
	(1) Mean	(2) Median	(3) St.Dev.	(4) Mean	(5) Median	(6) St.Dev.	(7) Mean	(8) Median	(9) St.Dev.
Reported Income	13,209	11,500	13,609	17,600	14,585	13,340	21,780	18,706	14,834
Total Credit	11,448	6,166	22,241	112,739	90,000	116,162	12,472	5,000	34,212
Prior Debt	5,626	0	20,780	18,814	0	78,694	7,304	0	34,372
Approved Loan	5,822	4,308	5,066	93,925	80,000	77,561	5,168	5,000	3,306
Credit Grade	89.3	88.0	10.0	136.3	136.0	22.0	90.1	91.0	7.3
Monthly Income	1,101	958	1,134	1,467	1,215	1,112	1,815	1,559	1,236
Mon. Debt Paym.	265	163	457	944	561	1,727	360	200	720
Monthly Rate	0.0095	0.0103	0.0016	0.0034	0.0031	0.0010			

Table A.5: Summary Statistics of Constrained Sample

The table reports summary statistics (means, medians, and standard deviations) of variables for the constrained sample, defined as the term-loan applicants, whose approved loan amount is lower than the amount requested. Means and standard deviations are reported for self-employed and wage workers, separately. Means are also provided for the 13 industries used in the analysis.

		Reported Income	Credit Capacity	Credit Grade	Self Employed	IncomeRisk	Lag Growth	Real Estate Wealth (Index)
All	Mean	17,253	12,520	92.4	0.22	0.0256	0.0085	1,165
	Median	14,400	7,900	92.0	0.00	0.0096	0.0054	1,021
	St.Deviation	14,095	27,056	9.1	0.41	0.0772	0.0284	477
Wage Workers	Mean	17,575	12,341	92.7	0.00	0.0154	0.0095	1,195
	St.Deviation	13,277	26,561	9.2	0.00	0.0506	0.0210	484
Self-Employed	Mean	16,083	13,171	91.1	1.00	0.0639	0.0048	1,055
	St.Deviation	16,681	28,771	8.8	0.00	0.1301	0.0463	436
Accounting & Finance	Mean	20,230	18,713	95.9	0.45	0.0305	0.0100	1,258
Bus. Services & Trade	Mean	18,481	13,720	92.9	0.13	0.0213	0.0094	1,231
Constr. & Transport	Mean	14,048	10,391	91.4	0.34	0.0353	0.0063	1,063
Education	Mean	20,003	14,978	101.8	0.06	0.0238	0.0120	1,193
Engineering & Science	Mean	21,580	19,312	99.7	0.43	0.0373	0.0096	1,263
Farming	Mean	11,249	7,383	87.1	0.57	0.0273	0.0078	787
Law & Order	Mean	18,908	13,521	97.6	0.04	0.0180	0.0075	1,104
Lodging & Restaurants	Mean	13,939	11,063	90.4	0.50	0.0580	0.0063	1,075
Manufacturing	Mean	17,745	12,468	91.0	0.32	0.0317	0.0099	974
Media & Entertainment	Mean	20,393	17,356	91.6	0.37	0.0342	0.0090	1,456
Medicine	Mean	19,969	15,709	96.3	0.11	0.0200	0.0099	1,247
Personal Services	Mean	17,474	13,561	96.0	0.19	0.0209	0.0083	1,119
Retail	Mean	15,874	13,974	89.8	0.29	0.0161	0.0080	1,221

Table A.6: True Income Estimated of the Self-employed in the Overdraft Sample

The table presents estimated coefficients [col. (1) and (2)] from equation (5). Columns (1a) and (2a) 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 applicants who have overdraft protection in their checking accounts (overdraft 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. *WageWorker* and *SE* are dummy variables that indicate the applicant's employment type. The vector of control variables includes the standard error of *Log(income)* over the prior 5 years by tax cell (*IncomeRisk*), the per-capita annual income growth by tax cell (*Lag(Inc.Growth)*), and the median presumed real estate value of the zip code of the applicant's residence (*RealEst.Wealth*). All specifications include 50 credit-grade buckets as risk fixed effects, branch fixed effects, and year fixed effects. Model (2) includes additionally wealth and income controls. The last row reports 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.

	(1) Estimate	(1.a) λ	(2) Estimate	(2.a) λ
	Dependent Variable: Credit Capacity = Outstanding Debt + Approved Loan			
Income*Wage Worker	0.1982**		0.1968**	
	[0.0869]		[0.0876]	
Income*SE	0.5872***	2.96*	0.5658***	2.87*
	[0.2137]		[0.2177]	
IncomeRisk			-2,110	
			[4,161]	
SE*IncomeRisk			8,932	
			[30,383]	
Lag(IncomeGrowth)			12,229	
			[20,181]	
SE*Lag(IncomeGrowth)			25,838	
			[96,644]	
Real Estate Wealth			0.1867	
			[0.8379]	
SE*Real Estate Wealth			5.129	
			[10.039]	
Branch F.E.	Yes		Yes	
Year F.E.	Yes		Yes	
Adj. R^2	0.074		0.070	
Tax-Evasion Rate		66.20%		65.10%

Table A.7: Tax-Evasion Estimates by Geography with Flexible (by Industry group) Sensitivities of Credit to Income

The table presents estimated coefficients [col. (1) and (2)] from equation (5) lambdas (λ), defined as the ratio of the coefficient on income for the self-employed divided by the coefficient for the wage worker [col. 1a) and (2a)] for applicants in urban and rural areas. The "Urban" columns include applicants inhabiting in Athens, Thessaloniki and cities with population higher than 100,000 people and the "Rural" columns all other areas. The sample consists of applicants, whose approved loan amount 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. $WageWorker$ 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 5 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 zip code of the applicant's residence ($RealEstateWealth$). All specifications include 50 credit-grade buckets as risk fixed effects, industry interacted with self-employment fixed effects, branch fixed effects, and year fixed effects. 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.

	Urban		Rural	
	(1)	(1a)	(2)	(2a)
	Estimate	λ	Estimate	λ
Income*Wage Worker Interacted with:				
*Construction, Engineering & Manufacturing	0.3537**		0.385***	
	(0.1742)		[0.136]	
*Lodging, Tourism and Retail	0.3613***		0.339***	
	[0.0578]		[0.0607]	
*Professional Services	0.3613***		0.339***	
	[0.0633]		[0.0590]	
Income*Self-Employed Interacted with:				
*Construction, Engineering & Manufacturing	0.5282***	1.49	0.554***	1.40
	[0.1342]		[0.107]	
*Lodging, Tourism and Retail	0.6427***	1.77***	0.662***	2.08***
	[0.0888]		[0.0681]	
*Professional Services	0.7277***	1.89***	0.698***	1.96***
	[0.0803]		[0.0826]	
Self Employed (SE)	-3,237*		-103.3	
	[1,913]		[1,686]	
IncomeRisk	1,491		2,287	
	[4,279]		[3,775]	
SE*IncomeRisk	9,099		-164.1	
	[9,063]		[4,332]	
Lag(Income Growth)	28,362		1,024	
	[20,933]		[13,556]	
SE*Lag(Income Growth)	-14,406		9,374	
	[22,986]		[14,768]	
Real Estate Wealth	0.6221*		0.796*	
	[0.3725]		[0.425]	
SE*Real Estate Wealth	1.4321		-1.059	
	[1.0930]		[0.870]	
Credit Grade F.E.	Yes		Yes	
Industry*SE F.E.	Yes		Yes	
Branch F.E.:	Yes		Yes	
Year F.E.:	Yes		Yes	
Adj. R^2	0.1837		0.1441	

Table A.8: Estimates of Hidden Income in the US Survey of Consumer Finances
The table presents self-employed summary statistics of household income and debt-servicing payments for the United States across the SCF waves 2001, 2004, and 2007, weighted to be nationally representative. The right part of the table presents the underreporting rate in the population, following Hurst et al. (2014) by way of the debt-burden rules. The debt-burden rule threshold is that debt can be no larger than 43% of income for homeowners and 20% of income for renters. Reported in brackets is the standard error from a test that the estimate equals zero underreporting.

	United States: Self-Employed(household level)					
	Monthly Income (1)	Monthly Debt Serv. Payments (2)	Ratio (Payments to Income) (3)	Underreporting Rate (% of Income Under-reported) (4)		
Mean	12,488	1,904	0.261	Estimate	0.191	[0.019]
Median	5,717	1,200		Lower	0.153	
St. Deviation	48,220	3,800		Upper	0.229	
		Number of Observations: 2,053				