Are Tax Lotteries Regressive? Income, Consumption and, Occupational Characteristics of Winners

Panayiotis Nicolaides *

June 16, 2023

Abstract

Tax lotteries can increase tax compliance by providing incentives to consumers to ask for a receipt. Whilst their effectiveness has been studied in economic literature, less is known about their progressivity. Using administrative data from the Greek tax lottery, I analyse the income, consumption and occupational characteristics of winners and I document that high income individuals win more frequently. A 10% increase in taxpayer [spousal] income is associated with a 0.11% [0.6%] increase in the winning probability. Self-employed or business owners record particularly large amounts of transactions and receive a large number of lottery tickets, which indicates either under-reporting of income or the channeling business expenditure through personal bank accounts. They exhibit 18% higher winning probability than other occupational categories. Using Monte Carlo simulations, I explore reforms for tax lottery design and I find that a monthly ticket ceiling per individual improves prize allocation and progressivity. Additional results with economic significance are (i) imperfect income sharing within household (ii) a marginal propensity to consume at 0.18.

JEL Classification: D13, D22, E21, H24, H25, H26, H31, H83

Keywords: Tax Lottery; Electronic Payments; Third-Party Information; Progressivity; Marginal Propensity to Consume.

---

*EU Tax Observatory, Paris School of Economics (p.nicolaides@psemail.eu). I would like to thank the Independent Authority for Public Revenue in Greece for the collaboration in anonymising and providing the data (which were received in October 2018). Christian Traxler provided close guidance in this paper. Seminar and conference participants at the Ministry of Finance in Greece, the Hertie School, Berlin, the 6th MaTax Annual Conference, University of Mannheim and the 7th Annual Tax Administration Research Centre Conference, University of Exeter provided useful comments and suggestions. Field studies in Greece were funded by the Hertie School through the German Academic Exchange Service (DAAD) funds.

1 Introduction

Tax lotteries are schemes designed to increase the compliance of firms with the payment of value-added tax (VAT hereafter). They provide incentives to address VAT's "last-mile" problem: third-party reporting in business-to-business transactions deters tax evasion, since one firm's input serves as the other's output (Pomeranz, 2015), but in business-to-consumer transactions no incentive exists in asking and keeping a receipt of purchase. Unrecorded transactions result in higher tax evasion and lower public revenue. The monetary incentives are targeted to the final consumer: assigning to each receipt the expectation of winning a prize. The additional third-party information should then facilitate enforcement (Kleven *et al.*, 2011). While the effectiveness of tax lotteries has been the subject of study in economic literature – notably in Naritomi (2019) for the Brazilian lottery and in Nicolaides (2023) for the Greek lottery – less is known about the winners' characteristics. Who wins the lottery determines the policy's regressivity or progressivity.

Using administrative data from the Greek tax lottery, this paper provides evidence on (i) the income, consumption and occupational characteristics of winners (ii) a quantification of the extent to which these characteristics affect their winning probabilities (iii) simulations on a ticket ceiling reform as a way to improve progressivity. The data allow for the reconstruction of a representative taxpayer population against which the population of tax lottery winners can be compared. They contain information on income, electronic consumption and occupation indicators for 68,897 tax units in the Greek population; 18,897 winning units in the Greek tax lottery during 19 monthly draws in 2017 and 2018, and a random sample of 50,000 non-winning tax units.

The comparison reveals an "income premium" in the lottery tickets: the chances of winning a prize increase proportionally to the level of income. The higher the income of a taxpayer, the higher the spending and the higher the chances of winning. Compared to the representative taxpayer population, the winners exhibit roughly seven times higher mean annual electronic transactions (€ 28,413 compared to € 3,931) and a higher mean income (€ 15,877 compared to € 9,403). In addition, annual electronic transaction differences are exacerbated between the two samples with extreme values in the winners' population; 369 winners exhibit more than € 1 million in annual e-transactions. By construction the scheme selects high income individuals, with a high propensity of payments, as winners.

Estimated elasticities from a linear probability model quantify the effect of income on (i) electronic transactions and (ii) the probability of winning. A 10% increase in income is associated with a 1.8% increase in the volume of electronic transactions and with a 0.11% increase in the probability of winning the tax lottery. Information from partners belonging to the same household and who file taxes jointly, captures also an intra-household income effect. A 10% increase in spousal income is associated with 0.7% increase in electronic transactions and a 0.06% increase in the winning probability.

Turning to the income sources provides some striking results. I document a particularly large "occupational premium" for the self-employed in winning the lottery. Self-employed winners record annual mean electronic transactions of € 181,520 compared to € 14,626 for winners belonging to other occupational categories (wage-earners, pensioners and agricultural workers). They also record an economically large discrepancy between their electronic transactions and their declared income. Whilst for the representative population, spending by electronic means remains below income, for self-employed individuals who won the lottery, electronic transactions are 10 times higher than their income. This suggests that the self-employed might either under-declare their income or use business-transactions through their personal bank accounts. As a result, this group accounts for 8.3% of all winners compared to a population share of 4.1%. The lottery selects winners from the self-employed income category with a higher probability.

When quantifying this effect, I find that the occupational premium outweighs the income premium. Being self-employed increases electronic transactions by about 75% and the probability of winning by 18% compared to other income categories. Intra-household occupational effects are also significant; having a self-employed spouse increases the electronic transactions of their partner by 45% and their chances of winning the lottery by 3.5%. These result hold after controlling for a taxpayer's own, as well as, for spousal income.

The winning premiums documented in this paper appear to be an inherent problem in tax lotteries with a direct effect on the policy's progressivity. By incentivising consumers to ask for receipts, it is by construction the high income individuals and high spenders who have higher winning chances and, thus, benefit the most from the scheme's prizes. In addition, the self-employed who might use their personal accounts for large business transactions have higher chances of winning. Both income and occupational premiums might raise fairness concerns and ultimately undermine the effectiveness of tax lotteries.

To examine solutions that could make the scheme more progressive by construction, I perform monthly draws of the lotteries using Monte Carlo simulations under two scenarios. Firstly, a reformed ticket structure introduced in 2019 that became more concave, therefore awarding more tickets per euro spent in lower spending levels, combined with an upper limit of € 50,000 per month per individual after which no lottery tickets were awarded. Secondly, a stricter ticket limit in the monthly amount of tickets at € 1,000 and at € 5,000 per individual (instead of € 50,000). I find that the reformed ticket structure has only marginal effects, whilst the progressivity is improved partially through the upper limit of € 50,000. The annual mean electronic transactions halves, indicating that the ceiling becomes binding for extreme consumption values. The annual mean income of winners is reduced only marginally. Simulation results from the stricter limits of € 1,000 and € 5,000 per individual per month indicate a fairer distribution of prizes. However, the stricter the ceiling, the less the incentive to ask for receipts once the monthly limit is reached. In the Greek tax lottery, a ceiling of € 5,000 per individual per month strikes a good balance between the incentive to ask for receipts and a fair distribution of prizes.

The evidence presented in this paper contribute to our understanding of tax lotteries. Other studies have documented the policy's effectiveness in raising tax revenue. Notably, Naritomi (2019) analyses the Brazilian tax lottery and identifies a 21% increase in reported sales and a lower, yet significant, increase of 9.3% in reported revenue for the state of Sao Paolo in Brazil. The study mentions whistle-blowing and collusion costs as potential mechanisms for the increase. The Greek tax lottery is analysed in Nicolaides (2023). In contrast to the Brazilian tax lottery, tickets are awarded only when taxpayers complete payments through electronic payments, and in addition, no registration of payments is necessary; the tickets are awarded automatically to the entire population proportional to their amount of electronic payments. The paper documents an increase in regional VAT by 0.01% per additional winner. The main mechanism is idiosyncratic effects from winners, who spent more in electronic payments after winning, as well as, spillover effects in electronic consumption from winners to non-winners.

Beyond the policy's effectiveness in raising revenue, none of the previous studies have considered so far the income, consumption and occupational characteristics of winners. These characteristics are important in assessing how regressive or progressive the policy can be. Evidence in this paper contribute to the literature by bringing forth an inherent regressivity in the design of tax lotteries. Firstly, since tickets increase proportional to spending, the higher the income of individuals, the higher the chances of winning. Secondly, winning is not occupationally-neutral; some income categories who might utilise consumption for professional purposes gain an advantage. These lottery characteristics result in high-income earners, high-spenders and the self-employed/business-owners being selected more frequently as winners.

This paper contributes also to a vast literature that studies the efficiency of intra-household allocations of income. The income pooling hypothesis predicts that only household (or joint) income should matter for allocation decisions, and not who receives it. Browning *et al.* (1994) using Canadian household expenditure data show that who gets what depends on the income of each individual. Additional evidence are presented in Lundberg *et al.* (1997) using a natural experiment in the United Kingdom with child benefits being allocated to wives increasing women and children clothing expenditure. A survey of results in developed countries is presented in Chiappori *et al.* (2020). These findings are particularly applicable in developing countries, since household decisions are key for development; evidence from a number of countries suggest imperfect income sharing within a household (Quisumbing and Maluccio, 2003). Adding to these evidence, I quantify the elasticity of electronic consumption with respect to a change in spousal income. A 10% increase in spousal income increases electronic consumption of their partner by 0.76%, lending support to evidence of imperfect income sharing.

Finally, the findings add to existing empirical evidence on the marginal propensity to consume. Carroll *et al.* (2014) estimate the aggregate marginal propensity to consume for Greece to range between 0.10, when fitting a net wealth distribution, and 0.35, when fitting a liquid assets distribution. I estimate a marginal propensity to consume of 0.18, which falls within this range.

The estimate in this study is important as it is derived directly from the real electronic transactions of individuals and declared income in tax returns.

The remaining paper is organised as follows. Section 2 provides the institutional information of the Greek tax lottery. Section 3 describes the data. A descriptive analysis of the winners' characteristics, followed by a parametric analysis that quantifies these effects is provided in Section 4 . Simulations of reforms are shown in Section 5. Lastly, Section 6 concludes.

2 The Greek Tax Lottery on Electronic Transactions

Tax lotteries have become a common tool to mobilise consumers as a source of third-party reporting, to expand the tax base and ultimately to increase tax revenues (Naritomi, 2019). This trend is also visible in Europe, where several countries introduced tax lotteries during the European debt crisis (Fooken *et al.*, 2015).[1] In 2017, when the Greek Tax Authority was granted a strong institutional and financial independence, a tax lottery scheme gained traction. Technical advice towards the Greek Tax Authority supported and widely promoted the use of electronic payments in the economy, as a way to fight tax evasion.[2] The design of the lottery was thus incorporated in a broader strategy that aimed at curbing tax evasion by encouraging the use of electronic payments over cash payments. For this reason, the Greek tax lottery focuses specifically on incentivising electronic transactions over cash payments.

Electronic Transactions. For a long time, cash has been the prevailing payment method in Greece. However, capital controls in July 2015 led to a substantial increase in electronic payments.[3] From 2015 onward, Greece recorded a massive increase of debit cards issuance, which have been associated with increased tax buoyancy in the years that followed, despite a large negative economic shock (Hondroyiannis and Papaoikonomou, 2017). Additional incentives to promote electronic payments were introduced in 2016 (Law No. 4446/2016). Among others, they included an annual tax credit to all taxpayers who spent (from 2017 onward) a certain fraction of their (reported) incomes

[1]In 2011, during the first years of the economic crisis, Greece embarked on its first attempt to establish a "random monetary reward" scheme. The first project was called "@podeixi" (the Greek word for "receipt" in Latin characters), developed by the National Center of Scientific Research Demokritos in Athens. It received a Greek patent in 2011 (Patent No. 1007355). For more information see `http://www.obi.gr/obi/Default.aspx?tabid=127&idappli=X410275`. Mimicking tax lotteries from other countries, this first proposal sought to reward individuals who asked for paper receipts (regardless of the payment method). Due to lack of government support, however, this initiative never materialised. The tax lottery appeared only once in an official document; a 2014 leaked email of the Minister of Finance (also known as "Hardouvelis Email") to the heads of monitoring institutions known as the "Troika" during Greece's second economic adjustment programme. It was intended to serve as an eleventh hour proposal to bridge fiscal differences before the structural programme expiry in early 2015.

[2]A comprehensive strategy for the promotion of electronic payments to tackle tax evasion, appeared as a key deliverable by Greece's creditors in the summer of 2015, the implementation of which was linked to financial disbursements in the Memorandum of Understanding of the European Stability Mechanism programme to Greece. See p. 9 in `https://ec.europa.eu/info/sites/info/files/01_mou_20150811_en1.pdf`.

[3]Cash withdrawals were limited to € 60 per day per individual, but electronic payments remained unlimited.

in electronic payments; this policy is analysed in Nicolaides (2022).[4] At the same time, acceptance of electronic payments and introduction of Point-Of-Sale (POS) terminals became mandatory (in gradual roll-out phases starting 2017, depending on the profession). Overall, the legislative measures had a strong, positive impact on card use (Danchev *et al.*, 2020).

These reforms were complemented by the introduction of a comprehensive IT system. Starting from January 2017, banks were required to automatically report the total volume of electronic transactions per individual to the tax authority. This information, on the one hand, enabled the tax authority to assess whether a taxpayer would pass the minimum threshold of electronic payments (see Nicolaides (2022)). On the other hand, the reporting system also served as key building block for a further incentive for electronic payments: a lottery that rewards electronic transactions.

Electronic Transactions Lottery. At the end of each month, the *aggregated volume* of all eligible electronic transactions (but not each single transaction) of each bank client are submitted to the authority. All electronic payments – online banking transactions as well as debit, prepaid, and credit cards payments – by Greek taxpayers to businesses with presence in Greece or in other EU countries are eligible.[5] Based on the unique tax identification number of each individual, the tax authority adds up the total volume of completed electronic transactions in a given month for each taxpayer.[6] The resulting sum serves as the key input for the lottery.[7]

The monthly volume of electronic transactions are converted into tax lottery tickets according to a given ticket-awarding mechanism (or *TAM* hereafter). In an attempt to increase the relative chances of low-consumption groups to win the lottery, the TAM has a concave structure: at higher levels of electronic transaction volumes, an additional euro would translate into fewer tickets. This point is documented in Table 1, which presents the detailed structure of TAM.[8] While the first euro of monthly electronic consumption would translate into one ticket, the € 1,001st would yield only

[4] For 2017, the cutoffs were 10% for declared annual income up to € 10,000, increasing to 15% for any income between € 10,000 to € 30,000 and 20% for incomes exceeding € 30,000 (with an upper limit for very high incomes). Such a tax credit already existed since 2011. Instead of being tied to electronic payments, however, individuals had to collect and keep paper receipts in case of a tax audit. If electronic payments would fall below the cutoff, individuals would face higher tax obligations (which are automatically calculated by the tax authority and saliently reported in tax returns and online bank accounts). The tax credit thus provided a clear incentive for consumers to use electronic payments in everyday transactions.

[5] Eligible electronic payments include, for instance, paying an airline ticket, buying a product online, or paying for online subscription services. Eligible credit and debit card payments include paying for any product or services at any shop within the EU. Non-eligible payments include paying for rent and any government related payments such as taxes and fines. Transactions with firms from non-EU countries are excluded since these cannot be fully identified and classified by the electronic systems of banks.

[6] Banks request the tax identification number when opening a bank account and, thus, have information on each individual. Multiple accounts and bank cards registered on an individual are aggregated using the their tax numbers. Joint bank accounts include the tax identification numbers of multiple individuals. However, the means of payment (a credit or a debit card, for instance) is always registered on the name of one individual. Banks take into account this differentiation when submitting data to the tax authority.

[7] Details on the functioning of the lottery can be found in Law No. 4446/2017, and Ministerial Decision 1161/17-10-2017. Communication by the Tax Authority on matters regarding the lottery can be found in `https://www.aade.gr/menoy/miniaies-synallages-kai-lahnoi`.

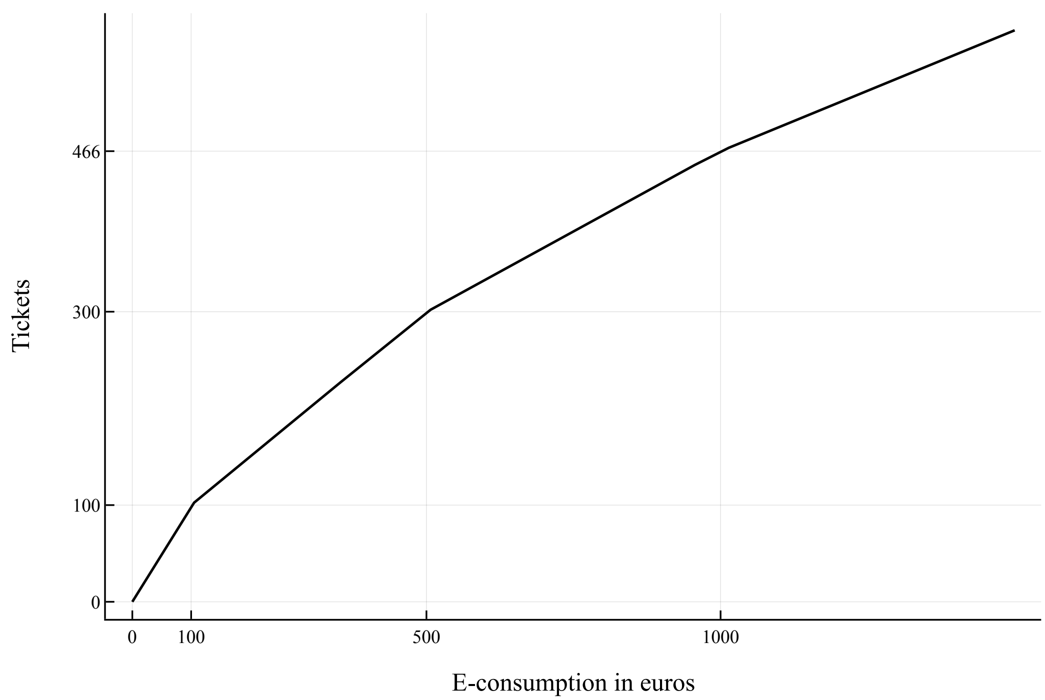
[8] This scheme was in place until May 2019. In Section 5 I discuss a reform and compare the initial with the post-reform TAM.

0.25 tickets. Note further that the TAM does not contain any upper bound. Figure 1 plots the resulting euro-to-ticket mapping.

Table 1 Ticket-Awarding Mechanism (Jan 2017 – May 2019)

Total amount of monthly e-transactions	Tickets awarded	Maximum number of tickets
€ 1 – 100	1 ticket per € 1	100
€ 101 – 500	1 ticket per € 2	300
€ 501 – 1,000	1 ticket per € 3	466
> € 1,000	1 ticket per € 4	No limit

Fig. 1 Euro-to-Ticket Scale



Notes: The graph illustrates the scale used to convert the aggregate level of monthly electronic consumption to eligible tickets in the lottery. Banks sent the aggregate level of electronic consumption completed by each individual and this is converted to ticket using the following scale. At €1-100, tickets correspond at 1 for every €1. At €101-500, tickets correspond at 1 for every €2. At €501-1,000, tickets correspond at 1 for every €3. For over 1,000, tickets correspond at 1 for every €4. There was no upper limit in tickets. Details about eligible payments and additional information on the institutional structure are explained in Section 2.

Applying the TAM on the volume of electronic transaction data from a given month m yields the total number of tickets that enter a draw in month $m+1$. A random draw then picks 1,000 winning tickets (with the constraint that an individuals can only win once in a given month).[9] Each winner receives € 1,000. The prize is tax exempt and cannot be confiscated.

The initial plan was to introduce the first tax lottery draw in January 2017. While the bank reporting system started operating in January, there were several legislative and technical issues that resulted in a delay. In Nicolaides (2023), this delay is utilised to identify the VAT revenue effect of the tax lottery. A public announcement of the lottery took place in early October 2017. At that point in time, the TAM and the prize structure were made public.[10] Subsequently, the first draw took place in November 2017, based on electronic transactions from October. In December 2017, draws from the previous months in 2017 took place en bloc, based on electronic transactions from January to September 2017. The tax authority announced 10,000 winners in December; corresponding to 1,000 winners for each month from January to September 2017 and 1,000 winners for December 2017. Search volumes in Google recorded in Greece at the time for the word "Lottery" in Greek are shown in Figure 2. While the search volume is close to zero in the months prior to the first lottery, the volume spikes at the end of November (1st lottery), while the highest volume was recorded at the end of December (at the time en bloc draws), indicating increasing public awareness.[11]

Since November 2017, the lottery is running on a monthly basis, each time based on the previous month's transaction data. On the day of the draw (usually the 24th day of a month), the authority publishes the tax IDs of all winners on a dedicated website. Thereafter, winners are also contacted by emails or via text messages on their phone. The prizes are transferred directly to winners' bank accounts about a week after winning.

The Greek tax lottery is unique and differs from other tax lotteries in numerous ways. Firstly, it is hardly based on self-selection. Typically, consumers must register in a system and collect receipts and in order to participate in a tax lottery (see, e.g., Naritomi, 2019). Instead, the Greek lottery is the first to almost automatically include the vast majority of taxpayer: everybody who (i) holds a bank account and (ii) makes an electronic transactions in a given month takes part in the lottery.[12] Secondly, tickets are mapped directly to electronic payments. Hence, the objective of the Greek lottery is to incentivise consumers to switch from cash to e-payments – which tend to be more difficult to conceal and should thus facilitate monitoring and tax enforcement. In fact, as

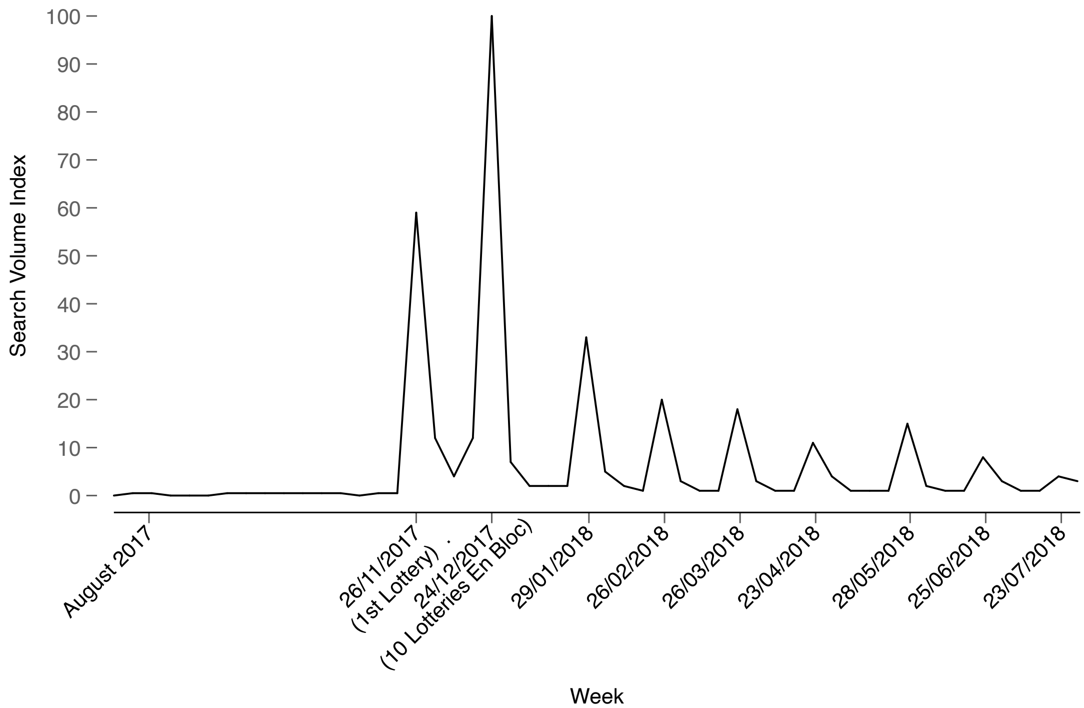
[9] To prevent that the lottery gets rigged, the tax authority follows a two-step process. First, the random draw is outsourced to a university. Based on the total number of monthly tickets, the university generates a series of random numbers which are submitted to the tax authority. Second, the authority modifies these numbers using a pre-determined formula, which is unknown to the university.

[10] The TAM was more widely discussed in the press when the first lottery draw took place.

[11] The search volume index records increases at the end of each month thereafter, in line with the time of monthly draws.

[12] According to the World Bank's Global Findex database, 85% of individuals in Greece above the age of 15 had a bank account in 2017. As some of these are joint 'family' accounts (see fn. 6), the formal banking system includes almost the entire population.

Fig. 2 "Lottery" Google Trends in Greece



Notes: The figure plots the Google search volumes (indexed from 0-100 on the y-axis) for the word "lottery" in Greek. The geographical area is constraint to Greece alone. The timeline is shown on the x-axis, containing weekly trends for every week starting with the first week of August 2017 and ending in the last week of July 2018.

discussed above, the lottery emerged jointly with a system that provides the tax authority with full information on an individual taxpayer's level of monthly electronic payments.

One caveat concerns the use of private bank accounts for business purposes. In an attempt to separate individual from business transactions, the tax authority obliged all firms to use separate business bank accounts (Joint Ministerial Decision No. 45231/2017). However, the implementation of this measure was significantly delayed and the enforcement of the requirement was gradually pushed from summer 2017 to spring 2019.[13] Hence, while the intention of the tax authority was to include only private electronic transactions ('consumption') by individuals, the lottery initially (i.e., before 2019) included a non-trivial volume of business transactions. The line separating business and individual transactions will be, as a result, particularly blurry for self-employed individuals and owners of small firms, since business transactions can be made through their personal bank accounts (which are tied to their unique tax identification number).

[13]The initial ministerial decision was published in April 2017 giving business owners 3 months to comply with the decision. Renewed decisions where issued 3 times with deadlines being pushed to 15/1/2019, 28/2/2019 and 30/4/2019.

3 Data

The data contain (a) information on the monthly level of electronic transactions for the period from January 2017 to July 2018 as well as (b) matched information from tax returns in 2017.[14] I observe the annual declared pre-tax income for submitted tax returns of tax units, which include either a single or a joint filing (the latter consists of income from the main taxpayer and the spouse) from economic activities in 2017. In case of joint filings, the monthly level of electronic transactions correspond to only one of the two individuals in the tax unit. The data provided by the tax authority were anonymous, while monthly electronic transactions were rounded to the nearest € 10 and annual declared income information to the nearest € 5.[15] For joint filings, I observe income values for both partners, enabling the calculation of the tax unit's declared income.[16] For single filings I observe the declared income of the single person in the filing, which is also a single-household income.[17] Lastly, it is compulsory to file tax returns even if an individual has exactly zero income. Thus, the data include many students above the age of 18 (in tertiary education) as well as the unemployed. With an unemployment rate of 21.5% in 2017, the latter group constituted a significant proportion of the working population. The zero income group, however, might also include tax units who conceal all of their incomes.

In addition to the declared amount, the data indicates the source(s) of income from five different categories: income from wages (subsequently *WG*); self-employed/business income (*SB*); agricultural income (*AG*), income from pensions (*PE*), or zero-declared income (*NO*).[18] WG includes income received from salaried activities. Hence, I observe a tax unit's reported annual gross salary. The SB category includes sole proprietorships, such as the self-employed, sole traders and small firms. This is the most common legal form of business activities in Greece.[19] The data contain annual net profit from business activities (but would not report loses). AG contains declared annual income from agricultural activities, such as for farm owners, agricultural workers and small cultivations. PE includes all individuals who receive pensionable income from main or auxiliary pensions. The data report the (pre-tax) annual pension income. NO contains individuals who have reported zero income in 2017. This category includes individuals who are obliged to submit tax returns, even if their income is zero, such as tertiary education students and the unemployed.

[14] The last day of tax return submission for the tax year of 2017 was July 30, 2018. The tax returns underwent a basic plausibility check and tax payment statements were issued by the tax authority in August 2018. The data in this paper were received in October 2018.

[15] The declared income does not include any income received from the government as a subsidy to the household, such as social welfare transfers for poor households, nor any tax credits added before the final tax calculation.

[16] For the 2017 tax returns, joint filing was mandatory for married couples. Law no. 4172/2013 provided that the main taxpayer of the household is the husband, responsible for submitting the tax return, while the wife must sign-in before finalising the submission and give consent to the declared amounts.

[17] However, I cannot distinguish individuals and households in the case were the main taxpayer has declared some level of income, and the spouse has declared 0 income.

[18] These income categories in Greece corresponded to different pension insurance funds contributions that existed in the past.

[19] In 2017, 1.2 million sole proprietorships existed out of a total of 1.4 million firms. These correspond to the SB category. See the Statistical Business Register in 2017 available from the Hellenic Statistical Authority at: `https://www.statistics.gr/en/statistics/-/publication/SBR01/-`.

However, it might also contain individuals from the SB and AG income categories, who report zero income. The percentages of single filings, joint filings for households and income categories, included in the sample, are shown in Table 7 in Appendix A.

Note that a given tax unit might of course declare incomes from multiple sources (categories). Below, I will use indicators that define the primary source of reported incomes. These dummies also serve as a proxy for the (primary) occupational activity.

Due to between category variation in third-party reporting (Kleven *et al.*, 2011), there are major differences in the opportunities to under-report incomes. For WG and PE income, the income values (as reported by employers or pension funds) appear automatically in the individuals' tax returns. SB and AG incomes, on the contrary, are self-reported. Hence, as noted above, some individuals with non-zero incomes from these sources might not report any income and thus end up in the NO category.

Sample. I obtained data for two different samples. First, the universe of 18,897 tax units with members that have won the lottery during the first 19 consecutive lottery draws (based on the e-transactions for the months from January 2017 to July 2018);[20] second, a randomly drawn sample of 50,000 tax units that did *not* win the tax lottery in any of these 19 draws. For the winners' sample, and for all 19 months covered, I observe the monthly volume of electronic transactions for the winning individual, while for the randomly-drawn sample I observe electronic transactions for one individual in the tax unit (as reported from banks to the tax authority). Among the first sample, I also observe the month of winning. For both samples, I observe the tax return filing, which contains one annual income value for single member tax units and two annual income values for joint filing households. In the case of joint filings, monthly e-transactions correspond to one of the two persons in the household. Table 7 in Appendix A presents basic summary statistics for the two samples.

Based on the information from these data, the analysis will focus on comparing the characteristics of winners against a reconstructed baseline population of tax units. Specifically, the winners from the 2017 draws will be used, since this has two advantages. Firstly, one can compare the annual declared income from 2017 with the monthly e-transactions $z_{i,m}$ for an entire calendar year, i.e., I compute $Z_i = \sum_{m=1}^{12} z_{i,m}$ for each taxpayer i covered by in the data. A second advantage is that the e-transactions values from this year are, to a large extent, not influenced by the lottery itself. To see this, recall that the tax lottery was announced in early October 2017 and the first draw took place in late November (see Section 2). Note further that the broader public only took notice of the lottery after the first draw and, in particular, after the draws that took place en bloc at the end of December 2017. Hence, electronic transaction values for the months January to September 2017

[20]The sample covers less than 19,000 winners (1,000 winners for each of the 19 lotteries) since some of the prizes were refused or not claimed by those who were drawn. This happened in 103 out of 19,000 cases, of which 40 from draws in 2017. In informal conversation with the tax authority it has been reported that because essentially all taxpayers are included by default in the lottery (and do not actively opt-in), some winners are unaware of its existence and regard the winning message by the tax authority as fraud.

were recorded before the announcement of the lottery; and even for the remaining months the tax lottery might play a minor role in shaping annual e-transactions.

To allow for a meaningful comparison of winners relative to the baseline population, one has to account for the different sampling of the two samples. The non-winners sample was drawn randomly from the population of taxpayers – conditional on not having won (given the TAM and the taxpayers monthly levels of e-transactions). The other sample contains the universe of winners (which were drawn under the same TAM conditional on the taxpayers' e-transaction pattern). To arrive at a sample that represents the population of taxpayers, I expand (or re-weight) the non-winner population such that they match the overall number of 'lottery tickets' (i.e., the aggregated amount of e-transactions) observed in 2017. As a plausibility check, I compare the obtained number of tax units with in the corresponding number in the population of taxpayers. The details of this approach are described in Appendix A.2.

Basic summary statistics from the resulting, expanded sample are presented in Table 2. The table compares the baseline tax unit population with the tax units of winners from 2017. Comparison takes place at the individual level within tax units, since I have e-transaction information only for one individual in each unit. The tabulation reveals several interesting characteristics. Firstly, the mean e-transactions and mean income of winners is much higher than the rest of the population. Secondly, the SB income category is over-represented and exhibits very high levels (and variance) of e-transaction. Thirdly, winners in the NO category, had a particularly high level of e-transactions. These discrepancies form the centre of the analysis in the following section.

4 Who Wins the Lottery?

This section examines the characteristics of lottery winners highlighted in Table 2. I first provide descriptive evidence on the selection implied by the TAM (considering the e-transaction volume, income, and the primary income source at the individual and tax unit level). Second, I parametrically quantify the selection along multiple dimensions.

4.1 Descriptive Analysis

4.1.1 E-transactions and Income

A comparison of the mean annual e-transactions is shown in Table 2. Individual winners exhibit roughly seven times as high mean annual e-transactions compared to the representative population. This reflects the basic property of the lottery: the chances of winning increase proportionally to the level of e-transactions. Since the TAM contains no upper bound (there is no maximum number of assigned tickets in a given month), the probability of winning *cet.par.* approaches unity if $z_{i,m} \to \infty$.

Table 2 Summary Statistics: 2017- Winning Tax Units versus Population of Tax Unit

	Winners			*Population*		
	Obs.	Income	E-Trans.	Obs.	Income	E-Trans.
by Primary Income Category:						
SB	988	20,753	181,520	266,317	12,120	11,420
Self-Employed/Business inc.	8.3%	(32,955)	(695,170)	4.1%	(25,891)	(60,163)
WG	5,773	18,357	10,857	2,890,322	11,418	4,064
Wage income	48.3%	(38,738)	(45,598)	44.7%	(13,941)	(6,138)
PE	2,704	14,631	10,964	1,573,228	11,875	3,322
Pension income	22.6%	(6,347)	(67,821)	24.3%	(6,046)	(5,350)
AG	503	47,423	15,532	340,746	17,582	3,817
Agriculture income	4.2%	(106,648)	(33,355)	5.27%	(38,113)	(6,627)
NO	1,818	0	27,618	1,397,996	0	2,935
Zero income declared	15.2%	(0)	(197,309)	21.6%	(0)	(15,109)
No Filing	174	-	37,119	-	-	-
(Tax return not submitted)	1.45%	-	(342,630)	-	-	-
Total	11,960	15,877	28,413	6,468,897	9,403	3,931
	100%	(37,277)	(229,919)	100%	(15,036)	(15,243)

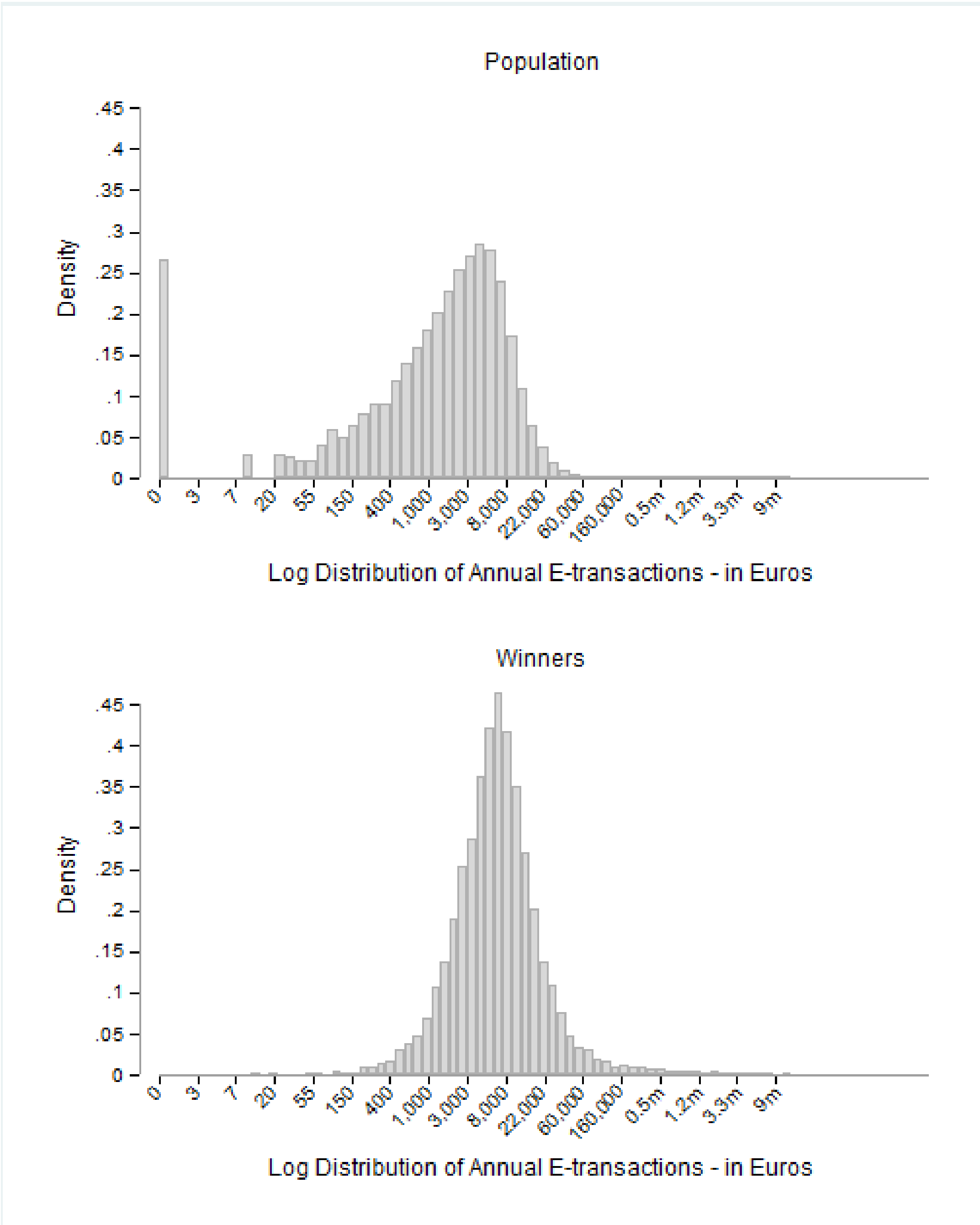
Notes: The table presents the number of observations, the mean declared income and the mean e-transactions Z_i in 2017 (nominal € values) for individuals in tax units with matched e-transactions information. They are presented by primary income source as has been declared in their tax returns. The groups pair taxpayers according to their main source of reported incomes coming from wages (WG), self-employed/business incomes (SB), agricultural income (AG) or income from pensions (PE). An additional category indicates zero-declared income (NO). Standard deviations are in parenthesis. The 'Winners' sample includes all individual winners from the monthly draws based on 2017 e-transactions. For joint-filings income and e-transactions information are matched for one individual, either main taxpayer or spouse. The 'Population' sample is a reconstructed sample of the tax unit population (see Appendix A.2).

While this holds for a given month m, as e-transactions fluctuate between months, the annual level Z_i is only an indirect indicator for the selection implied by the TAM.[21]

The difference in e-transaction (more specifically, in $\log(Z_i)$) can also be seen in Figure 3. Among the total taxpayer population, the distribution is right skewed and bi-modal: almost 25% of

[21]The difference between mean e-transactions for winners relative to the population is even more evident in a monthly comparison. The taxpayer population's monthly mean e-transactions followed an upward trend in 2017, fluctuating between € 278 (in the beginning of the year) to € 445 (at the end of the year). The mean e-transactions of those who have won in a particular month fluctuated around € 4,000 (without observing any upward trend).

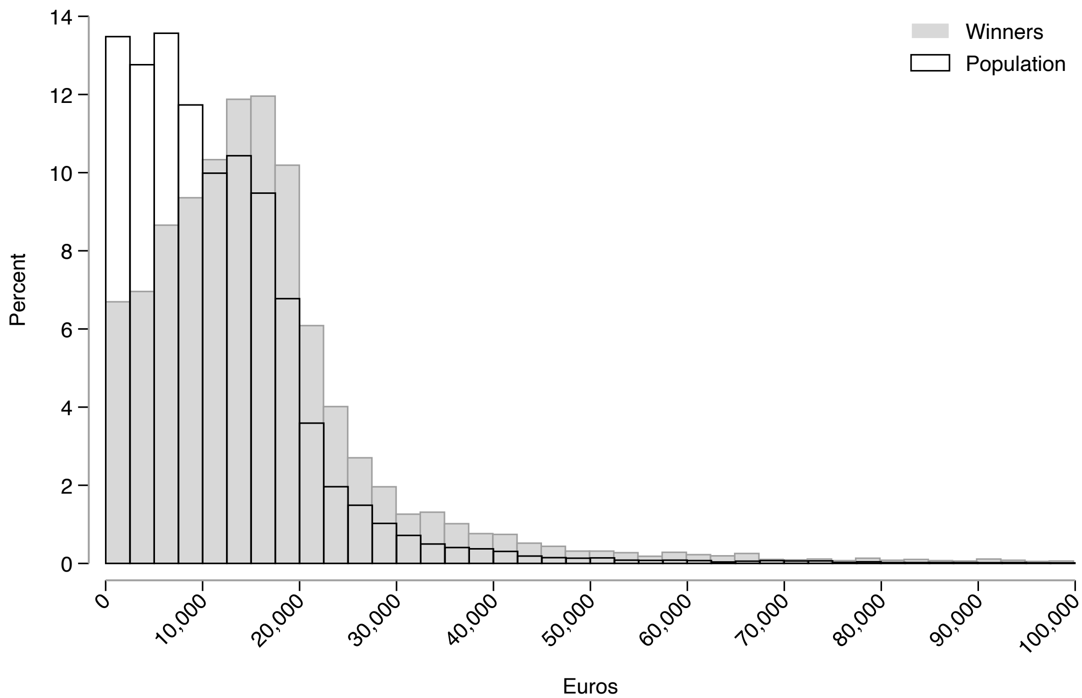
Fig. 3 Distribution of Annual Electronic Transactions, 2017



Notes: The figures present the log of annual e-transactions distributions in 2017 between individuals in the tax unit population and individuals who have won the lottery in 2017. The x-axis is a log scale representing the equivalent values in €. Tickers are rounded to the nearest thousand in €. The population distribution includes the individuals from 6.4 million tax units. The winners distribution includes 11,960 winners in tax lotteries that took place in 2017. Monthly e-transactions of individuals were summed up over the 12 months to create the annual e-transactions. Monthly values in the data were rounded to the nearest € 10 by the tax authority. The distributions are drawn on the same scale, sharing the same y- and x- axes.

taxpayers have basically zero e-transactions. The mean annual e-transaction volume is € 3,931 (median € 1,940). For winners, log-e-transactions are symmetrical and normally-distributed, with a significantly higher mean of € 28,413 (median € 6,400). In contrast to the general population, where there is hardly any mass in the range $Z_i >$€ $22,000$, there is a heavier right-hand tail for winners, with a non-trivial share of annual e-transactions volumes well above € 60,000. For 2017, I observe 334 winners with more than € 1 million annual e-transactions, 34 with more than € 2 million and one extreme value of more than € 9 million e-transactions (who has won twice in 2017).

Fig. 4 Distribution of Declared Income 2017



Notes: The figure compares the distribution of declared income in 2017 of lottery winners with the corresponding distribution of individuals in the tax unit population. The population has been reconstructed from a random sample of 50,000 tax units. The graph is truncated at € 100,000, as right-tails diminish quickly in the distribution.

The differences in e-transaction are mirrored in the levels of reported income. The mean declared income among winners is € 15,877, whereas it is € 9,403 among the tax unit population (median values are € 12,113 and € 6,850, respectively). These differences are economically and statistically significant. They are also illustrated in Figure 4. As can be seen a substantial proportion of individuals in the tax unit population report income below € 10,000. Among lottery winners, there is less mass in this income range but over-proportionately many cases with higher income levels. Overall, the figure clearly reflects that the lottery winners are (judged against the general population) higher-income taxpayers.[22]

I then compare the gap between annual e-transactions and incomes within winners and the tax unit population. Among the latter group, the log values of the two variables display a positive correlation of 0.29. For winners, the mean e-transactions are about twice as high as the mean declared income (see Figure 8 in the Appendix). This difference, however, is largely driven by some outliers (with very high e-transactions values). The correlation between the log values of the two variables in the winners' sample remains positive but falls to 0.11, indicating that e-transactions become weakly associated to income. Lastly, the e-transaction/income ratio is 1.79 among winners and 0.42 among the general population. Winners spent almost two times their income on e-transactions, while the general population spent less than half. For every third winner (33.5%) I observe a ratio above unity, i.e., an e-transaction volume that is above the declared income.

4.1.2 Income Sources

This section examines the extent to which the observed patterns are driven by differences in the (primary) income source. Firstly, note that Table 2 documents significant differences in mean incomes between winners and the general population within each group with a given primary income. The table further documents that individuals from the SB category are massively over-represented among lottery winners. Relative to a population share of 4.1% , this group accounts for 8.3% of all winners.[23] While WG, PE and AG individuals are observed at similar percentages as in the population of tax units, taxpayers in the zero declared income (NO) group are under-represented.[24]

Comparing the groups with different primary income sources, Table 2 reports very high levels of e-transactions among the SB category: the mean among winners is € 181,520 (median of €17,565), with a high standard deviation. Among the total population, individuals with SB incomes have a mean (median) e-transaction volume of € 11,420 (€ 4,410). Hence, in addition to the fact that individuals from the SB group win the lottery more often than others, the lottery also selects (within the SB group) winners with unusually high e-transactions volumes.

The discrepancy of e-transaction for SB individuals is depicted in Figure 5, where declared income and e-transactions of SB are compared against the pooled groups WG, PE and AG.[25] For the latter income categories, mean annual e-transactions are about one-third of mean declared incomes in the tax unit population. Among the SB population, the corresponding share is around 90%. Hence, these income categories display a different pattern. As shown in Figure 5 (b), these differences

[22] For illustration purposes, Figure 4 is truncated at € 100,000. However, there are a few observations of winners with very high incomes well above € 100,000.

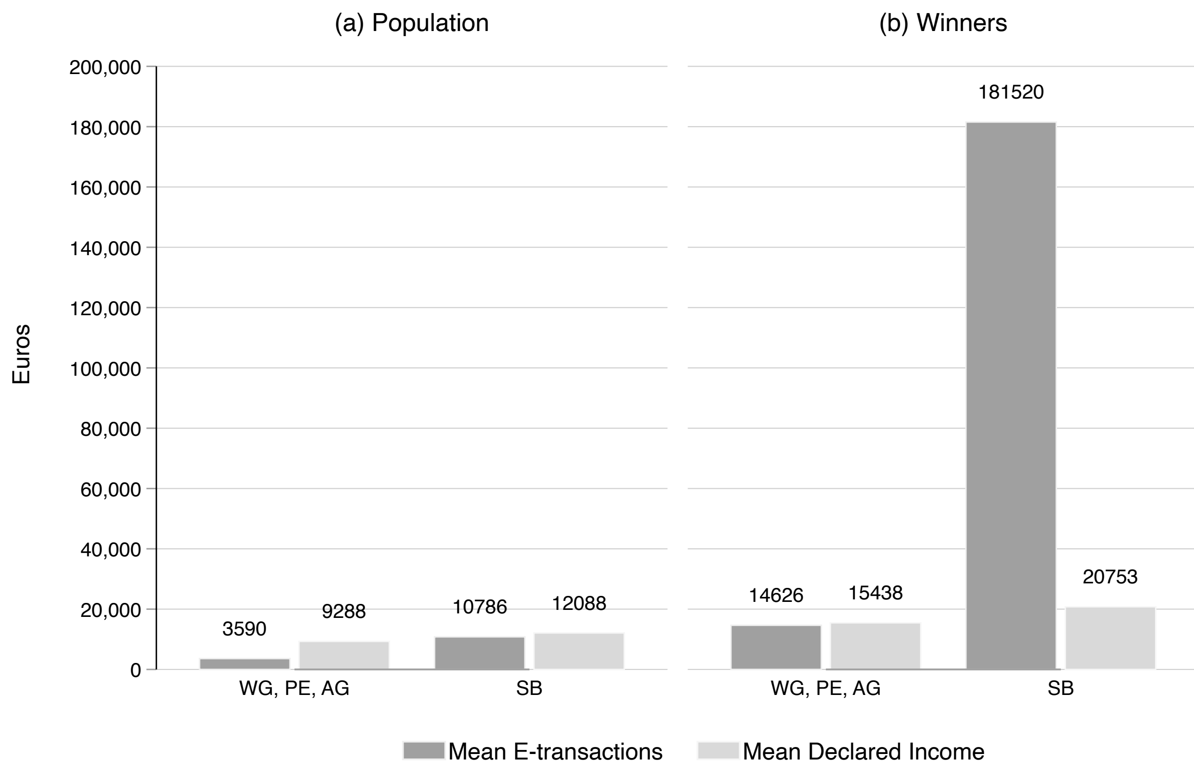
[23] See Section 3 for more information on the SB category.

[24] The latter group, which still accounts for around 15% of all winners, can be composed by heterogeneous types: (a) students or unemployed individuals who have non-zero e-transactions; (b) individuals with self-employed or business income that report zero or negative incomes (losses).

[25] To allow for a meaningful comparison, the NO category individuals are excluded.

are even more pronounced among winners: the mean e-transactions of SB winners are ten times as high as their mean declared income. Among the winners from the other income categories, the e-transaction/income ratio is below one.[26]

Fig. 5 Declared Annual Income and E-transaction by Income Sources



Notes: The figures compares mean e-transactions and mean declared income for groups with different primary income sources: self-employed/business income (SB) vs other non-zero incomes from wages, pensions and agricultural activities (WG, PE, and AG). Zero-declared income individuals are excluded from this comparison. Figure (a) is based on the population of tax units and figure (b) presents the lottery winners from 2017.

The stark difference between e-transactions and declared income among winners from the SB group suggests that these individuals are using electronic means and their private bank accounts when paying for business-purpose expenses. The flow of business transactions results in particularly high e-transaction levels compared to both, their own declared income as well as the e-transaction volumes of other income groups. As a result, one can observe disproportionately more tax units with primary SB income among winners. Recall from Section 2, that the use of private bank accounts

[26]Out of 988 winners from the SB category, 64% exhibit e-transactions higher than their income. Among the SB group in the general population, the equivalent percentage is 39% – which is still much high compared to taxpayers with other income sources. For example, in the WG group, about 9% of the population (16% of winners) have annual e-transactions levels higher than their incomes.

for business proposes was prohibited in 2019. Hence, it was still legal in 2017 for some business expenses to be channeled through private bank accounts.[27]

4.1.3 Income Sources within Households

Another characteristic that affects winning chances is the household composition and their income source types. I first examine the influence of having a spouse with SB income. As long as some couples share their (private) bank account, and if partners with SB income use the accounts for business (e-)transactions (see above), one should expect to observe higher e-transactions levels for individuals jointly filing with a SB (rather than a non-SB) spouse. To assess this case, I focus on individuals who filed jointly in 2017. Overall, I observe that 37% tax units in the data file jointly, which is very close to the official percentage of 40% for 2017.[28]

Figure 9 in Appendix A compares e-transaction and income levels of individuals who jointly file with spouses that either receive their primary income from SB or with spouses in the WG/PE/AG income group.[29] (To facilitate interpretation, the sample underlying this graph excludes individuals from the SB and the NO income group). Having a partner in the SB income group is associated with higher levels of e-transactions. This holds for the general population (Panel (a) of Figure 9) but, more strongly among the group of winners (Panel b).[30] At the same time, the partner's income source does not make a difference for the reported income. Overall, the data indicates that (many) jointly filing couples seem to share private bank accounts and that SB partners seem to use these accounts for business transactions.[31]

4.1.4 Descriptive Analysis Summary

The descriptive analysis documents several key implications of the lottery's design. First, as the chance of winning is proportional to the volume of e-transactions, the lottery selects (more) high e-transaction taxpayers as winners. Since e-transactions positively correlate with incomes, I document an over-representation of higher-income taxpayers among winners. Whilst the concave structure of the TAM was introduced in the scheme to ensure better winning chances for the

[27] The observed pattern might also originate from the illegal underreporting of incomes: since the SB group has (relative to third-party reported incomes) more opportunities to conceal income (Kleven *et al.*, 2011), the vast e-transaction/income gap may therefore – at least in parts – reflect income tax evasion. The data do not allow us to quantify this channel.

[28] This information is included in the annual statistics published by the tax authority in `https://www.aade.gr/menoy/statistika-deiktes/eisodima/etisia-statistika-deltia`.

[29] Appendix Figures 10, 11 and 12 illustrate the same type of sample split for individuals with WG, PE and AG spouses, respectively.

[30] The differences are even greater if the spouse receives any part of income from SB activities (instead of having SB as a primary income source; column 1). Table 9, column (2), documents that this difference in annual e-transactions is economically and statistically highly significant. The difference is hardly affected by controlling for annual declared income in 2017, as shown in column (3).

[31] The pattern might also be shaped by individuals who record certain private, household expenses (such as the purchase of a personal computer) as business input costs in order to exempt these costs from VAT.

low-income taxpayers, the results indicate that this was not sufficient in ensuring equal chances. A sizable gap between the income of lottery winners relative to the population exists, suggesting a regressive policy.

Second, e-transaction patterns appear to be strongly affected by the use of private bank accounts for business purposes. High e-transactions associated with business activities result in more SB individuals being selected as winners. In 2017, the TAM thus rewarded a subgroup of individuals from occupations (with SB incomes) that resulted in very high e-transaction levels, well above their declared income. Third (and consistent with the previous point), I observe spillovers within jointly filing couples: a taxpayer, whose spouse receives SB income, has a higher amount of e-transactions. This increases the chances of winning the lottery.

4.2 Parametric Analysis

This section quantifies the patterns documented in the descriptive analysis. In particular, I explore the role of individual and spousal income level and income sources for (i) the level of e-transactions and (ii) the probability of winning the lottery. First, I consider models of the structure:

$$\log(Z_i) = \beta_0 + \beta_1 \log\left(Y_i\right) + \beta_2 \log\left(Y_{j|i}\right) + \beta_3\, SB_i + \beta_4\, SB_{j|i} + \beta_5\, Joint_i + \varepsilon_i \tag{1}$$

where Y_i indicates the reported income and SB_i is a dummy variable indicating business or self-employed being the primary income source. The sub-index $j|i$ measures these variables for i's spouse j. $Joint_i$ is a dummy variable indicating an individual who has filed jointly with a spouse. Note that β_1 and β_2 capture taxpayer i's elasticity of e-transactions with respect to their own and their spouse's income, respectively. Perfect income sharing within a household (plus equal propensities to spend money electronically) would imply $\beta_1 = \beta_2$.

Columns (1)–(3) in Table 3 reports OLS estimates that follow the structure of equation (1). The estimated β_1 suggests that a 10% higher income correlates with a 1.8% increase in e-transactions. This measure is similar to a marginal propensity to consume estimate for electronic consumption only. It captures how much electronic consumption changes to a change in income. Carroll *et al* estimate the aggregate marginal propensity to consume for Greece to range between 0.10, when fitting a net wealth distribution, and 0.35, when fitting a liquid assets distribution (Carroll *et al.*, 2014). The estimate in Table 3 falls within this range. The coefficient hardly changes in Column (2), when spousal income is controlled for. The correlation with the partner's income is significantly lower: a 10% higher income of the spouse is associated with a 0.7% increase in individual i's e-transactions. F-test reject the null $\beta_1 = \beta_2$ with $p < 0.001$. The results are quantitatively similar in column (3), where I focus on taxpayers who filed jointly. The findings indicate imperfect income sharing within a household (Browning *et al.*, 1994; Lundberg *et al.*, 1997) or differential propensities to engage in e-transactions.

The estimates further document that, consistent with the descriptive evidence from above, income from self-employment or business activities is associated with significantly higher levels of e-transactions. The estimated semi-elasticities imply that receiving the primary income from this source (SB) is associated with an approximately 75% higher level of e-transactions. A similar estimate is obtained for a spouse with primary income from SB: the corresponding semi-elasticity for the $SB_{j|i}$ dummy is around 45% (see Column 2). It is worth stressing that this holds while controlling for the taxpayer's own, as well as, for the spousal income. Hence, the pattern reflects an occupational rather than a mere income correlation.

Table 3 Estimation Results

	Log(e-transactions)			P(winning)			
	(1)	(2)	(3) joint-filers	(4)	(5)	(6) joint-filers	(7)
Log-Income	0.181***	0.179***	0.205***	0.011***	0.011***	0.016***	0.002
(β_1)	(0.003)	(0.003)	(0.008)	(0.000)	(0.001)	(0.001)	(0.002)
Log-Income Spouse		0.073***	0.076***		0.006***	0.008***	0.001
(β_2)		(0.005)	(0.006)		(0.001)	(0.001)	(0.001)
Self-empl./business	0.773***	0.745***	0.519***	0.179***	0.177***	0.186***	0.021
(β_3)	(0.050)	(0.050)	(0.074)	(0.014)	(0.014)	(0.022)	(0.025)
Self-empl./business Spouse		0.445***	0.466***		0.036**	0.035**	-0.005
(β_4)		(0.063)	(0.063)		(0.017)	(0.017)	(0.021)
Joint Filing		-0.130***			-0.010		-0.008
		(0.044)			(0.007)		(0.008)
Tickets 2017							0.000***
							(0.000)
Constant	5.582***	5.371***	5.002***	0.098***	0.080***	0.009	0.019
	(0.026)	(0.028)	(0.090)	(0.003)	(0.004)	(0.014)	(0.012)
F-Tests (p-values):							
$\beta_1 = \beta_2$		0.000	0.000		0.000	0.000	0.461
$\beta_3 = \beta_4$		0.000	0.594		0.000	0.000	0.338

Notes: The table presents estimation results following the structure of equation (1). The dependent variable in columns (1)–(3) are e-transactions ($\log(Z_i)$) and, in columns (4)–(7), an indicator for winning the lottery. Coefficients and standard errors in columns (4) – (7) are multiplied by 100. The sample is $N = 6,468,609$ observations, except for columns (3) and (6), where the sample is constrained to $2,406,683$ jointly filing taxpayers. Robust standard errors (clustered at the level of 50,000 + 11,960 unique taxpayers) in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

As a second step, I replace the dependent variable from equation (1) with a dummy W_i which indicates that the taxpayer has won in the lottery during 2017. The results from linear probability model estimates are presented in Columns (4)–(7) in Table 3 and they document a positive correlation between the own and the partner's income with the chance of winning, which in itself reflects the positive correlation between income and e-transactions (see columns 1 – 3). It is striking to observe that individuals with primary income from SB or (jointly filing) taxpayers with a spouse that receives SB income have a significantly higher chance to win in the lottery. The point estimates from column (5) indicate that an individual with SB income has a 0.00177 log-point higher chance of winning the lottery – which is non-trivial given that the baseline chance of winning are the same magnitude (roughly 12,000 tax units over 6.5 million tax units). Put differently, having an SB income is *cet.par.* associated with a twice as high chance of winning the lottery. The association with the partner's primary income source being SB is smaller but still statistically significant at the 5% level. Results are qualitatively unchanged when I estimate only in the sample of jointly filing taxpayers.

Finally, it is worth noting specification (7) of Table 3 which adds the annual number of tickets assigned to an individual in 2017. It is reassuring to observe that, controlling for the total number of tickets, renders all other variables statistically insignificant. This suggests that the lottery is not rigged. The correlations between winning and the level and source of incomes are merely shaped by the association of these variables with the amount of e-transactions. It is the latter which then translates into tickets and, ultimately, the probability of winning the tax lottery.

5 Lottery Reforms

The evidence from above indicates that individuals (and households) reporting incomes from business activities or self-employed labor, benefit from the lottery design; private bank accounts are used for business purposes, which inflates the (seemingly private) e-transaction volumes of these taxpayers. This 'occupational premium', however, applied in the beginning of the lottery. Since May 2019, all firms, self-employed and sole proprietorships were required to switch all business activities to separate professional bank accounts. While the level of enforcement of (and compliance with) this rule remains unclear, there is at least an appropriate regulatory framework in place.

The analysis further documents that the lottery – by design (i.e., via the TAM) – tends to select higher-income individuals as winners. While the primary objective of the tax lottery is not to engage in income redistribution (but rather to incentivise the transition from cash to electronic payments), the tax authority and the Ministry of Finance (who together are formally in charge of the lottery design) nevertheless responded to this observation.[32] In an attempt to generate more lower-income winners, they reformed the ticket-awarding mechanism in May 2019. The initial TAM (from Table 1) was replaced by a more concave ticket scheme, displayed in Table 4. Monthly e-transactions up to € 200 (pre-reform: € 100) would be now one-to-one converted into tickets. For e-transaction volumes above € 1,000, an addition Euro would only yield 1/6 of a ticket (pre-reform: 1/4). Moreover, the new TAM introduced an upper bound with a maximum of 8,682 tickets in a given month (which is, in the new scheme, equivalent to e-transactions of € 50,000). Below I explore the potential impact of this reform for different taxpayer's propensity to win the lottery.

[32] Following en bloc draws in December 2017 (see Section 2), some Media reported on cases where individuals won more than once. See for example, `https://www.dimokratianews.gr/dimokratia/forolotaria-me-symptoseis-adianoites/` (in Greek).

Table 4 Post-Reform Ticket-Awarding Mechanism (May 2019 onward)

Total amount of monthly e-transactions	Tickets awarded	Maximum number of tickets
€ 1 – 200	1 ticket per € 1	200
€ 201 – 500	1 ticket per € 2	350
€ 501 – 1,000	1 ticket per € 3	516
> € 1,000	1 ticket per € 6	8,682

5.1 Comparing Pre- and Post-Reform TAM

Based on the distribution of 2017 e-transactions in the general population, I simulate who would have won under the new TAM.[33] By using the 2017 distribution of e-transactions, this approach ignores potential behavioral responses to the lottery's post-reform TAM. In a non-static framework behavioural adjustments could have a non-trivial impact on the post-reform distribution of e-transactions and on the results presented below.[34] Starting from the data, I first transform the monthly e-transactions of the population (from 2017) into lottery tickets – using both pre- and post-reform TAMs. I then simulate (based on 100 iterations) the 1,000 lottery winners of the 12 lottery draws in a calendar year (this ensures that the simulation results take into consideration all months and are not affected by specific spending patterns in particular month). For each of the 1,200 iterations, I record the winners' characteristics and then compare the 1.2 million simulated (12 months $\times$ 100 iterations $\times$ 1,000) winners under the initial TAM with the winners from the post-reform TAM.

The main results from the simulations are shown in Table 5. Winners' selection in pre-reform TAM is comparable to winners' selection in the actual lottery; both the mean and standard deviation values of annual e-transactions and income are similar to the respective values in Table 2.[35] In the simulations one can observe the results of the occupational premium of the policy; higher mean e-transactions than mean income prevails among winners, € 20,794 against € 13,870; lower median than mean values in both income and e-transactions; and a very high standard deviation value for annual e-transactions at € 227,467. The latter indicates the existence of extreme values, confirming that the pre-reform TAM selects individuals with high e-transaction volumes in the population.

[33] Recall that the data available do not no cover the post-reform period.

[34] The upper ceiling of monthly tickets, for instance, creates an incentive to 'smooth' e-transactions above € 50,000 over different months. However, the data from the pre-reform period do not indicate any bunching of monthly e-transactions at the kink-points of the initial TAM. This null-observation might be due to the limited salience of monthly e-payment volumes.

[35] These values are not precisely the same because the actual lottery is similar to only one iteration (1000 winners) for each month in 2017. The precision in the simulation increases, since the lottery is performed 100 times each month, and thus, converging to more precise values.

Table 5 Main Simulation Statistics

	Pre-reform TAM				Post-reform TAM			
	Mean	p10	p50	p90	Mean	p10	p50	p90
Annual E-transactions	20,794	1,000	4,890	16,730	10,480	990	4,720	14,980
	(227,467)				(69,923)			
Annual Income	13,870	0	11,225	25,370	13,476	0	11,190	24,695
	(23,226)				(21,495)			

Notes: The table presents the main statistics from lottery simulations using the pre-reform TAM (left-hand side) and post-reform TAM (right-hand side). Each simulation aggregates 1,200,000 observations of winners, drawn based tickets assigned from each structure (100 lottery iterations, drawing 1,000 winners in each iteration, for each of the 12 months in 2017). The first column in each TAM presents the mean values and standard deviation in parentheses. The median values are presented in the "p50" columns, together with the lowest and highest percentiles in "p10" and "p90" respectively.

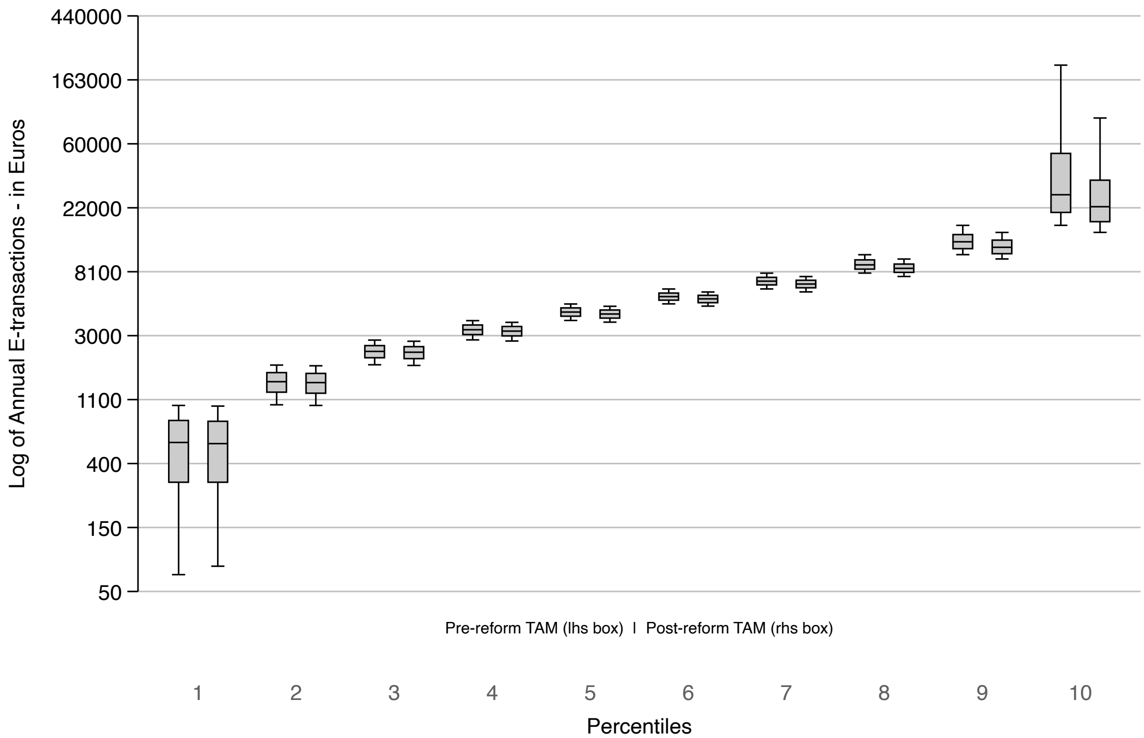
Importantly, the results in Table 5 indicate a change in the winners population once the post-reform TAM is applied. This change works predominantly through the € 50,000 limit in monthly e-transactions. The annual mean (standard deviation) e-transactions for winners drops from € 20,794 (227,467) to € 10,480 (69,923), while the mean declared income is reduced only marginally, indicating that the upper ceiling can limit the distortion of winners exhibiting higher e-transactions than their declared income. Mean annual e-transactions for the 9th decile reduce accordingly from € 16,730 to € 14,980.[36] Moreover, while decreasing e-transactions in the highest decile are non-trivial, the post-reform TAM produces a lower e-transactions level for low-to-middle deciles. This point is illustrated in Figure 6, which presents box plots of the log of e-transactions for every decile in pre- and post-reform TAM. Deciles in the middle of the distribution present small reductions in log e-transactions, while the largest reduction happens at the highest decile. An overall comparison between the income and e-transactions distributions can be seen in Figures 13 and 14, respectively.

Finally, I provide a measure to assess the extent to which the reform can limit the distortion of winners exhibiting higher e-transactions volumes than their declared income. I plot the e-transactions distribution curves for each TAM (on the y-axis), ranking individuals by their annual reported incomes (on the x-axis) in Figure 7. This produces an e-transactions distribution of winners, where if the slope is higher [lower] than the 45-degree line, winners at a particular decile exhibit proportionally more [less] e-transactions than income.[37]

[36]Notice also that the lowest decile exhibits 0 income, while having positive annual e-transactions. This group consists of individuals who declare zero income and happen to win the lottery.

[37]If all slopes at all deciles are equal to 45-degree, this serves as a point of equality, where the percentage of e-transactions for each decile equals their percentage in the income distribution.

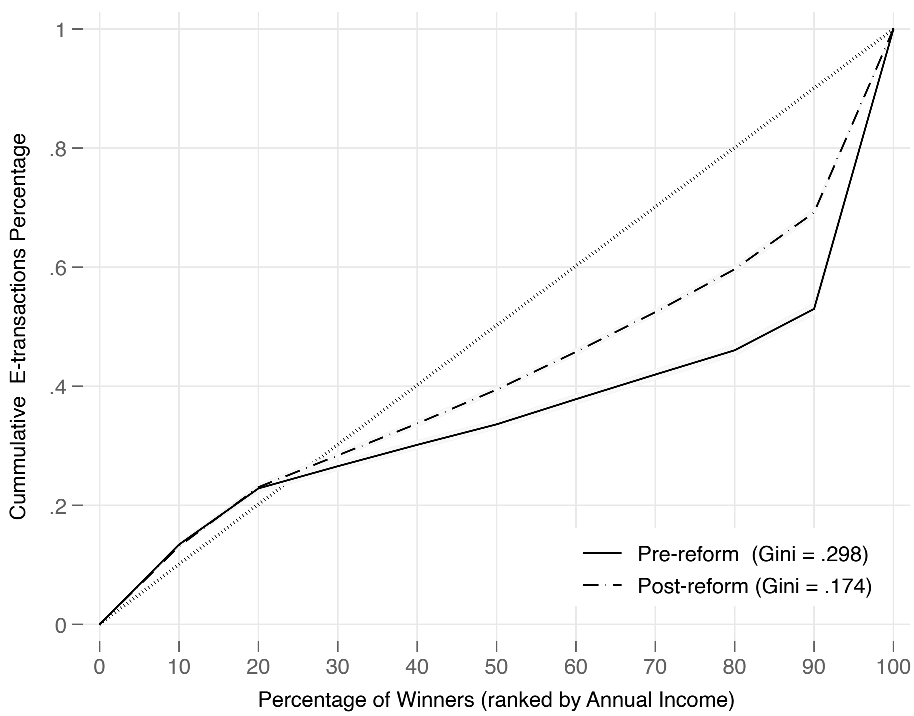
Fig. 6 E-transaction Box Plots



Notes: The figure presents box plots of the log-value of annual e-transactions for each decile in the pre- and post-reform TAM simulations. The pre-reform TAM are presented on the left-hand side of each decile and the post-reform TAM on the right-hand side. Each simulation contains 1.2 million observations of winners. The plots exclude extreme values, which are present in 1st and, in particular, the 10th deciles. The y-axis is drawn in the log value of annual e-transactions but has the values have been computed to Euros, rounded up to the nearest tenth for values of up to € 8100, and to the nearest thousand for higher values.

First, note that the lowest quintile of the population (according to their annual e-transactions) covers more than 20% of winners. This property, which holds for both the initial and the new TAM, seems to reflect the TAM's concavity, favoring lower income households who exhibit proportionately higher monthly e-transactions than their declared income (for example, by being net welfare benefit receivers). The slope is higher than the 45-degree line in the first decile, while it is close to 45-degrees for the second decile, indicating that the lowest benefits the most in both TAMs. Second, for the pre-reform TAM, the top decile in the income distribution accounted for 47% of all e-transactions. With the post-reform TAM, this percentage drops to 34%. The observed decline originates from a decreased slope in the highest decile, as well as, increased slopes in lower deciles. The change suggests that the post-reform TAM – in particular, it's upper ticket limit – is partly effective in reducing the winning chances associated with extreme e-transactions volumes. Yet, high income individuals still exhibit higher chances of winning. Third, the reform-beneficiaries are located between the 2nd and the 9th decile of the e-transaction distribution (higher slopes in the post-reform, than in the pre-reform TAM). Overall, the Gini coefficient associated with the e-transaction curve, when ranked by income, can serve as measure of this distortion; it falls from 0.30 to 0.17. Hence, the reform had a non-trivial effect on achieving a more equal distribution of prizes.

Fig. 7 E-transactions Distribution



Notes: The figure plots e-transaction distribution curves for 2 × 1.2 million winners in Monte Carlo Simulations of the lottery. The x-axis represents the population percentiles of winners, ranked by annual income in 2017. The y-axis plots the cumulative percentage of e-transactions resulting from simulated lottery winners in pre- and post-reform TAM. The dotted line is a 45-degree line, at which the winners' income percentage equals the winners' e-transactions percentage. For the pre- [post-]reform curve, tickets were assigned based on the initial [post-reform] TAM presented in Table 1 [Table 4].

5.2 Other Reforms

Following the simulation approach from above, I also consider the impact of more radical upper limits in the monthly amount of tickets; at € 1,000 and € 5,000, respectively.[38] I find that both ceilings are effective in limiting the winning chances associated with very high monthly e-transactions. As shown in Table 6, the e-transactions of the highest decile fall to € 15,350 at the € 5,000 ceiling and to € 13,550 at the € 1,000 ceiling. As in the pre-/post-reform TAM the annual income distribution of winners changes only marginally. Changes to the distribution of income and e-transactions are illustrated over the whole population of winners in Figures 15 and 16 respectively.

[38]The simulations adopt the euro-to-ticket structure from the pre-reform TAM. The ceilings thus limit the maximum number of monthly tickets at 467 and 1,467, respectively.

Table 6 Main Simulation Statistics - Stricter Ticket Ceilings

	Ceiling € 1,000				Ceiling € 5,000			
	Mean	p10	p50	p90	Mean	p10	p50	p90
Annual E-transactions	6,994	970	4,630	13,550	8,316	990	4,800	15,350
	(24,783)				(33,949)			
Annual Income	13,088	0	11,210	24,160	13,553	0	11,225	24,950
	(19,307)				(21,259)			

Notes: The table presents the main statistics from lottery simulations using a € 1,000 monthly ticket ceiling per individual (left-hand side) and a corresponding € 5,000 ceiling (right-hand side). Each simulation aggregates 1,200,000 observations of winners, drawn based tickets assigned from each structure (100 lottery iterations, drawing 1,000 winners in each iteration, for each of the 12 months in 2017). The first column in each TAM presents the mean values and standard deviation in parentheses. The median values are presented in the "p50" columns, together with the lowest and highest percentiles in "p10" and "p90" respectively.

By drawing e-transaction distribution curves over the income distribution of winners in Figure 17, one can investigate the beneficiaries of such reforms. The stricter ceilings reduce even further the chances of winning for the highest decile of the winners' income distribution (the slope at the top decile approaches the slope of the 45-degree line), while individuals in the 2nd to the 9th decile stand to benefit. The Gini coefficients fall further to 0.162 for the € 1,000 ceiling and to 0.148 for the € 5,000 ceiling. The latter records a more equal distribution than the former, because, when the € 1,000 ceiling is used, a much higher fraction of individuals ends up receiving the maximum amount of tickets in several months. (see Figures 18 and 19). As the very strict ceiling of € 1,000 becomes more binding in the population, the chances of winning become more detached from the individuals' monthly level of e-transactions. Overall, the € 5,000 monthly ceiling seems to strike a better balance between, on one hand, limiting the distortion caused by high e-transaction volumes and, on the other, achieving a fairer distribution of prizes in the population.

6 Conclusion

This paper documents how certain design characteristics of the Greek tax lottery (a scheme that provides monetary rewards to consumers when they ask for payment receipts) assign higher winning probabilities to certain groups of individuals against others. Using a reconstructed taxpayer population in Greece and by comparing their income and annual electronic spending against those of winners, I document an "income premium" and an "occupational premium" in lottery winners.

When it was introduced in 2017, the Greek tax lottery selected winners from high income categories, who spent more in electronic consumption and were awarded with more tickets than lower income categories. The probability of winning increased by 0.11% in response to a 10% increase in income. The lottery exhibited also an "occupational premium", selecting winners from the self-employed/business-income category more frequently. Being self-employed increased the chances of winning by 18% compared to wage-earners, pensioners and agricultural workers, after controlling for individual and household income. This might have resulted from business transactions channeled through personal bank accounts, or from under-declaration of self-employed income. In addition, intra-household income and occupational effects are economically and statistically significant. The probability of winning increases by 0.7% in response to a spousal income increase by 10%; it increases further by 3.5% when the spouse is self-employed.

These results have important implications for tax lotteries. Since tax lotteries provide incentives to the final consumer to ask for a receipt at the point of purchase, it is high-income/high-spending consumers who stand to benefit the most from the lottery's monetary rewards. In addition, spending for business purposes cannot be distinguished easily from personal spending, resulting in some occupational categories winning more frequently than others. These effects are amplified by intra-household effects, which move in the same direction as idiosyncratic effects. Both occupational and income premiums might dampen the effectiveness of the policy. Firstly, if the policy is perceived as unfair it might discourage individuals from participating. Secondly, by awarding winning prizes to high income individuals makes the policy regressive by construction. Prizes in low income individuals might act as a stronger incentive mechanism due to a higher prize-to-income ratio. Thirdly, the policy becomes less salient in low income individuals who experience winning less often than high income individual, but who might be using electronic payments less frequently than high-income individuals.

To mitigate the lottery's shortcomings and limit ticket premiums in certain groups of the population, the analysis in this paper considered a reform of the lottery in 2019. A more concave ticket-to-euro structure and a limit of € 50,000 per individual per month in the amount of tickets awarded were introduced. Using Monte Carlo simulations in a static framework, tax lottery draws were performed to determine the policy's effect. The results establish that the € 50,000 ticket ceiling is effective in improving fairness by limiting extreme consumption values for high spending individuals, whilst the more concave ticket-to-euro structure has only a marginal effect. In the absence of a ceiling high-income/high-spending individuals would have had higher chances of winning. In a second step, I considered stricter limits that suggest a € 5,000 limit in the amount of tickets awarded, per month per individual, would have resulted in an even fairer distribution of prizes, without placing excessive limits on the incentivisation of taxpayers to ask for receipts.

The evidence in this paper provide guidance for tax lottery design. As regressivity seems to be a characteristic of this policy, inherent by construction and difficult to predict prior to the policy's implementation, placing a ticket ceiling will benefit a tax lottery's effectiveness. The ceiling can limit excess ticket premiums to certain groups of individuals and thereby improve a tax lottery effectiveness.

References

Browning, M., Bourguignon, F., Chiappori, P.-A. and Lechene, V. (1994). Income and Outcomes: A Structural Model of Intrahousehold Allocation. *Journal of Political Economy*, **102** (6), 1067–1096.

Carroll, C. D., Slacalek, J. and Tokuoka, K. (2014). The Distribution of Wealth and the MPC: Implications of New European Data. *The American Economic Review*, **104** (5), 107–111.

Chiappori, P.-A., Giménez-Nadal, J. I., Molina, J. A. and Velilla, J. (2020). *Household Labor Supply: Collective Evidence in Developed Countries*, Cham: Springer International Publishing, pp. 1–28.

Danchev, S., Gatopoulos, G. and Vettas, N. (2020). Penetration of Digital Payments in Greece after Capital Controls: Determinants and Impact on VAT Revenues. *CESifo Economic Studies*, **66** (3), 198–220.

Fooken, J., Hemmelgarn, T. and Herrmann, B. (2015). Improving VAT Compliance – Random Awards for Tax Compliance, Directorate General Taxation and Customs Union, European Commission, Taxation Paper No. 51.

Hondroyiannis, G. and Papaoikonomou, D. (2017). The effect of card payments on VAT revenue: New evidence from Greece. *Economics Letters*, **157**, 17 – 20.

Kleven, H. J., Knudsen, M., Kreiner, C., Pedersen, S. and Saez, E. (2011). Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark. *Econometrica*, **79** (3), 651–692.

Lundberg, S., Pollak, R. and Wales, T. (1997). Do Husbands and Wives Pool Their Resources? Evidence from the United Kingdom Child Benefit. *Journal of Human Resources*, **32** (3), 463–480.

Naritomi, J. (2019). Consumers as Tax Auditors. *American Economic Review*, **109** (9), 3031–3072.

Nicolaides, P. (2022). Income Tax Incentives for Electronic Payments: Evidence from Greece's Electronic Consumption Tax Discount, Working Paper.

— (2023). Winning the Tax Lottery: Evidence from a Superdraw on Christmas Eve, Working Paper.

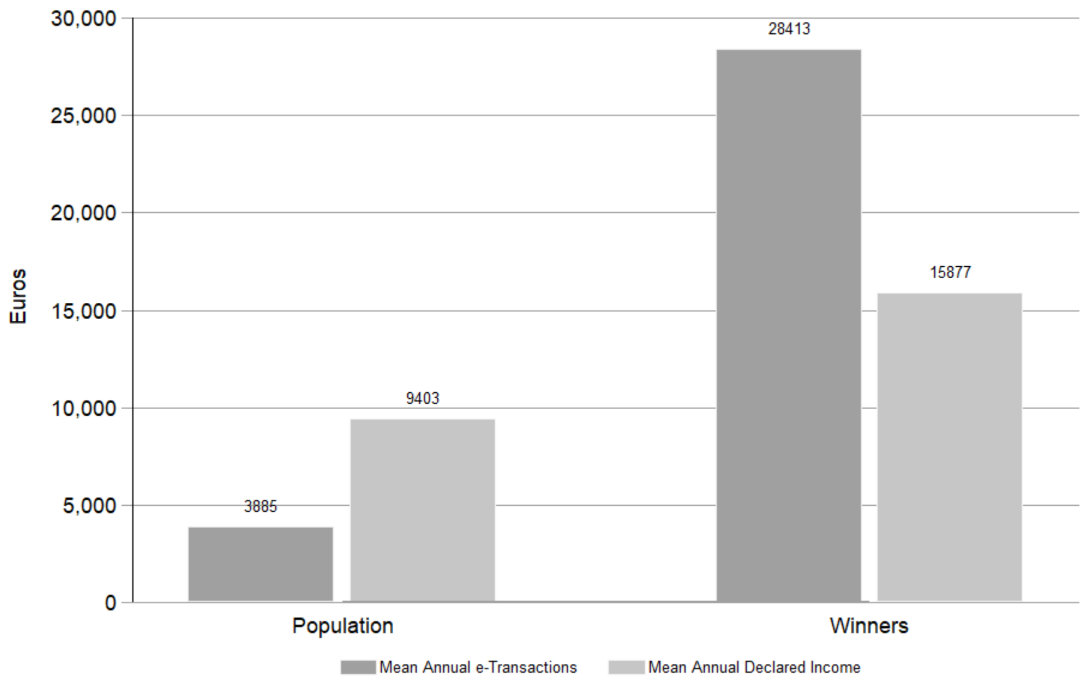
Pomeranz, D. (2015). No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax. *American Economic Review*, **105** (8), 2539–69.

Quisumbing, A. R. and Maluccio, J. A. (2003). Resources at Marriage and Intrahousehold Allocation: Evidence from Bangladesh, Ethiopia, Indonesia, and South Africa. *Oxford Bulletin of Economics and Statistics*, **65** (3), 283–327.

A Appendix

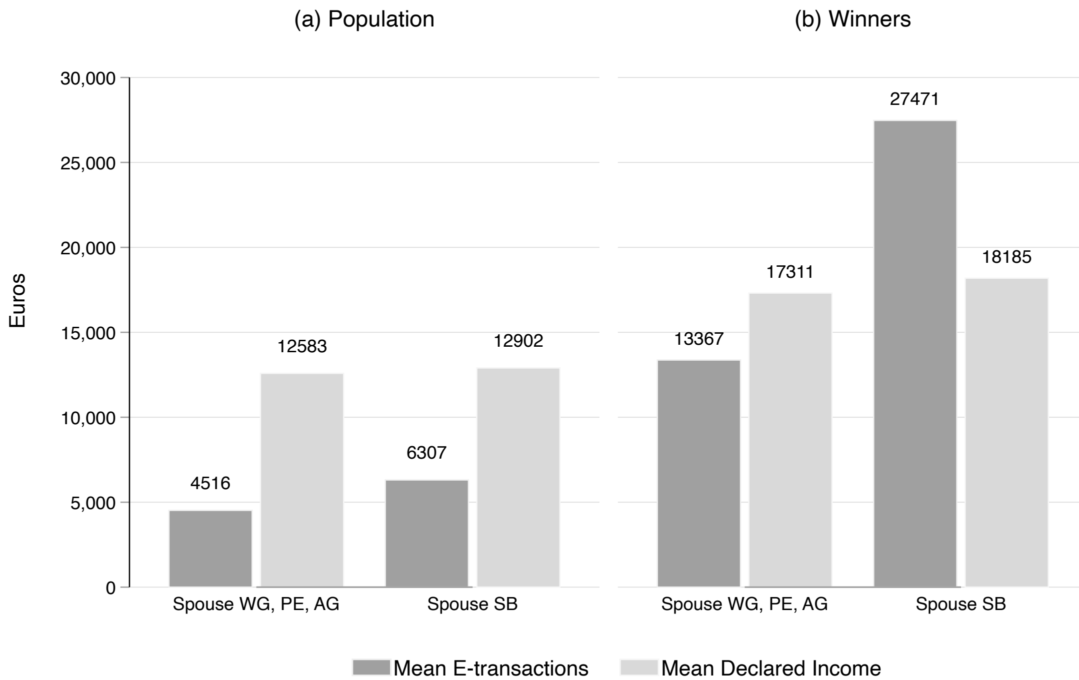
A.1 Complementary Figures and Tables

Fig. 8 Mean Annual E-transactions vs. Annual Declared Income



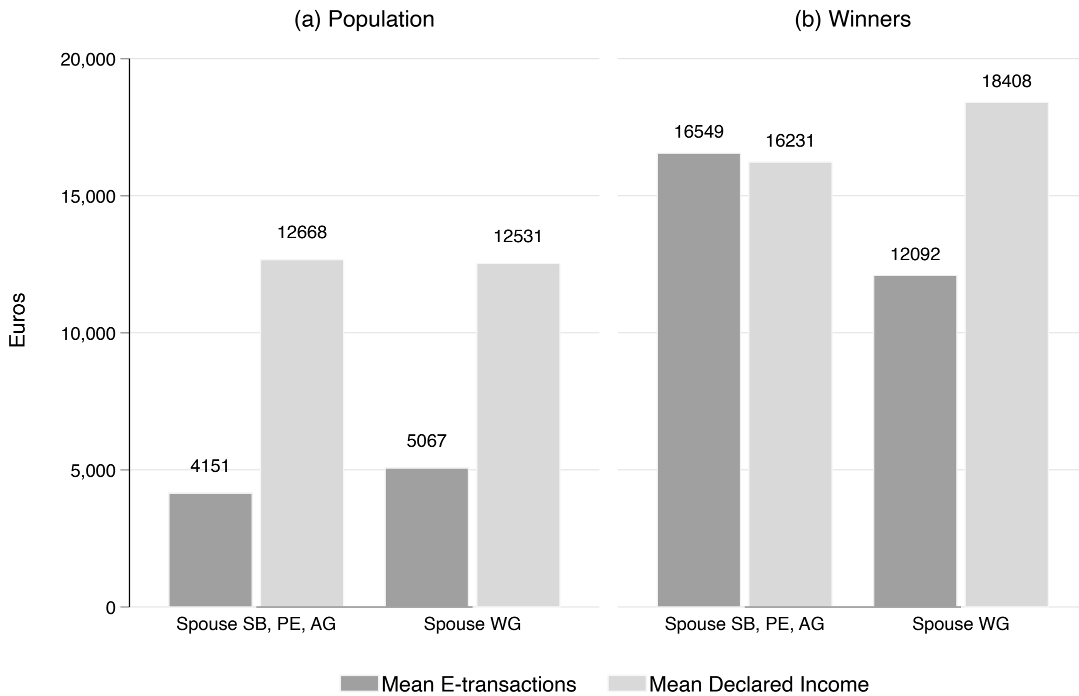
Notes: The figure compares the annual mean electronic transactions against the mean annual declared income of winners and of the representative taxpayer population. Only winners from lotteries in 2017 are included in the winners' sample. Non-parametric estimates of the differences are provided in Table 8

Fig. 9 Declared Annual Income and E-transactions, Taxpayers with SB spouses



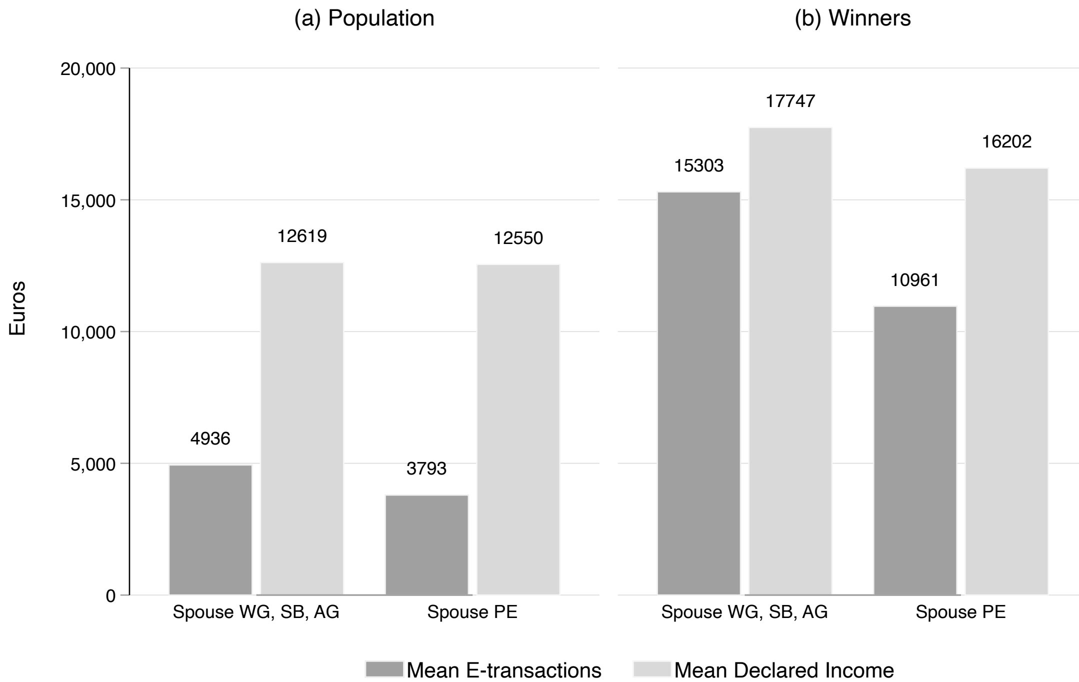
Notes: The figure compares the mean annual declared income and mean annual e-transactions of the taxpayer population, figure (a) on the left-hand side and for winners, figure (b) on the right-hand side. The left-hand side columns of each figure include individuals who have a spouse with primary income from WG, PE and AG, against individuals who have a spouse with primary income from SB. Individuals with primary SB income and NO income are excluded from the sample.

Fig. 10 Declared Annual Income and E-transactions, Taxpayers with WG spouses



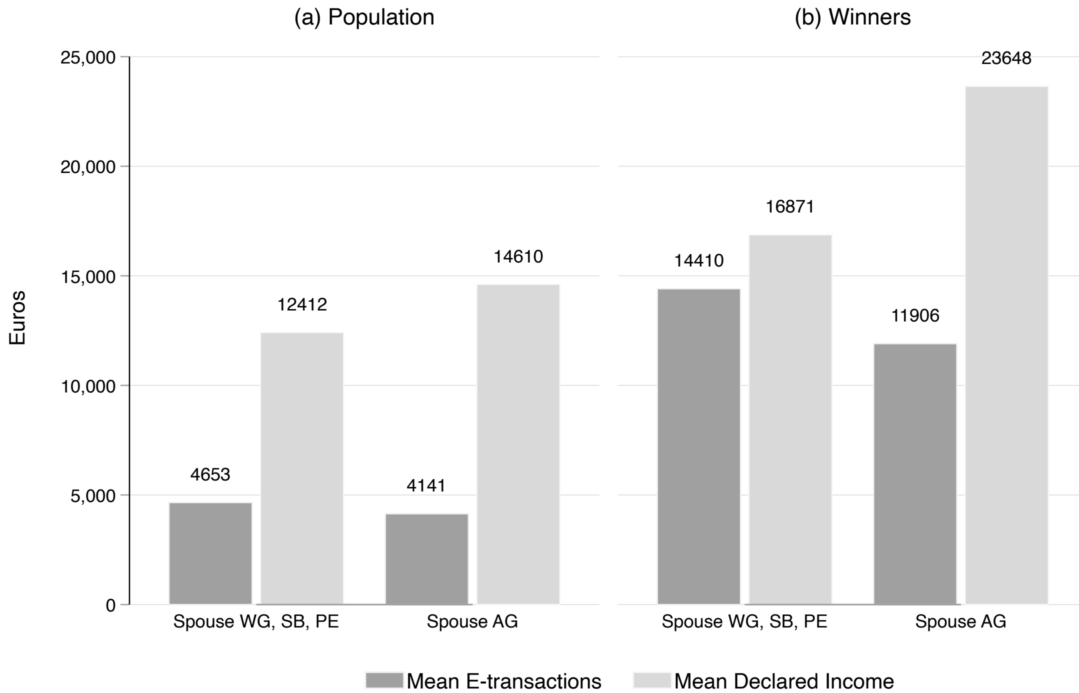
Notes: The figure compares the mean annual declared income and mean annual e-transactions of the taxpayer population, figure (a) on the left-hand side and for winners, figure (b) on the right-hand side. The left-hand side columns of each figure include individuals who have a spouse with primary income from SB, PE and AG, against individuals who have a spouse with primary income from WG. Individuals with primary SB income and NO income are excluded from the sample.

Fig. 11 Declared Annual Income and E-transactions, Taxpayers with PE spouses



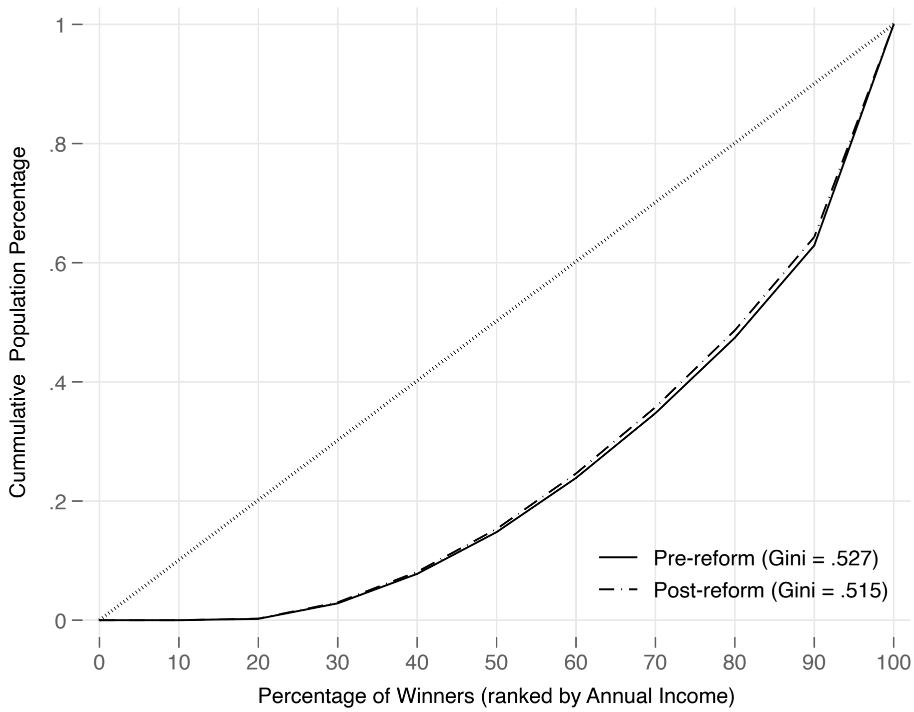
Notes: The figure compares the mean annual declared income and mean annual e-transactions of the taxpayer population, figure (a) on the left-hand side and for winners, figure (b) on the right-hand side. The left-hand side columns of each figure include individuals who have a spouse with primary income from WG, SB and AG, against individuals who have a spouse with primary income from PE. Individuals with primary SB income and NO income are excluded from the sample.

Fig. 12 Declared Annual Income and E-transactions, Taxpayers with AG spouses



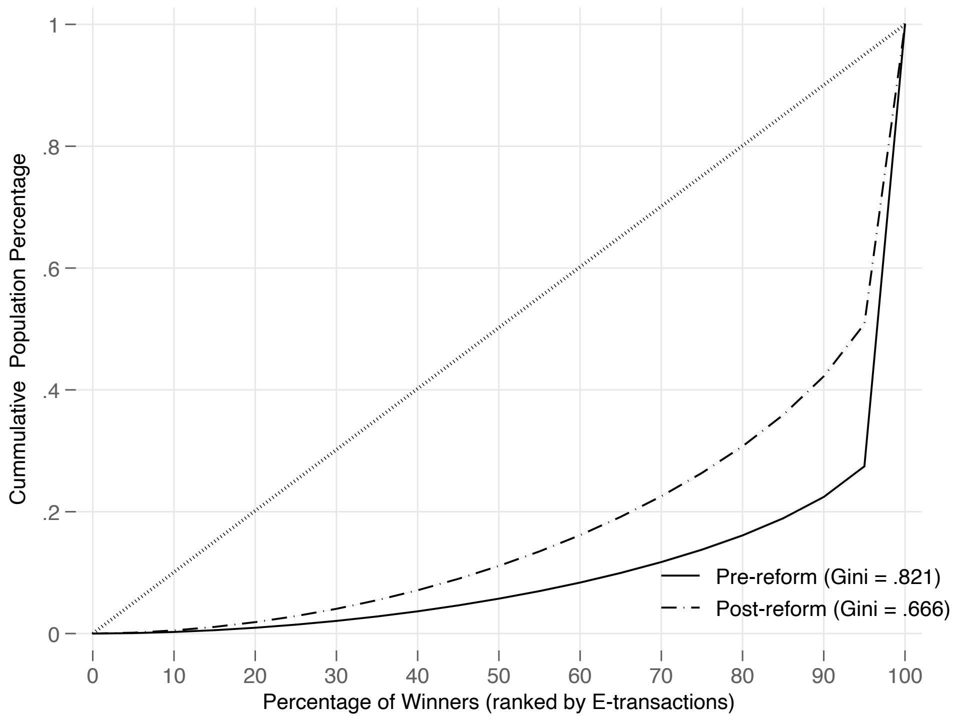
Notes: The figure compares the mean annual declared income and mean annual e-transactions of the taxpayer population, figure (a) on the left-hand side and for winners, figure (b) on the right-hand side. The left-hand side columns of each figure include individuals who have a spouse with primary income from SB, PE and WG, against individuals who have a spouse with primary income from AG. To allow for a meaningful comparison, SB individuals are excluded from the sample since these have exhibited a very high volume of e-transactions as shown in Fig. 5. NO income category and single filings are excluded from the sample.

Fig. 13 Income Distribution



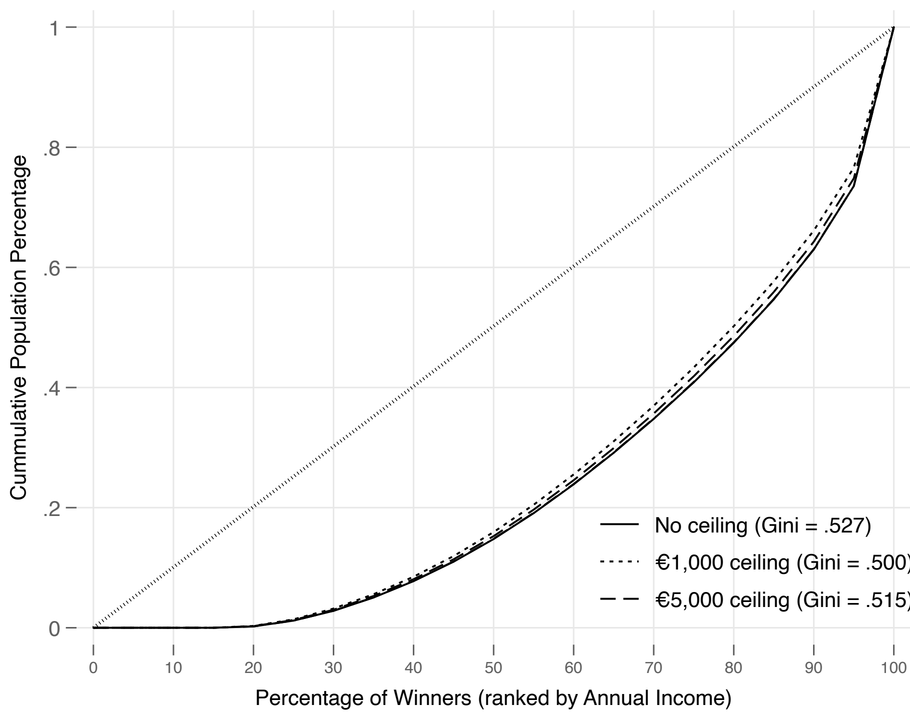
Notes: The figure plots e-transaction distribution curves for 2 $\times$ 1.2 million winners in Monte Carlo Simulations of the lottery. The x-axis represents the population percentiles of winners, ranked by annual income in 2017. The y-axis shows the percentage of individuals who have won the lottery in the simulations. The dotted line is a 45-degree line, at which the population percentage equals the winners percentage in the distribution. For the pre- [post-]reform curve, tickets were assigned based on the initial [post-reform] TAM presented in Table 1 [Table 4].

Fig. 14 E-transactions Distribution



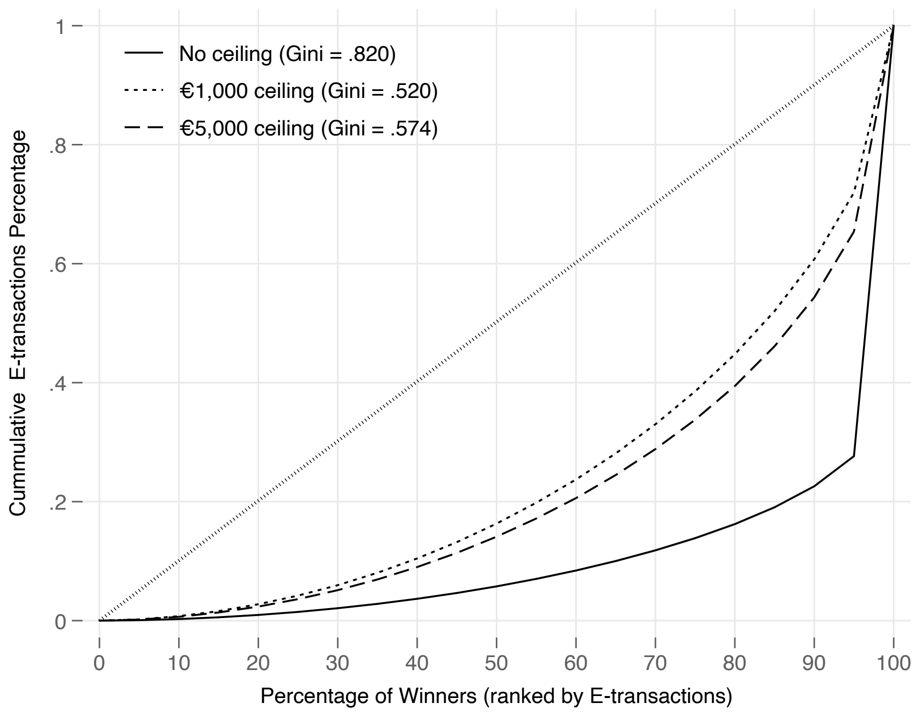
Notes: The figure plots e-transaction distribution curves for 2 × 1.2 million winners in Monte Carlo Simulations of the lottery. The x-axis represents the population percentiles of winners, ranked by annual e-transactions in 2017. The y-axis shows the percentage of individuals who have won the lottery in the simulations. The dotted line is a 45-degree line, at which the population percentage equals the winners percentage in the distribution. For the pre- [post-]reform curve, tickets were assigned based on the initial [post-reform] TAM presented in Table 1 [Table 4].

Fig. 15 Income Distribution for Stricter Ceilings



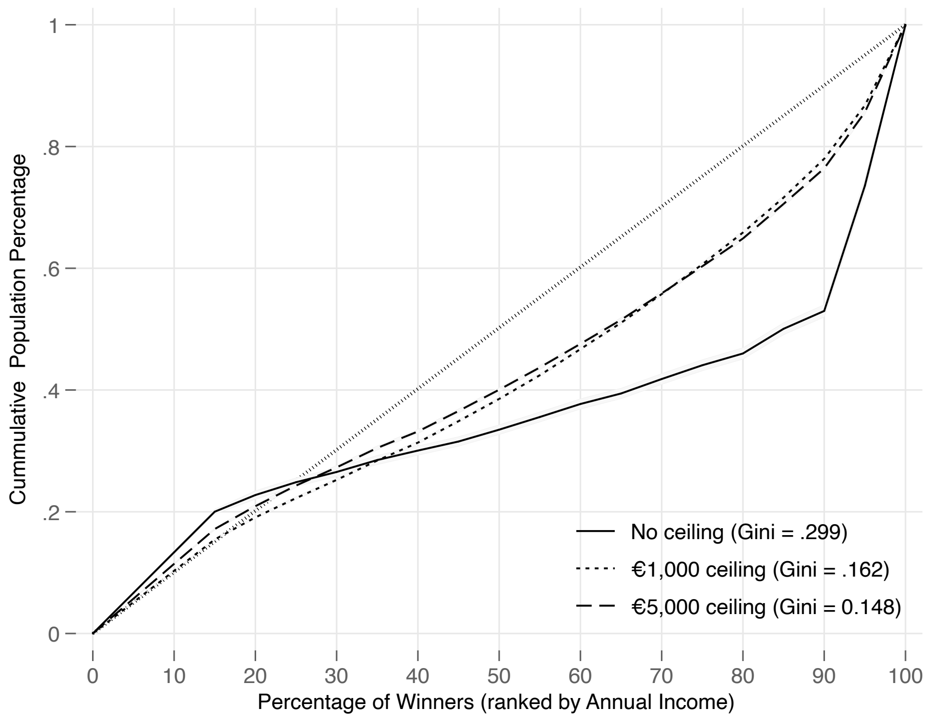
Notes: The figure plots income distribution curves for 1.2 million winners in Monte Carlo Simulations of the lottery. The x-axis represents the population percentiles of winners, ranked by their declared annual income in 2017. The y-axis shows the percentage of individuals who have won the lottery in the simulations. The dotted line is a 45-degree line, at which the population percentage equals the winners percentage in the distribution. The "no ceiling" curve is a simulation of the lottery assigning tickets using the pre-reform TAM in Table 1. The "€ 1,000 ceiling" curve retains the pre-reform TAM characteristics, but introduces a maximum ceiling in monthly tickets. For e-transactions beyond € 1,000 per month no more tickets are awarded to individuals.Similarly, the € 5,000 curve, retains the characteristics of pre-reform TAM, but introduces a ceiling at the € 5,000 monthly e-transaction level.

Fig. 16 E-transactions Distribution for Stricter Ceilings



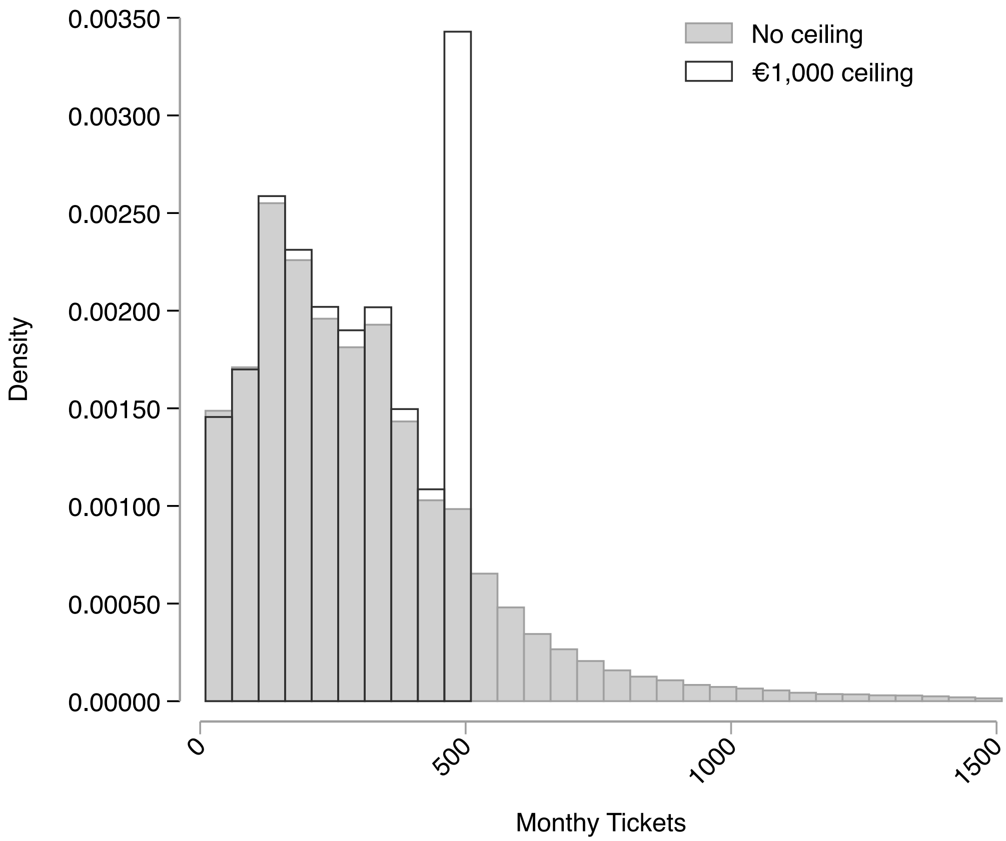
Notes: The figure plots e-transactions distribution curves for 1.2 million winners in Monte Carlo Simulations of the lottery. The x-axis represents the population percentiles of winners, ranked by e-transactions in 2017. The y-axis plots the cumulative percentage of e-transactions resulting from simulated lottery winners in pre- and post-reform TAM. The y-axis shows the percentage of individuals who have won the lottery in the simulations. The dotted line is a 45-degree line, at which the population percentage equals the winners percentage in the distribution. The "no ceiling" curve is a simulation of the lottery assigning tickets using the pre-reform TAM in Table 1. The "€ 1,000 ceiling" curve retains the pre-reform TAM characteristics, but introduces a maximum ceiling in monthly tickets. For e-transactions beyond € 1,000 per month no more tickets are awarded to individuals. Similarly, the € 5,000 curve, retains the characteristics of pre-reform TAM, but introduces a ceiling at the € 5,000 monthly e-transaction level.

Fig. 17 E-transactions Distribution (ranked by Income) for Stricter Ceilings



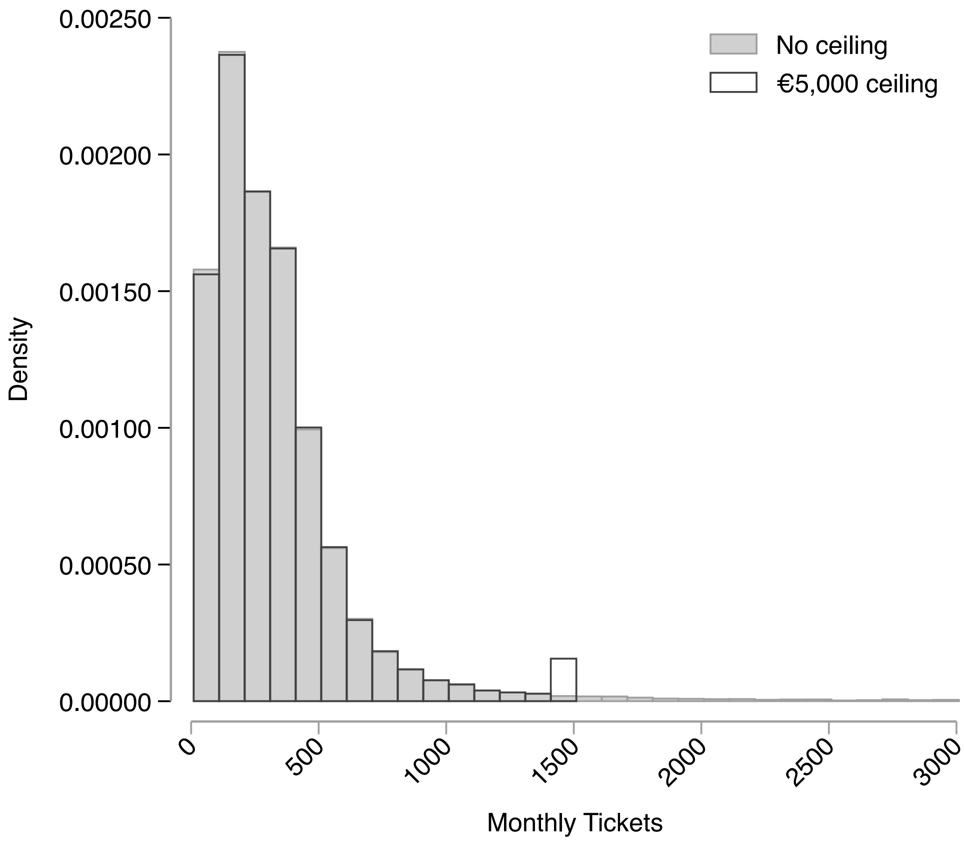
Notes: The figure plots e-transactions distribution curves for 1.2 million winners in Monte Carlo Simulations of the lottery. The x-axis represents the population percentiles of winners, ranked by annual income in 2017. The y-axis plots the cumulative percentage of e-transactions resulting from simulated lottery winners in pre- and post-reform TAM. The dotted line is a 45-degree line, at which the winners' income percentage equals the winners' e-transactions percentage. The "no ceiling" curve is a simulation of the lottery assigning tickets using the pre-reform TAM in Table 1. The "€ 1,000 ceiling" curve retains the pre-reform TAM characteristics, but introduces a maximum ceiling in monthly tickets. For e-transactions beyond € 1,000 per month no more tickets are awarded to individuals. Similarly, the € 5,000 curve, retains the characteristics of pre-reform TAM, but introduces a ceiling at the € 5,000 monthly e-transaction level.

Fig. 18 Winner's Distribution of Tickets in Simulations



Notes: The figure compares the distribution of winners' tickets assigned in Monte Carlo Simulations and the effect on tickets by placing a maximum ticket ceiling at € 1,000. This translates to a maximum number of 467 monthly tickets. Both simulations contain 1,200 iterations of the lottery (100 for each month in 2017), drawing 1,000 winners in each iteration. Both distributions contain 1.2 million winners. For the "No ceiling" distribution tickets were assigned using the the pre-reform TAM in Table 1. The "€ 1,000 ceiling" distribution retains the pre-reform TAM characteristics, but introduces a maximum ceiling in monthly tickets. For e-transactions beyond € 1,000 per month no more tickets are awarded to individuals. The distributions is truncated at 1,500 tickets, as right-tails diminish quickly in the distribution beyond this point.

Fig. 19 Winner's Distribution of Tickets in Simulations



Notes: The figure compares the distribution of winners' tickets assigned in Monte Carlo Simulations and the effect on tickets by placing a maximum ticket ceiling at € 5,000. This translates to a maximum number of 1,467 monthly tickets. Both simulations contain 1,200 iterations of the lottery (100 for each month in 2017), drawing 1,000 winners in each iteration. Both distributions contain 1.2 million winners. For the "No ceiling" distribution tickets were assigned using the the pre-reform TAM in Table 1. The "€ 5,000 ceiling" distribution retains the pre-reform TAM characteristics, but introduces a maximum ceiling in monthly tickets. For e-transactions beyond € 1,000 per month no more tickets are awarded to individuals. The distributions are truncated at 3,000 tickets, as right-tails diminish quickly in the distribution beyond this point.

Table 7 Basic Summary Statistics

	Samples:		Single/Joint Filing:	
	Non-Winners Freq (Percent)	Winners Freq (Percent)	Single Filers Freq (Percent)	Joint Filers Freq (Percent)
By Primary Income Source:				
SB : Self-Employed/Business Income	2,052	1,609	1,855	1,806
	(4.10)	(8.52)	(4.46)	(6.61)
WG : Wage-Earner	22,335	9,107	17,205	14,237
	(44.67)	(48.19)	(41.38)	(52.11)
PE : Pensions (Main and Auxiliary)	12,163	4,201	8,979	7,385
	(24.33)	(22.23)	(21.60)	(27.03)
AG : Agriculture	2,635	831	1,463	2,003
	(5.27)	(4.40)	(3.52)	(7.33)
NO : Zero-declared Income	10,815	2,861	12,072	1,604
	(21.63)	(15.14)	(29.04)	(5.87)
No Filing : Tax return not submitted	-	288	-	288
	-	(1.52)	-	(1.05)
Total	50000	18897	41574	27323

Notes: The table presents basic summary statistics for the tax unit samples, per income source category. The left-hand side columns present the number of observations and percentages (in parentheses), of the non-winners and winners samples in the tax lottery. The winners sample includes winners in 19 consecutive months, from January 2017 to July 2018. The non-winners sample has been randomly drawn from the population of tax units. The right-hand side columns present the frequencies and percentages of single and joint-filing in each primary income source category. Joint-filers can be indirectly deduced from the sample. based on annual declared income from both spouses in a household. The case where the main taxpayer declares positive income and the spouse zero income cannot be identified in the sample.

Table 8 Non-parametric Estimates, by Primary Income Category

	(1)	(2)	(3)	(4)	(5)	(6)
Annual Income	SB	WG	PE	AG	NO	No Filing
Winner in 2017	8,665***	6,952***	2,760***	29,884***	0	0
	(1,192)	(518)	(134)	(4,807)	(0)	(0)
Constant	12,088***	11,404***	11,870***	17,538***	0	0
	(568)	(92)	(55)	(733)	(0)	(0)
Annual E-transactions						
Winner in 2017	170,733***	6,807***	7,655***	11,732***	24,715***	53,286
	(22,125)	(601)	(1,305)	(1,491)	(4,628)	(33,389)
Constant	10,786***	4,050***	3,309***	3,800***	2,903***	4,925***
	(860)	(37)	(41)	(126)	(125)	(1,137)
Observations	266,317	2,890,322	1,573,228	340,746	1,397,996	288

Notes: The table presents estimation results per primary income source. Results on top of the table use annual declared income as independent variable and at the bottom, annual electronic transactions. The NO and No Filing categories in columns (5) and (6) do not record results for annual income regressions, since no income was declared. Annual e-transactions are observed for the latter categories. Robust standard errors (clustered at the individual level, depending on the number of individuals in each income source category) in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 9 Estimates - Spouse's Primary Income Source

	(1) e-trans	(2) e-trans	(3) e-trans	(4) e-trans	(5) e-trans	(6) e-trans
Spouse's Primary Income SB	2,006***		1,778***	1,781***	1,791***	1,747***
	(317)		(320)	(256)	(268)	(256)
Spouse's Any Income from SB		2,321***				
		(357)				
Annual Income in 2017			0.153***	0.139***		0.138***
			(0.027)	(0.021)		(0.021)
Winner in 2017					8,851***	8,196***
					(1,369)	(1,374)
Winner in 2017 & Spouse SB					12,313**	12,236**
					(6,088)	(6,081)
Constant	4,970***	4,943***	3,049***	2,788***	4,516***	2,774***
	(119)	(119)	(328)	(266)	(79)	(265)

Notes: The table presents estimation results for individuals with SB spouses. The sample is restricted to 2,406,971 individuals in the population who filed jointly in regressions (1) - (3). Observations are restricted to 2,279,469 in (4) - (6), which include joint-filers, but exclude individuals who declared SB as their primary income source. Robust standard errors (clustered at the level of 27,323 and 25,517 unique taxpayers for regressions (1) - (3) and (4) - (6) respectively) in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

A.2 Construction of the Baseline Population of Tax Units

In addition to the information from two samples, I observe the total number of lottery tickets issued in each calendar month, $\bar{T}_m$. Given that lottery tickets are derived from monthly e-transactions via the TAM described in Table 1, I can compute $T_{i,m,s}$, the number of tickets from individual i in month m in sample s, where $s \in \{1, 2\}$ indicates the winner and non-winner sample, respectively. In addition, non-winners in 2017, who were winners in 2018 must be added in the expansion. To avoid a different subscript for the year, I utilise $\hat{T}_{i,m,1}$.

Given this, the following identity must hold:

$$\sum_{m=1}^{12} \bar{T}_m = \sum_{m=1}^{12} \sum_{i=1}^{N_1} T_{i,m,1} + \sum_{m=1}^{12} \sum_{i=1}^{N_1} \hat{T}_{i,m,1} + \omega \sum_{m=1}^{12} \sum_{i=1}^{N_2} T_{i,m,2} \tag{2}$$

where N_s indicates the size of the samples s (with $N_1 = 18,897$ and $N_2 = 50,000$).

From this, it is straightforward to derive ω, the weight or expansion factor that I have to use to arrive at a sample that matches the population in terms of lottery tickets, since it is the only unknown. One can observe the total number of tickets in 2017, $\sum_{m=1}^{12} \bar{T}_m$ and the total number of tickets in the samples.

A further plausibility check is that $N_1 + \omega N_2 \cong N$. The calculation derives ω to be 129. Expanding the random sample gives a total tax unit population of 6.45 million (50,000×129), to which 18,897 winners are added. This is very close to official statistics from the tax authority, indicating 6.37 million tax returns being filed for 2017.[39]

In addition to expanding the population, I have explored other ways that would allow a direct comparison and solve the pre-selection of winners problem. One of this has been to obtain an additional random sample without pre-selecting winners from the population. That would had produced a random sample over the entire population. However, since the number of winners is very small, only very few winners would have been selected, therefore inhibiting a comparison of their characteristics. The random sample drawn in this case would have required to be particularly large in order to reach a point where a large number of winners are selected, to allow for a meaningful comparison. Thus, pre-selecting the winners was essential for providing insights to the lottery. Lastly, the random sample of 50,000 non-winners is large enough as to allow for a good approximation of characteristics to the actual population of taxpayers.

[39] Annual statistics for the 2017 filing are published by the Tax Authority at `https://www.aade.gr/menoy/statistika-deiktes/eisodima/etisia-statistika-deltia`.