A Dividend of Chance: Determinants and Responses of Winners in the Greek Tax Lottery

Panayiotis Nicolaides *

July 4, 2024

Abstract

This paper studies the determinants and responses of winners in the Greek tax lottery. The scheme incentivises consumers to switch from cash to electronic payments, thereby generating third-party information that can increase tax compliance. Tickets map directly to electronic consumption, with 1,000 winners awarded € 1,000 every month. Using administrative tax and consumption data, I reconstruct a representative taxpayer population to examine the characteristics of winners. High-income/high-consumption taxpayers win more frequently: A 10% income increase is associated with a 0.11% increase in winning probability. The self-employed record particularly large amounts of transactions, increasing their winning probability by 0.18% compared to other income categories. Utilising a unique event of retroactive draws in Christmas 2017, I document heterogeneity in the winners' responses in electronic consumption along the income and occupation dimensions. Counter to the scheme's design that links higher winning probabilities to higher spending, I provide evidence of (a) temporary increases in electronic consumption for taxpayers in low-to-middle income quantiles and, non-responsiveness for the highest income taxpayers (b) temporary increases for wage-earners and pensioners and, a non-responsiveness for the self-employed. These results have fairness and efficacy implications, which can be mitigated through a ticket ceiling that limits the winning probabilities for high consumption individuals.

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

Keywords: Tax Lottery; Electronic Payments; Third-Party Information; Value-Added Tax

*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. I am indebted to Christian Traxler for providing guidance for this paper and to Georges Siotis, who introduced me to the idea of a tax lottery in Greece. 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 can increase the compliance of firms in the payment of value-added tax (VAT hereafter), by providing incentives to final consumers to ask for a receipt of purchase. In business-to-business transactions, VAT has been shown to have a self-enforcing tax compliance element through paper trail: one firm's input serves as the other's output (Pomeranz, 2015). However, in business-to-consumer transactions, consumers do not have a similar incentive to ask for a receipt.[1] Tax lotteries in retail transactions have been employed by tax administrations as a way to close this gap. They encourage consumers to ask for a receipt in exchange for the possibility to win a prize, thus generating additional third-party information (Kleven *et al.*, 2011) that can increase VAT tax compliance (Naritomi, 2019). Using tax administration and consumption data from Greece, this paper studies the Greek tax lottery and documents (a) income and occupation characteristics that determine winners (b) heterogeneity in treatment effects to winning along the income and occupation dimensions.

The Greek tax lottery assigns tickets automatically to the entire population of Greek retail consumers provided they use credit/debit cards to complete transactions. It aims at incentivising a change in habits (from cash to electronic payments) and thereby in increasing the visibility of transactions that are processed through banking institutions. The latter become a third-party to the transaction, as opposed to transactions that take place in cash. Allocated tax lottery tickets map directly to the aggregate level of monthly electronic payments per individual, who then enter automatically in a draw to win € 1 million in prizes every month (€ 1,000 for 1,000 winners). I use 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 data allow for the reconstruction of a representative taxpayer population against which tax lottery winners can be compared to.

The analysis documents that due to the lottery's design, which assign tickets proportionally to the level of electronic consumption, important disparities in winners occur. When compared to the taxpayer population, winners exhibit roughly seven times higher mean annual electronic consumption (€ 28,413 versus € 3,931) and a higher mean annual income (€ 15,877 versus € 9,403). A 10% increase in income is associated with a 1.8% increase in the volume of electronic consumption and with a 0.11% increase in the probability of winning the tax lottery. Self-employed winners, exhibit an extreme large annual mean electronic consumption level of € 181,520 compared to € 14,626 for winners belonging to other occupational categories (wage-earners, pensioners and agricultural workers). Conditional on income, being self-employed increases the probability of winning by 0.18%, which is non-trivial given the large number of electronic transactions in the economy. The scheme's design awards more tickets to high-income/high-consumption individuals and, to the self-employed who are then selected as winners more frequently.

[1] The lack of paper trail in VAT at retail transactions has been commonly known as VAT's "last mile" problem.

However, an analysis of treatment effects heterogeneity by income quantiles and occupations reveals that post-winning responses are absent in individuals who are most likely to win, resulting in windfall gains without an associated treatment-related benefit. To examine responses, I rank individuals in household income quantiles (from lowest to highest) and in 5 occupation categories (wage-earners, pensioners, agriculture, self-employed and zero income).

The identification strategy relies on a natural experiment: an unanticipated superdraw on Christmas Eve in 2017. The tax authority planned monthly draws to start in January 2017, but due to a technical delay, the lottery was announced in October 2017. Earmarked prizes of € 9 million corresponding to months of January to September, could only be allocated to winners until the end of the budgetary year. To utilise the available funds, 9 retroactive draws took place on the 24th of December 2017 with tickets corresponding to electronic spending completed in the months of January to September without prior anticipation by taxpayers.

Results from event studies provide evidence of temporary responsiveness (increases) in electronic consumption at the first four income quantiles in response to winning and, of non-responsiveness at the highest income quantile. The lowest quantile exhibits statistically significant increases in electronic consumption by 16.6% in the first month, 20.9% in the second month and 19.5% in the seven month following winning. The 2nd quantile exibits statistically significant increases by 23% and 20% in the first two months. The 3rd income quantile has a statistically significant increase in electronic consumption only in the first month by 16.9%. The 4th quantile increases its electronic consumption by 9.1% in the first month and 9.5% in the third month after winning. By contrast, the highest income quantile does not respond to winning. The fact that lower income quantiles respond the most while the highest income quantile does not respond, runs counter to the scheme's design that links higher spending to higher winning probabilities.

Similar effects are present in an occupational comparison. Wage-earners respond by increasing their electronic consumption by 11%, 7.5% and 7.7% in the first three months after winning. Similar responses are observed in pensioners, ranging from 5.9% to 13.3%, and lasting for 6 months after winning. The self-employed do not respond to winning, whilst high consumption self-employed individuals decrease their electronic consumption permanently by 45-65%. One explanation for this large decrease is the use of personal bank accounts for business expenditure. Winning increases the salience of the lottery (and of the system of information exchange between banks and the tax authority) resulting in a reduction of electronic transaction volumes processed through their personal bank accounts.

The fact that individuals with the highest winning chances exhibit the least treatment benefit, appears to be an inherent problem in the scheme, with important efficacy and fairness implications. Linking spending to lottery tickets does not necessarily result in effective incentives for increasing electronic payments. Windfall gains are allocated mostly at the highest-income/highest-consumption individuals without additional third-party information being generated.

A simple policy solution to this problem is the introduction of a ceiling on the monthly tickets that can be assigned per individual, in order to limit the awarded tickets of high spenders. I perform repeated simulations of the monthly draws with conditions placed on the number of tickets per individual at two ceiling levels: € 1,000 and € 5,000 per month. Results indicate that both ceilings reduce the chances of winning for high consumption individuals, yet the stricter the ceiling the more binding it becomes for the taxpayers, which risks limiting the incentives offered by the tax lottery. Therefore, a fundamental trade-off for the tax lottery's design is to limit the excessive consumption while maintaining the incentives for electronic payments.

Despite of the introduction of tax lotteries in a number of countries in recent years, evidence on their efficacy, fairness and design characteristics remains particularly slim in the literature. A number of studies have documented the policy's effectiveness in raising tax revenue (Naritomi, 2019; Nicolaides, 2023). In the Brazilian tax lottery, Naritomi (2019) examines differences in wholesale versus retail sale of goods and services. The introduction of the lottery leads to a 21% increase in reported sales and a lower, yet significant, increase of 9.3% in reported revenue. The study identifies whistle-blowing and collusion costs as potential mechanisms for the increase. For the Greek tax lottery, Nicolaides (2023) 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. This paper contributes to the tax lottery literature by providing a detailed examination of the winners' characteristics in terms of income, electronic consumption and occupation, as well as, the heterogeneity in treatment effects after receiving a prize. Who wins has important implications for the scheme's efficacy and for limiting potential windfall gains to subgroups of the population. A final contribution to the tax lottery literature is a simple policy solution in the form of a ticket ceiling to improve the lottery's design.

The paper is organised as follows. Section 2 provides the institutional information of the Greek tax lottery and Section 3 describes the data. The winners' determinants are documented in Section 4. The methodology and results from treatment effects of winning are presented in Section 5. A ticket ceiling and related simulations are discussed in Section 6. Lastly, Section 7 concludes.

2 Institutional Background

Tax lotteries have become a common tool to mobilise consumers as a source of third-party reporting, thereby expanding the tax base and ultimately increasing 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).[2] 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.[3] Thus, the design of the lottery was 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.

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.[4] 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 (with 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 in electronic payments; this policy is analysed in Nicolaides (2022).[5] 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 on a monthly basis. The reporting system serves as a key building block in the tax lottery: it links monthly electronic payments per individual to lottery tickets.

[2]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 in Greece's second structural programme, which eventually expired in early 2015.

[3]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`.

[4]Cash withdrawals were limited to € 60 per day per individual in the summer of 2015, but electronic payments remained unlimited.

[5]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.

Lottery Tickets At the end of each month, banks send to the tax authority the *aggregate volume* of electronic payments (but not single transactions) completed by each Greek tax resident.[6] All tax residents are included in the lottery by default, as long as they complete payments with electronic means.[7] The monthly volume of electronic transactions are converted into tax lottery tickets according to a given ticket-awarding mechanism (or *TAM* hereafter). The TAM has a slightly concave structure: at higher levels of electronic transaction volumes, an additional euro would translate into fewer tickets.[8] This point is documented in Table 13 in Appendix D, which presents the detailed structure of TAM. While the first euro of monthly electronic consumption would translate into one ticket, the € 1,001$^{\text{st}}$ would yield only 0.25 tickets. Note further that the TAM does not contain any upper bound. Figure 38 in Appendix D plots the resulting euro-to-ticket mapping.[9] Eligible payments that are converted into tickets are limited to everyday consumption expenses. Excluded are purchases of intangible or tangible assets, motor vehicles and payments of house rent, mortgages, taxes and fines. All other purchases award tickets if they are completed with credit cards, debit cards and electronic payments.

Prizes Every month 1,000 winners win €1,000 each (€1 million in prizes per month). To ensure the fairness of the draws, the tax authority has implemented a double-blind draw system, where at first a research institute performs the draws and returns the winning numbers and then the tax authority applies a transformation to the numbers. In addition, individuals can only win once every month. For payments in a given month m, draws take place at the end of $m + 1$.[10] Winning tickets are announced to the public after the draw and winners are informed automatically via email and a text message to their mobile phones. They receive the prize in their bank accounts about a week after winning. A dedicate website allows the public to view their tickets for all lottery months, as well as, any winning tickets. Each winner receives € 1,000. The prize is tax exempt and cannot be confiscated.

[6] It is compulsory for all Greek tax residents above the age of 18 to acquire a tax ID, called AFM. This number acts as the main identifier of citizens by the state, much like an identity number. The matching of individuals between banks and the tax authority takes place through the tax ID. On one hand, when filing taxes individuals must declare their IBAN to complete the filing process. It is compulsory for all individuals above the age of 18 in Greece to file, even if they had no income during the financial year. To improve tax compliance during the economic crisis the filing process became completely electronic and automated with pre-filled information (paper declarations were eliminated). On the other hand, banks demand a tax ID when opening a bank account. This ensures matching when banks send the payment information to the tax authority.

[7] Individuals can opt out of the lottery by making a request to the tax authority. The request does not prevent banks from sending their payment information.

[8] A concave structure must have reflected the legislator's concern that high income taxpayers would had been awarded more tickets. This feature is investigated later in the paper.

[9] Note that the scale is public knowledge. At the introduction of the lottery it was rewarding 1 ticket per €1 for the first €100 spent; 1 ticket per €2 for the additional €400 (i.e. from €100 to €500); 1 ticket per €3 for the additional €500 (i.e. from €500 to €1,000); and 1 ticket per €4 for any payments above €1,000. For example, suppose that in a given month an individual spends €200 in electronic payments. The individual would receive 150 tickets (100 for the first €100 and 50 for the rest).

[10] For example, for all payments completed in October, banks collect payment information from October 1st to 31st, aggregate them and send them to the tax authority early November. Payments are converted to eligible tickets and the draw take place at the end of November. Winning numbers are announced immediately after the draw. The same procedure applies for the rest of the months.

Superdraw in December 2017 At Christmas Eve in 2017 a unique and unexpected superdraw took place with 9,000 winners and €9 million in prizes. Since the lottery was initially planned to begin in January 2017, the tax authority budgeted €12 million in prizes for the entire year, €1 million for each month. A public announcement of the lottery took place on the 9th of October 2017, where the TAM and the prize structure were made public. The first lottery took place at the 30th of November with payments completed in October and a second lottery was planned for 30th of December for payments completed in November. A €9 million earmarked amount corresponding to lotteries in the previous months remained unused and could only be allocated before the end of the financial year.[11] On the 24th of December 2017, the tax authority decided to run 9 consecutive draws, each corresponding to monthly payments completed from January 2017 to September 2017.[12] I utilise the superdraw event in this paper as an identification strategy to evaluate the taxpayers' responses to winning.

The Greek tax lottery differs from other tax lotteries in that it is hardly based on self-selection. Typically, consumers must register in a system and collect receipts in order to participate in a tax lottery (see, e.g., Naritomi, 2019). Instead, the Greek lottery automatically includes the vast majority of taxpayers provided that (i) they hold a bank account and (ii) they make an electronic transaction in a given month.[13] While the intention of the tax authority was to include only the private consumption by individuals to the draws, the lottery included a non-trivial volume of business transactions because of this automatic inclusion.[14] The line separating business and individual transactions will be, as a result, particularly blurry for self-employed individuals, since business transactions can be made through their personal bank accounts. The self-employed might have an additional incentive for business transactions, since work-related expenses can reduce the tax liability (for instance, office rents, phone expenses, traveling and more). From 2014, the law required that only expenditure above € 500 could count as eligible for tax deductions and that those should have taken place by electronic payments (through banks).

[11]This was because of budgetary reasons. Accrual amounts to individual winners could only be made until 31st of December, even in payments took place a few days into the new financial year. As with any public organisation, the budget is annual and earmarked amounts cannot be transferred to the following year.

[12]A visual illustration of the lottery's timeline in 2017 is shown in Figure 39 in Appendix D. Search volumes in Google search engine recorded in Greece at the time for the word "Lottery" in Greek are shown in Figure 40 in Appendix D. 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, indicating increasing public awareness. The search volume index records increases at the end of each month thereafter, in line with the time of monthly draws.

[13]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, the formal banking system includes almost the entire population.

[14]This peculiarity covers the entire time horizon of our data. It was solved only in 2019, by obliging individuals to hold separate bank accounts for business transactions. The data in this study cover the period prior to this change.

3 Data

Tax Filings The data contain (a) anonymous taxpayer information on the monthly level of electronic consumption for the period from January 2017 to July 2018, (b) matched to tax filings in 2017.[15] I observe the annual pre-tax income 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 consumption correspond to only one of the two individuals in the tax unit. Monthly electronic consumption was rounded to the nearest € 10 and information in tax filings to the nearest € 5.[16] For joint filings, I observe income values for both partners, enabling the calculation of the tax unit's income.[17] For single filings, I observe the income of the single person in the filing, which is also a single-household income.[18] It is compulsory to file tax returns even if an individual has zero income.[19]

Occupation Categories In addition to the income amount, the data indicate the source(s) of income in broad categories: from wages (subsequently WG); from pensions (PE); self-employed income (SB); and agricultural income (AG).[20] WG includes income received from salaried activities, which is the tax unit's reported annual gross salary. PE includes all individuals who receive pensionable income. SB includes sole proprietorships, such as the self-employed and sole traders. AG contains declared annual income from agricultural activities, such as for farm owners, agricultural workers and small cultivation. An additional category contains individuals who have reported zero income in 2017 (subsequently NO). 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. However, it might also contain individuals from the SB and AG income categories, who report zero income. Percentages of these categories for single-filing and joint-filing households, are shown in Table 3. Note that a given tax unit might of course declare income from multiple sources. Below, I will use indicators that define the primary source of reported incomes. These binary variables also serve as a proxy for each tax unit's primary occupational activity.

As a result of between-category variation in third-party reporting, there are major differences in the opportunities to under-report income. For WG and PE income, the income values (as reported

[15] The last day of tax filings 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.

[16] The 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.

[17] For tax filings in 2017, 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.

[18] 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.

[19] A consequence of this is that 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 income.

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

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.

Samples I use data from two different samples. First, the universe of tax lottery winners. This includes 18,897 individuals from tax units that have won the lottery during the first 19 consecutive lottery draws, from January 2017 to July 2018. Second, a randomly drawn sample of 50,000 tax units that did *not* win the tax lottery. For the winners' sample, and for all 19 months covered, I observe the monthly volume of electronic consumption for the winning individual, while for the randomly-drawn sample I observe electronic consumption for one individual in the tax unit. For both samples, I observe the annual income during tax filing. Table 3 presents basic summary statistics for the two samples.

Construction of Taxpayer Population To allow for a meaningful comparison of the winners' characteristics, one has to account for the different sampling for winners and non-winners. The winners sample was pre-selected from the taxpayer population (therefore, it is not a random sample). The non-winners sample was drawn randomly from the population of taxpayers – conditional on not having won. 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 electronic consumption) observed in 2017.

The following procedure is used to obtain the sample weights. Firstly, I observe the total number of lottery tickets issued in each calendar month, $\bar{T}_m$. Secondly, given that lottery tickets are derived from monthly electronic consumption, one can compute $T_{i,m,s}$, which is 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. As a final step, 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:

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

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

From this identity, the final step is to derive ω, the weight or expansion factor for the non-winners sample that matches the population in terms of lottery tickets, which 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.[21] Given that the winning sample is only a tiny fraction of the overall population, the non-random sampling of winners does not result in any noticeable distortions of the taxpayers' characteristics.

Basic summary statistics from the resulting, expanded sample are presented in Table 4. The table compares the baseline tax unit population with the tax units of winners from 2017. Several interesting characteristics can be observed. Firstly, the mean electronic consumption and mean income of winners is significantly higher than the rest of the population. Secondly, the SB category is over-represented in the winners and exhibits very high levels (and variance) of electronic consumption. Thirdly, winners in the NO category, had a particularly high level of electronic consumption. These discrepancies are central to the analysis of the winners' determinants in Section 4.

4 Determinants of Lottery Winners

This section provides evidence on income, consumption and occupation disparities in winners. Descriptive evidence show that (a) winners exhibit higher income and electronic consumption compared to the representative taxpayer population (b) self-employed winners generate a particularly large volume of electronic consumption as compared to their income. The evidence indicates that the lottery has selected subgroups of the population as winners based on income and occupation characteristics. In a second step, I parametrically estimate the effect of income and occupation disparities on the probability of winning the tax lottery.

4.1 Income and Consumption Disparities

The income comparison takes place using annual income in the 2017 tax filings, while electronic consumption can be compared by summing the monthly electronic consumption $z_{i,m}$ for an entire calendar year, i.e., by obtaining $Z_i = \sum_{m=1}^{12} z_{i,m}$ for each taxpayer i in the data.

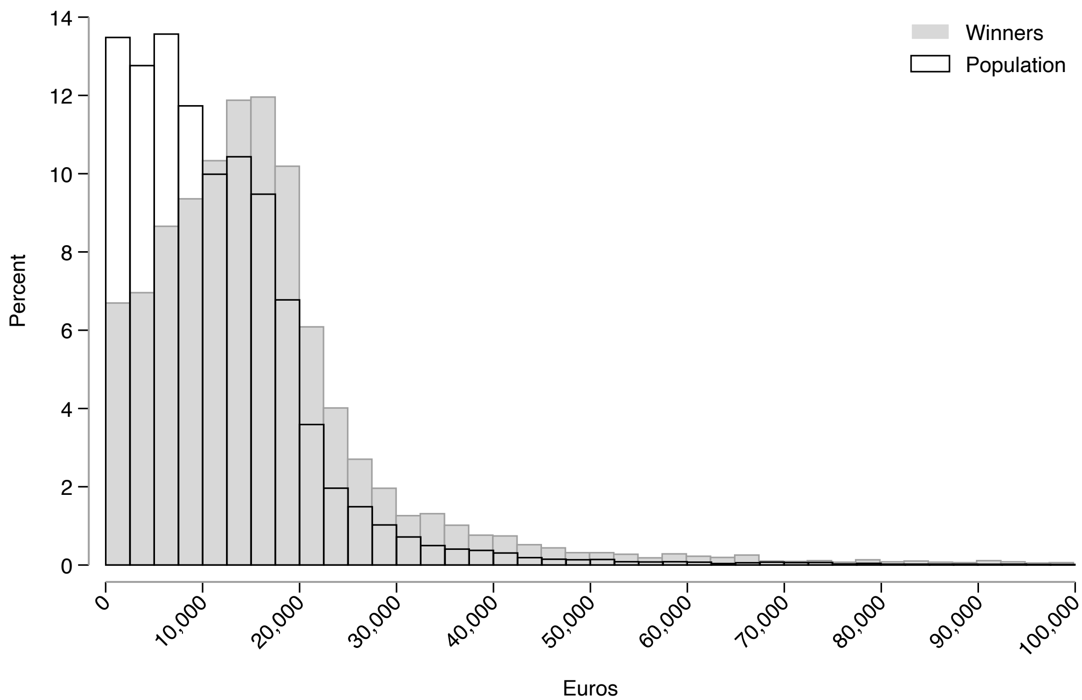
The winners exhibit a higher level of income compared to taxpayer population. Their mean income is € 15,877, as opposed to € 9,403 (median values are € 12,113 and € 6,850, respectively). This point is shown in Table 4 and illustrated graphically in Figure 1, which compares their income distributions. For the taxpayer population, a high visual mass can be observed below the € 10,000 level, after which the distribution tails-off fast as income increases. By contrast, the distribution of lottery winners exhibits lower mass below the € 10,000 level, after which the mass increases substantially with income. Tailing off takes place after about € 16,000. Overall, Figure 1 clearly

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

reflects the fact that lottery winners are (judged against the taxpayer population) higher-income taxpayers.[22]

Income disparities are reflect also in electronic consumption, where differences between winners and taxpayer population are even more acute. As can be seen in Table 4, winners exhibit roughly seven times as high mean annual electronic consumption compared to the taxpayer population, at €28,413 versus €3,931 (with median values at € 6,400 and € 1,940, respectively). This reflects the basic property of the lottery: the chances of winning increase proportionally to the level of electronic consumption. 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 $z_{i,m} \to \infty$. While this holds for a given month m, as electronic consumption fluctuates between months, the annual level Z_i is only an indirect indicator for the selection implied by the TAM.[23]

Fig. 1 Income Distribution in 2017



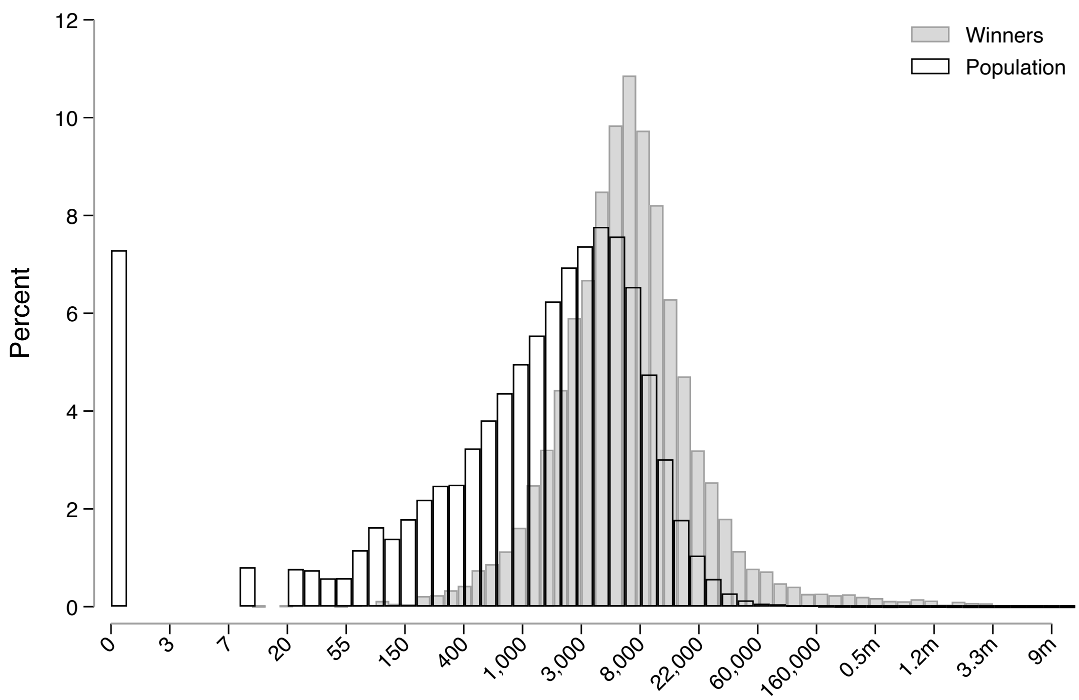
Notes: The figure compares the distribution of declared income in the 2017 tax filings for lottery winners against the corresponding distribution of income for the taxpayer population. The latter has been reconstructed from a random sample of 50,000 tax units as described in Section 2. The graph is truncated at € 100,000, as right-tails diminish quickly in the distribution.

[22]Note that for illustration purposes, Figure 1 is truncated at € 100,000. The winners distribution exhibits a longer tail as income increases, with a number of observations with high incomes well above € 100,000.

[23]The difference between mean electronic consumption for winners relative to the population is even more evident in a monthly comparison. The taxpayer population's monthly mean electronic consumption 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 electronic consumption of those who have won in a particular month fluctuated around € 4,000 (without observing any upward trend).

The distributional difference in electronic consumption (more specifically, in $\log(Z_i)$) is shown in Figure 2. For the taxpayer population, the distribution is bi-modal: a large visual mass of taxpayers (about 7%) exhibit zero electronic consumption for the entire year. The remaining taxpayers are concentrated around the €3,931 mean value. By contrast, the winners' distribution is symmetrical and normally-distributed, with more taxpayer mass as income increases. There is a heavier right-hand tail, with a non-trivial share of electronic consumption volumes well above € 60,000. For 2017, there were 334 winners with more than € 1 million annual electronic consumption, 34 with more than € 2 million and one extreme value of more than € 9 million e-consumption (who has won twice in 2017). By contrast, the taxpayer population distribution, exhibits hardly any mass in the range of $Z_i >$€ $22,000$.

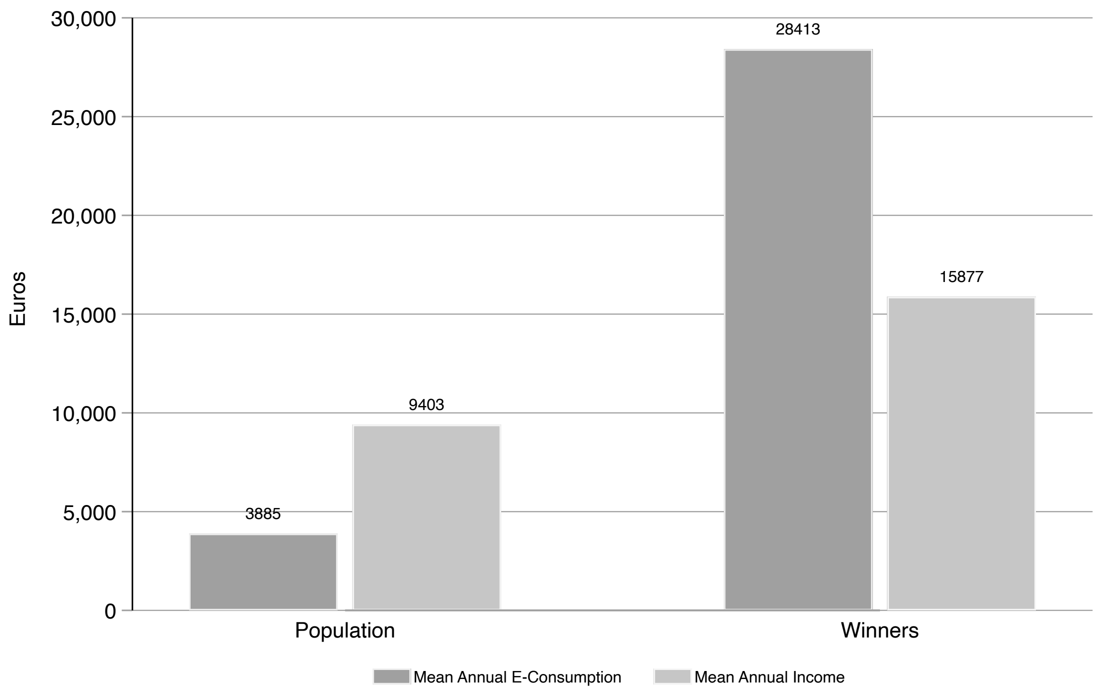
Fig. 2 Distribution of Annual Electronic Consumption



Notes: The figures presents distributions of the log of annual electronic consumption in 2017 for individuals in the taxpayer population and for lottery winners. 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 in the reconstructed taxpayer population. The winners distribution includes 11,960 winners in tax lotteries that took place in 2017. The monthly electronic consumption of individuals was summed up over the 12 months to create the annual electronic consumption. Monthly values in the data were rounded to the nearest € 10 by the tax authority.

A peculiar characteristic for winners is that their electronic consumption is significantly higher than their income. This is shown in the winners' column of Figure 3, while the corresponding comparison for the taxpayer population is shown in the population column. The e-consumption-over-income ratio is 1.79 for winners and 0.42 for the taxpayer population. This indicates that winners spent almost twice as much as their income using electronic payments, while the taxpayer population spent less than half. While this peculiarity in winners' consumption pattern is largely driven by some outliers with very high electronic consumption values, it does hold for a significant number of winners. Specifically, for every third winner (33.5%) I observe an e-consumption-over-income ratio above unity, i.e., their annual electronic consumption volume exceeds their income.

Fig. 3 Electronic Consumption and Income



Notes: The figure compares the annual mean electronic consumption against the mean annual income for winners and for the taxpayer population. Only winners from lotteries in 2017 are included in the winners' sample. Non-parametric estimates of the differences are provided in Table 5.

4.2 Occupation Disparities

In addition to the income and consumption disparities documented above, the income source of taxpayers is a determining factor in winning the tax lottery, and in particular being self-employed. Recall that our sample can be divided in occupation categories, for which the income source in tax filing is recorded: wage-earners (WG), pensioners (PE), self-employed (SB) and agricultural workers (AG). Table 4 documents significant within-occupational-category differences in mean income between winners and the taxpayer population. SB taxpayers are over-represented among lottery winners. Relative to a population share of 4.1%, this group accounts for 8.3% of all winners.

WG, PE and AG taxpayers have roughly similar representation in the winners as in the taxpayer population. Taxpayers in the zero declared income (NO) group are under-represented compared to their population percentage.[24]

Comparing winners against the taxpayer population per occupation category, reveals that income differences are statistically significant for all WG, PE, SB and AG categories. Regardless of the income category, winners exhibit higher income, as shown in non-parametric estimations in Table 5. The income of winners is almost double for the SB and WG categories compared to the taxpayer population and, almost triple for the AG category and about one-forth higher for the PE category. A larger income variance for these categories is also observed. The median income differences per occupation category, for winners and the population respectively are: for SB € 11,073 and € 6,260; for WG € 14,325 and € 9,145; for PE € 14,858 and € 11,275; and for AG € 17,090 and € 7,575.

A similar comparison for electronic consumption reveals an extremely large difference between winners and taxpayer population in the SB category. As documented in Table 4, the mean electronic consumption for winners is € 181,520 (median of €17,565). Among the taxpayer population, the corresponding value € 11,420 (median of € 4,410). Hence, in addition to the fact SB taxpayers win the lottery more frequently than others, the lottery also selects (within the SB group) winners with unusually high electronic consumption volumes.

This discrepancy in SB individuals is depicted in Figure 4, where income and electronic consumption are compared for winners (on the right) and for the taxpayer population (on the left). To aid the comparison, the corresponding graphs for pooled groups of WG, PE and AG are also plotted.[25] Firstly, note that the mean electronic consumption represents about one-third of income if the taxpayer not SB and about as much as their income if the taxpayer is SB. Hence, these income categories display a different pattern: the self-employed exhibit high volumes of electronic consumption, which interestingly is as high as their income. Importantly, as shown in Figure 4 (b), these differences are extremely large among winners: for SB winners, electronic consumption is ten times as high as their income. For winners in other income categories, electronic consumption is about as high as their income.[26]

Additional evidence examining the relationship of electronic consumption and income for the categories examined above are presented in scatter plots in Figure 5. As can be seen, in the Winners - SB scatter plot at the bottom right-hand corner, a proportionally larger number of SB winners exhibit high volumes of electronic consumption, than winners from other categories. When

[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-consumption; (b) individuals in the SB category who report zero.

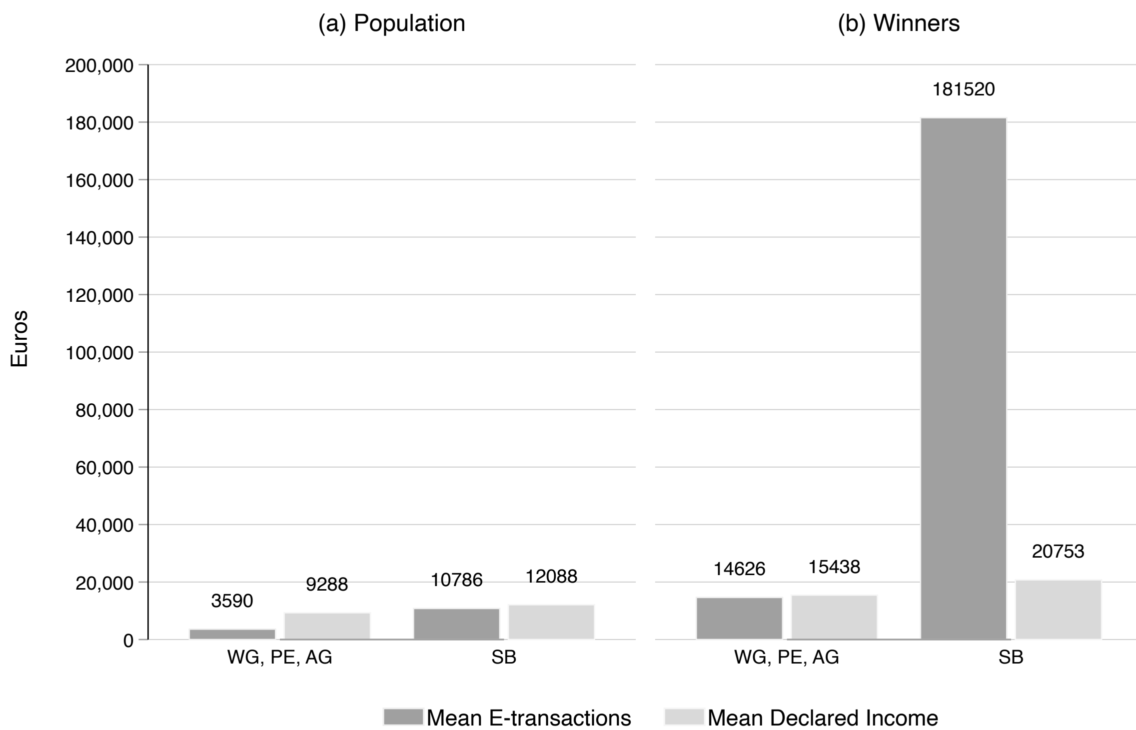
[25] To allow for a meaningful comparison, the NO category individuals are excluded because they declare zero income.

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

the SB winners are compared to the SB population, one can observe that SB winners exhibit high consumption volumes, as well as, higher income.

The extreme divergence between electronic consumption and income for SB lottery winners suggests that they (a) use private bank accounts when completing business transactions and/or (b) income is under-reported. Regarding the former, the flow of business transactions through personal bank accounts can result in particularly high electronic consumption levels, which then generates a large number of tax lottery tickets. Recall from Section 2, that the use of private bank accounts for business proposes was prohibited in 2019. Hence, it was still possible in 2017 for some business expenses to be channeled through private bank accounts, but one should expect that this practice would have eclipsed slowly from 2019 onwards. Regarding the latter, the observed pattern might also originate from illegal under-reporting of income: since the SB group has (relative to third-party reported income) more opportunities to conceal (Kleven *et al.*, 2011), the vast electronic consumption/income gap may therefore – at least in part – reflect income tax evasion.[27] These two reasons rationalise the disproportionate representation of SB among tax lottery winners.

Fig. 4 Annual Income and Electronic Consumption, by Income Category



Notes: The figure compares mean electronic consumption and mean income for groups with different primary income sources: self-employed (SB) vs other non-zero incomes from wages, pensions and agricultural activities (WG, PE, and AG). Individuals who declare zero income (NO) are excluded from this comparison. Figure (a) is based on taxpayer population and figure (b) presents the lottery winners from 2017.

[27]Note that the data do not allow quantification of this channel.

4.3 Occupation Disparities within Households

Examination of intra-household income sharing and household income source composition, provides corroborating evidence of income and occupation disparities. As long as some couples share their (private) bank account, and if partners with SB income use the accounts for business transactions, one should expect to observe higher electronic consumption levels for individuals jointly filing with an 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 6 compares electronic consumption and income levels of individuals who file jointly and who have an SB spouse against those who have a WG/PE/AG spouse.[29] (To facilitate interpretation, the sample underlying this graph excludes individuals from the SB and the NO income categories). Having an SB spouse is associated with higher levels of electronic consumption. This holds for the taxpayer population (Panel (a) of Figure 6) 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 the spouses of SB individuals seem to use these accounts for business transactions.[31]

4.4 Estimation of Winning Probability

This section estimates the relationship between the winners' determinants and the probability of winning. In particular, I explore the relationship between individual/spousal income and, occupation categories for (i) the level of electronic consumption and (ii) the probability of winning the lottery. To capture this, I consider models of the following structure:

$$\log(Z_i) = \beta_0 + \beta_1 \log(Y_i) + \beta_2 \log\left(Y_{j|i}\right) + \beta_3\, SB_i + \beta_4\, SB_{j|i} + \beta_5\, Joint_i + \varepsilon_i \qquad (2)$$

where Z_i is the annual electronic consumption, Y_i indicates the annual income and SB_i is a binary variable indicating self-employed income. The sub-index $j|i$ measures these variables for i's spouse j. $Joint_i$ is a binary variable indicating an individual who has filed jointly with a spouse. Note

[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 7, 8 and 9 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 as shown in column (1). Table 6, column (2), documents that this difference in annual electronic consumption is economically and statistically highly significant. The difference is hardly affected by controlling for 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.

that β_1 and β_2 capture taxpayer i's elasticity of electronic consumption with respect to their own and their spouse's income, respectively.[32]

Columns (1)–(3) in Table 1 reports ordinary least squares estimates that follow the structure of Equation (2). The estimated β_1 suggests that a 10% higher income correlates with a 1.8% increase in electronic consumption. This measure is similar to a marginal propensity to consume estimate for electronic consumption. 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 1 falls within this range. The coefficient hardly changes in Column (2), when spousal income is controlled for. The correlation is significantly lower: a 10% increase in spousal income is associated with a 0.7% increase in individual i's electronic consumption. An F-test reject the null $\beta_1 = \beta_2$, indicating imperfect income sharing within the household (Browning *et al.*, 1994; Lundberg *et al.*, 1997) or differential propensities to engage in electronic consumption. The results are quantitatively similar in Column (3), where only taxpayers who filed jointly are considered.

Table 1 Estimation Results

	Log(e-consumption)			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-employed	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-employed 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 in 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 from Equation (2). The dependent variable in columns (1)–(3) are the logarithm of annual electronic consumption ($\log(Z_i)$) and, in columns (4)–(7), the probability of 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 $2,406,683$ joint-filing taxpayers. Robust standard errors (clustered at the level of 50,000 non-winning tax units + 11,960 winners from 2017) in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

[32] Perfect income sharing within a household (plus equal propensities to spend money electronically) would imply $\beta_1 = \beta_2$.

The estimates further document that, consistent with the descriptive evidence from above, being an SB is associated with significantly higher levels of electronic consumption. The estimated semi-elasticities imply that SB income is associated with an approximately 75% higher level electronic consumption compared to other income categories. An SB spouse increases electronic consumption by 45%. It is worth stressing that this holds while controlling for the taxpayer's own, as well as, for spousal income. Hence, the pattern reflects an occupational rather than a mere income correlation.

In a second step, I estimate the effect of the observed disparities on the probability of winning the tax lottery. The dependent variable from Equation (2) is replaced with a binary variable W_i which indicates that the taxpayer has won in the lottery in 2017. The results from a linear probability model are presented in Columns (4)–(7) in Table 1. Note that for presentation reasons, the estimates have been multiplied by 100 (probabilities of winning are rather small). The results document a positive correlation vis-à-vis a taxpayer's income and spousal income with the probability of winning the tax lottery. A 10% increase in taxpayer (spousal) income raises the probability of winning by 0.11% (0.06%).

It is striking to observe that SB taxpayers or (jointly filing) taxpayers with an SB spouse have a significantly higher chance to win the tax lottery. Controlling for income, the associated semi-elasticity in Column (5) indicates that being SB increases the probability of winning by 0.177% compared to other income categories. Having an SB spouse results in a higher probability of winning the lottery by 0.036%, which is statistically significant at the 5% level. Results are qualitatively unchanged with estimates only from the sample of jointly filing taxpayers (in Column (6)).

Finally, it is worth noting the specification in Column (7) of Table 1, which includes the annual number of tickets assigned to an individual in 2017 in the regression. 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 the probability of winning vis-à-vis the level of income and the income category, are merely shaped by the association of these variables with the amount of electronic consumption. The latter translates into tickets and, ultimately, determines the probability of winning.

5 Treatment Effects Heterogeneity

The income and occupation disparities documented so far imply that certain subgroups are selected more frequently as winners. Given that the lottery aims at increasing third-party reporting by incentivising more electronic consumption, a winners' subgroup selection would had been beneficial had winners responded by increasing their electronic consumption after receiving the € 1,000 prize. This section provides evidence of treatment effects heterogeneity in taxpayers based on their income and occupation. Low-to-middle income quantiles exhibit higher responsiveness to winning (yet temporary), while taxpayers in the highest income quantile do not increase their electronic consumption after winning. In an occupational comparison, SB individuals (who are more likely to win) do not respond after winning and high consumption SB respond negatively to winning. WG and PE winners have the highest responsiveness among occupational categories. As a result, the more frequent selection of high income individuals and the SB subgroup as winners by the lottery's design has efficacy implications for the scheme.

To facilitate this comparison, taxpayers are separated in quantiles based on household income and provided their annual electronic consumption was positive (i.e. at least € 1 was spent by electronic payments, which would have ensured lottery participation). Since the data include values of individual electronic consumption per month from January 2017 to July 2018, one can compare monthly consumption responses for winners, who are "treated" by receiving € 1,000, against non-winners, who can act as "control". This comparison takes place per income quantile and per occupation category.

5.1 Identification Strategy

The identification strategy utilises the superdraw that took place on Christmas Eve in December 2017 (see Section 2). The superdraw created conditions that resemble a natural experiment. Firstly, it was impossible for taxpayers to self-select in the lottery: eligible electronic payments, which were converted into lottery tickets, were completed before the tax lottery's announcement. This ensures that any lottery-loving behaviour by some individuals or increased lottery-salience for some others were not factors affecting their participation. Secondly, as the draws were retroactive (for payments that took place in previous months), taxpayers could not affect their winning chances. Thirdly, the allocation of prizes (which determines the treatment and control groups) was random conditional on electronic consumption. Lastly, the setup is simple since it involves a common treatment level (€ 1,000) and single timing (information on winning arriving on Christmas 2017 and prize money in early January 2018 by bank transfer in taxpayers' bank accounts).

Hence, to ensure a random treatment, it suffices to control for electronic consumption levels in the population which determined the probability of winning. As was shown in Section 4.4, higher spending increases the chances of winning, which then determines the probability of assignment in the treatment group. Accounting for the probability of winning is necessary to create two comparable groups where a similar spending pattern during the lottery months occurred. This allows one to define a valid counterfactual of non-winners who exhibited similar payment behaviour to winners.

5.2 Matching Procedure

Matching winners and non-winners follows a two-step procedure. First, I implement coarsened exact matching based on income quantiles and (separately) on occupation categories.[33] This ensures that

[33]Note that matching on cross-categories of income quantiles and occupation is not considered in this study, since some occupational categories are small (SB and AG in particular), which does not provide enough statistical power after matching, propensity scoring and event study regressions are implemented.

the comparison of winners and non-winners takes account of similar levels of household income or occupation categories, which can determine spending habits and the level of electronic consumption.

Secondly, for each coarsened exact matched subsample (i.e. 5 quantile groups and 5 occupation categories), I obtain a propensity score of being treated using the following procedure. The propensity score produces a metric for the probability of an individual being a winner. Recall that when assessing the probability of winning in Column(7) of Table 1, the number of tickets was the most important variable, rendering any other variable statistically insignificant. Let W_i be a binary variable for individual i with the value of 1 if winning occurs and 0 otherwise. Let T_i, m represent tickets received in months $m \in [1, 9]$ (January to September 2017, which were the months of the Superdraw – see Section 2). The following logit model calculates the probabilities of winning:

$$P\left(W_i = 1\right) = \frac{1}{1 + \exp\left(-\left(\beta_0 + \sum_{m=1}^{9} \beta_m T_{i,m}\right)\right)} \tag{3}$$

Logistic regressions are fitted for the 5 quantile income groups and for 5 occupation groups (WG, PE, SB, AG and NO), using maximum likelihood with Firth's bias reduction (Firth, 1993; Heinze and Schemper, 2002).[34] Results are presented in Table 11 for the income quantiles and in Table 12 for the occupation quantiles. These show the increase in probability of winning for every ticket obtained in the months of January to September 2017. The numbers are relatively small, which shows that one extra ticket does not increase the probability significantly (given a high aggregate volume of transactions in the economy). Most of the values are positive and statistically significant.

Predicted values are plotted in Kernel density functions in Appendix C. Figures 26 to 30 present Kernel density functions, showing the probability of winning, for each of the income quantiles. Note that all quantiles are winsorized at 5% on both tails to avoid extreme values in the sample. In comparing the graphs, note, that as income quantiles increase, so does the probability of winning for both winners and non-winners.

Similar graphs are also plotted for occupational categories in Figures 31 to 36. For SB taxpayers in particular, the density function indicates a mass of taxpayers with very high probability of winning, which reflects the existence of extreme consumption individuals. This is shown in the right-hand tail of the winners (not winsorized) density function in Figure 36, which exhibits significant excess mass. To investigate how high consumption SB individuals react to winning, I form two subgroups of SB, one for propensity scores with 0.8 and above (high probability of winning) and another with less than 0.8. All density functions have a common support; they include individuals with the same probability of winning in each of the coarsened matched subsamples.

In a final step, I produce inverse probability weights from the propensity score, to re-weight the probability of being a winner. Weights are obtained within each coarsened exact matched sample.

[34] There are 1,000 winning tickets every month and roughly €10 billion in electronic consumption monthly, which makes winning a rare event. Firth's bias reduction ensures convergence of the maximum likelihood estimator.

This procedure produces comparable sets of winners and non-winners, taking account of their spending patterns and their income or occupation characteristics.

5.3 Estimation

I utilise the superdraw in December 2017 as an event to investigate the treatment-effects heterogeneity in electronic consumption across income quantiles and occupation categories. For each of the 5 income quantiles and 5 occupation categories, I compare sets of matched individuals (winners and non-winners) re-balanced using inverse probability weights. Overall, I observe 12 months before the event and 7 months after.

Recall that monthly electronic consumption is denoted by $z_{i,m}$ for individual i, while the binary indicator for winning is denoted by W_i. The following regression captures differences in electronic consumption between winners and non-winners:

$$\underbrace{z_{i,m}}_{E-Consumption} = \alpha + \overbrace{\beta W_i \times Post_m}^{Winner's\ \ Indicator} + \chi_i + \lambda_m + \epsilon_{i,m} \tag{4}$$

Variables χ_i and λ_m capture individual and time fixed effects respectively. Inverse probability weights from the propensity score estimation are used to re-weight individuals and control for the probability of selection in the winners group. Robust standard errors, clustered at the individual level, are used in all specifications.

Note that income quantiles exclude, firstly, self-employed individuals because of the high volumes of electronic consumption documented in Section 4. Since their electronic consumption is multiple times their income, inclusion in the income quantiles would have massively distorted the results. Their monthly spending is significantly higher than other income categories, whilst the use of transactions for business expenditure would not have facilitated a proportional comparison with everyday expenditure for households. Secondly, NO category individuals are excluded since they declare zero income (allocation in income quantiles cannot take place). Therefore, income quantiles include WG, PE and AG individuals, who exhibited positive income and positive electronic consumption.

Two estimators are used in Regression 4. First, a within-estimator, with taxpayer and month fixed effects, produces level differences in electronic consumption. Second, a Poisson pseudo maximum likelihood estimates the monthly semi-elasticity of electronic consumption with respect to winning.[35] It is useful to obtain both estimators as they act in tandem to understand taxpayer responses. The former allows for an examination of electronic consumption levels, as well as, an assessment of electronic consumption responses in absolute terms. The Poisson pseudo maximum likelihood

[35] This approach is preferred to logarithmic transformation of electronic consumption values since some months may contain zero values for some individuals.

provides log-point estimates, which can be transformed into percentages; a comparable measure of responsiveness in electronic consumption across income quantiles and occupation categories.

5.4 Responses by Income Quantile

Regression results for each income quantile are presented in Table 7 and Table 8, for linear and Poisson pseudo maximum likelihood, respectively. Corresponding graphs plotting level differences and log-point estimates are shown in Figure 10 to Figure 14 for each quantile, respectively. Linear estimates capture electronic consumption differences between winners and non-winners, while log-point estimates show the responsiveness of electronic consumption with respect to winning.

Examination of the monthly consumption evolution reveals parallel trends in all quantiles. The matching procedure produces comparable samples with no statistically significant differences in the 1st, 3rd, 4th and 5th quantile. The graphs capture clearly that winners and non-winners have a similar spending pattern prior to receiving the prize. A level divergence can be observed in the 2nd quantile, yet maintaining a parallel trend. It is reassuring to observe that prior to receiving the prize the winners and non-winners compared, have exhibited similar levels of electronic consumption (and therefore, had similar chances of winning).

Treatment effects in the month when the prize is received are present in the 1st, 2nd, 3rd and 4th quantiles, but not in the 5th (highest income) quantile. In January 2018, winners exhibited about € 80 increase in electronic consumption, followed by about € 50 in subsequent months for the lowest quantile. In addition, the lowest quantile maintains higher electronic consumption level than prior to winning, yet statistically significant at the 10% level in February and March, and at the 5% level in July 2018. By contrast, following a prize, differences between winners and non-winners are statistically insignificant for the highest income quantile, indicating non-responsiveness of electronic consumption to winning.

Log-point estimates in Table 8 (plotted at the bottom graph in Figures 10 to 14 for each quantile), provide a comparable measure of this response. The lowest quantile exhibits statistically significant increases in electronic consumption by 16.6% in the first month, 20.9% in the second month and 19.5% in the seven month following winning. The 2nd quantile exhibits statistically significant increases by 23% and 20% in the first two months. The 3rd income quantile has a statistically significant increase in electronic consumption only in the first month by 16.9%. The 4th quantile increases its electronic consumption by 9.1% in the first month and 9.5% in the third month following winning. Log-point differences are statistically insignificant for the 5th quantile.

Overall, the results from income quantile event studies, suggest that winners respond by increasing their electronic consumption temporarily at low-to-middle income quantiles, while taxpayers in the highest income quantile do not respond to winning. This result runs counter to the scheme's design that links higher spending to higher winning probabilities as documented in Section 4. Winning does not result in treatment-related benefits at the highest income individuals.

5.5 Responses by Occupation

Estimates from occupation categories are presented in Table 9 for linear estimates and in Table 10 for Poisson pseudo maximum likelihood estimates. Corresponding graphs for level and log-point estimates are depicted in Figure 15 for WG, Figure 16 for PE, Figure 19 for AG and Figure 20 for NO. As was noted above, SB individuals are grouped by those who exhibit high-consumption and all the rest, in Figure 18 and Figure 17, respectively.

Parallel trends are evident in both graphs and event study estimates. AG, NO and SB categories exhibit parallel trends at the same level of electronic consumption, whilst level differences are lower by about € 30-40 for WG and PE, yet still maintaining a parallel pattern between winners and non-winners. High consumption SB in Column (6) of Tables 9 and 10, exhibit some level differences in some of the months which originate from the fact that the sample is particularly small and more prone to monthly fluctuations.

Linear regression results indicate that following winning, WG, PE and NO individuals respond by increasing their electronic consumption by about € 100 in the first month of winning. By contrast, taxpayers in the AG and SB categories (excluding SB with high consumption), do not increase their electronic consumption following winning; differences in electronic consumption are statistically insignificant. Importantly, SB individuals with high electronic consumption, who were more likely to win, respond by reducing their electronic consumption significantly following winning. Note that the regression constant at € 112,850, gives an indication of the level of their consumption during the year. High consumption SB taxpayers respond by permanently reducing their electronic consumption by about € 30,000 following winning. This might indicate, once they win, an increase in salience of the tax lottery and of the visibility that the tax authority has in their transactions.

A comparison of the responsiveness' magnitude in electronic consumption across occupation categories is presented in Table 10. Log-point estimates are plotted at the bottom part of the occupation category figures mentioned above. Similar to linear estimates, it is reassuring to observe parallel trends in most of the responses. Importantly, individuals in the WG category responds by temporarily increasing their electronic consumption by 11%, 7.5% and 7.7% in the first three months after winning. Similar responses are observed in PE individuals, ranging to increases from 5.9% to 13.3%, and lasting for 6 months after winning. NO individuals respond the most by increasing consumption by 19.8% in the first month and up to 28.3% in the seventh month after winning.

By contrast, AG and SB individuals (excluding SB with high consumption), do not increase their electronic consumption after winning. Differences remain statistically insignificant. The responsiveness of SB individuals with high consumption is negative as was noted above. They reduce their consumption permanently by about 45-65% decrease after the prize is received. This is also shown at the bottom graph of Figure 18.

Similar to the income quantiles, the fact that SB individuals, and especially those with high consumption levels, do not respond to winning by increasing their electronic consumption has important implications for the efficacy of the tax lottery. As documented in Section 4, the lottery assigns a higher probability of winning to SB individuals due to their high volume of electronic consumption, yet winning the lottery does not result in a treatment-related benefit. The responsiveness of electronic consumption to winning is positive (yet, temporary) in the WG, PE and NO categories. These results run counter to the scheme's design of mapping tickets to the level of electronic consumption.

6 Ticket Ceilings to Improve Tax Lottery Design

The evidence above indicate that high income and SB individuals are more likely to win the tax lottery, yet they are the least responsive. This affects both the fairness and efficacy of the scheme as windfall gains end up in high-income/high-consumption taxpayers without any treatment-related benefit from winning the lottery.

A policy solution to improve the lottery's design is to implement a ticket ceiling in order to limit the awarded tickets of high spenders, who originate from the highest income quantile and from the SB category. I consider the effect of two monthly limits; at € 1,000 and € 5,000. These ceilings can limit the maximum number of tickets an individual can receive at 467 and 1,467 respectively (given the TAM). Figure 21 and Figure 22, plot the resulting distributions of tickets, indicating the level of tickets at which it becomes binding for individuals. As can be seen, the € 1,000 ceiling affects more individuals as they cannot receive more tickets once this is reached. A smaller number is affected by the € 5,000 ceiling.

Tax lottery simulations investigate the effects of the ticket ceilings. I transform the monthly electronic consumption of the population (from 2017) into lottery tickets and then I simulate (based on 100 iterations) the 1,000 lottery winners of the 12 lottery draws in a calendar year (this ensures that the simulations 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 winners (12 months × 100 iterations × 1,000 winners).

Simulation results are shown in Table 2. Both ceilings are effective in limiting the winning chances associated with very high monthly electronic consumption. At the highest decile, electronic consumption falls to € 15,350 for the € 5,000 ceiling and to € 13,550 for the € 1,000 ceiling.

Changes to the distribution of income and electronic consumption are illustrated over the entire population of winners in Figures 23 and 24 respectively. The graphs plot cumulative distribution curves ranking individuals by their income and electronic consumption, respectively. The curves show how the distribution of winners changes when the two ceilings are implemented. For comparison purposes, a simulation of a lottery without ticket ceiling is also plotted. As can be seen in Figures 23, the ceilings has very small effect to the income distribution of winners, reducing high income winners only marginally. They are, however, binding in electronic consumption, as can be seen in Figure 24. The electronic consumption is reduced dramatically for the highest electronic consumption decile, indicating that the ceiling can improve lottery design by limiting high consumption individuals from winning.

Table 2 Main Simulation Statistics - Ticket Ceilings

	Ceiling € 1,000				*Ceiling € 5,000*			
	Mean	p10	p50	p90	Mean	p10	p50	p90
Annual E-Consumption	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 (100 lottery iterations, drawing 1,000 winners in each iteration, for each of the 12 months in 2017). The first column presents the mean values and standard deviation in parentheses. The median values are presented in the "p50" columns, together with the lowest and highest deciles in "p10" and "p90" respectively.

By plotting electronic consumption cumulative distribution curves over the *income* distribution of winners in Figure 25, one can investigate the beneficiaries of such reforms. This produces an electronic consumption distribution of winners, where if all slopes at all deciles are equal to 45-degree, this serves as a point of equality, where the percentage of electronic consumption deciles equal the related percentage in the income distribution. Both ceilings reduces 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 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 21 and 22). 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 electronic consumption.

7 Conclusion

This paper has documented income and occupation determinants of winners in the Greek tax lottery, along with treatment effects heterogeneity. By mapping electronic consumption to allocated tickets, the scheme assigned higher winning probabilities to high-income/high-consumption subgroups of the population. The probability of winning increases by 0.11% in response to a 10% increase in income, while self-employed individuals were selected more frequently as winners due to a particularly high volume of electronic consumption. Being self-employed increased the chances of winning by 0.18% compared to wage-earners, pensioners and agricultural workers, after controlling for income.

By contrast, post-winning responses in these groups are absent, resulting in windfall gains without treatment-related benefit. Using an unanticipated superdraw in 2017, this paper implemented event studies comparing the electronic consumption of winners to non-winners across income quantiles and (separately) in five occupation categories. Results suggest a temporary increase in electronic consumption for low-to-middle income quantiles and a non-responsiveness for taxpayers in the highest income quantile. Similarly, wage-earners and pensioners increase their electronic consumption temporarily, whilst the self-employed do not exhibit any statistically significant differences post-winning. Self-employed individuals with particularly high consumption reduce their electronic consumption after winning, as the tax lottery becomes salient.

These results have important efficacy and fairness implications for the scheme. Since tax lotteries provide incentives to the final consumer to ask for a receipt at the point of purchase, it is high-income/high-consumption taxpayers and, the self-employed who stand to benefit the most from the lottery's monetary rewards but without any treatment-related benefit. Firstly, linking spending to lottery tickets does not necessarily result in effective incentives for increasing electronic consumption and for generating third-party reporting, which is the scheme's main objective. Secondly, windfall gains are allocated mostly at the highest income and highest spending individuals, with implications to the scheme's fairness. Thirdly, since the amount of prizes is constant, the policy becomes less salient in low-to-middle income individuals who experience winning less frequently, yet they respond the most in electronic consumption after winning.

The efficacy and fairness of the tax lottery can be improved by implementing a ticket ceiling per individual, after which additional tickets are not allocated. Simulation results in a static framework indicate that ticket ceilings can limit excessive electronic consumption, resulting in a fairer distribution of prizes.

Declarations

Competing interests

Partial financial support for field work in Greece was received from the Hertie School, Berlin through the German Academic Exchange Service (DAAD) funds. The author has no other known competing interests.

Data availability statement

The data used in this study consist of (i) 50,000 randomly-drawn tax units from the 2017-2018 taxpayer population in Greece and, (ii) 18,897 winning tax units. Both samples were anonymised and are non-identifiable. These were provided by the Independent Authority of Public Revenue in collaboration with the Greek Ministry of Finance in October 2018. The data were drawn and anonymised at the tax authority premises, to ensure confidentiality. However, the anonymised data are still considered confidential and cannot be shared publicly. Access to the data for replication purposes or use in future projects can be granted in a safe computer at the Paris School of Economics upon reasonable request to the author.

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.

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.

Firth, D. (1993). Bias Reduction of Maximum Likelihood Estimates. *Biometrika*, **80** (1), 27–38.

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.

Heinze, G. and Schemper, M. (2002). A solution to the problem of separation in logistic regression. *Statistics in Medicine*, **21** (16), 2409–2419.

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.

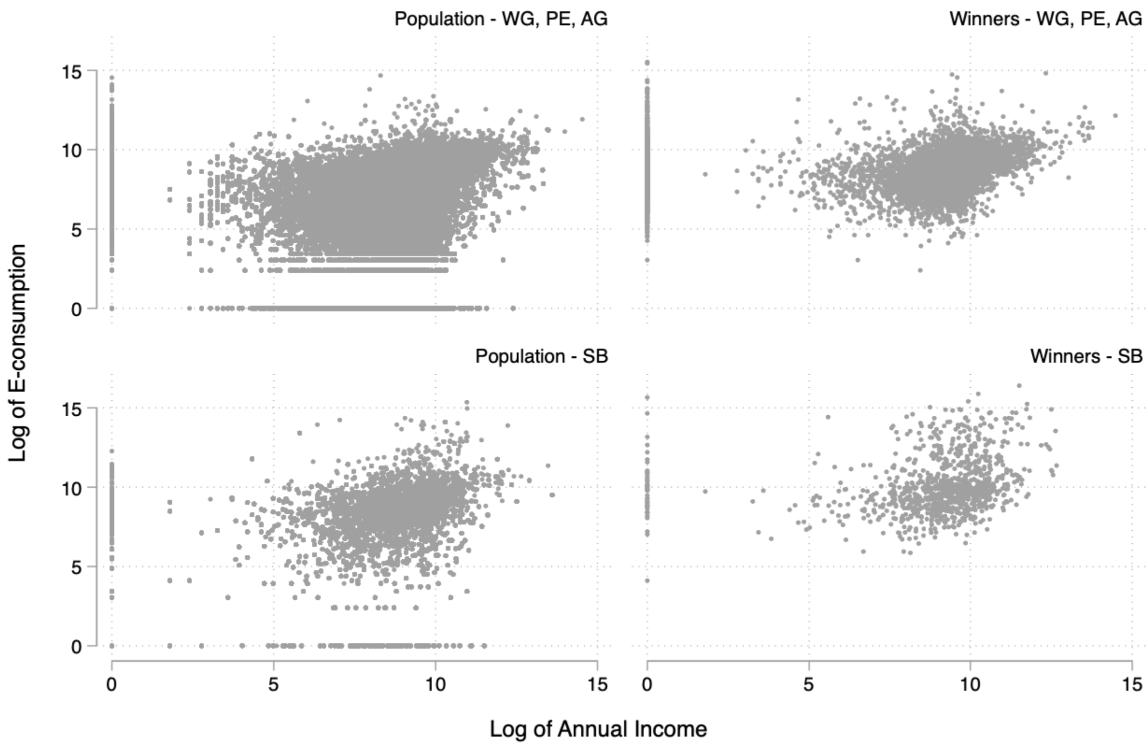
— (2023). Winning the Tax Lottery: Evidence from a Superdraw on Christmas Eve, Working Paper.

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

A Figures

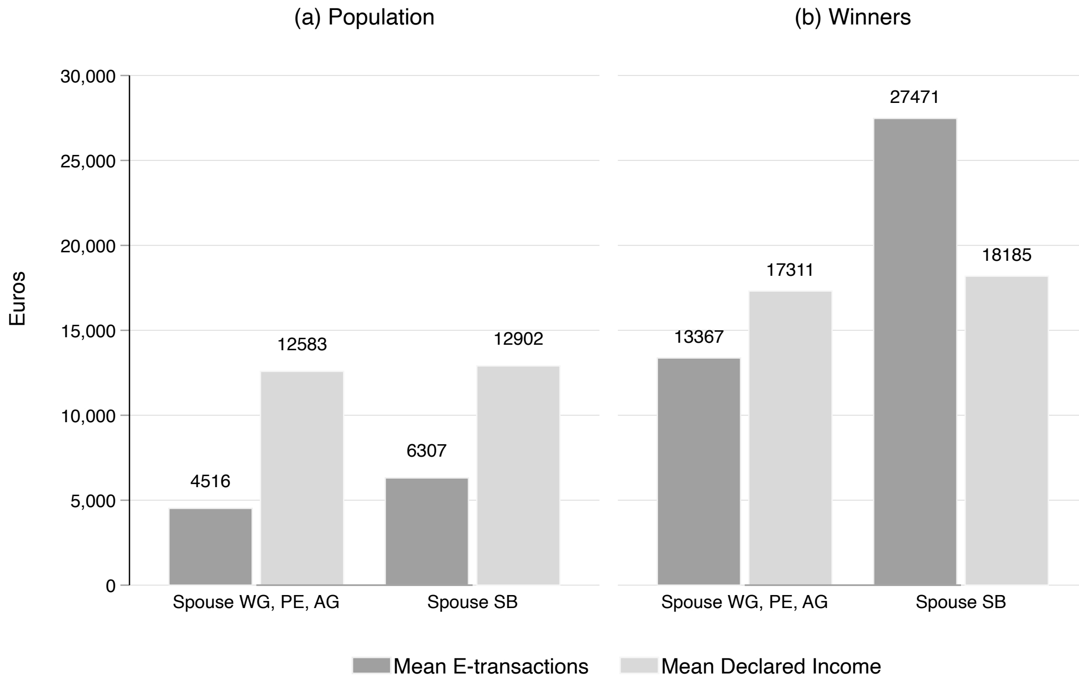
A.1 Descriptives

Fig. 5 Scatter Plots of Annual Income and Electronic Consumption by Income Source



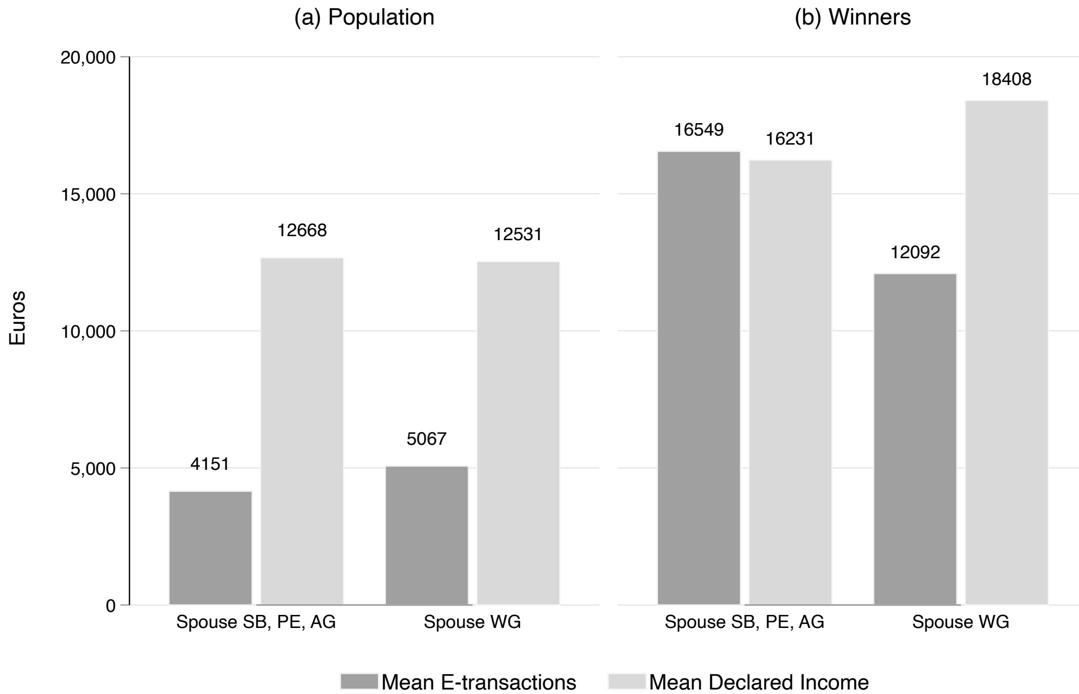
Notes: The figures plots the relationship between the logarithm of annual electronic consumption (y-axis) against the logarithm of annual income (x-axis) for groups with different primary income sources: self-employed (SB) vs other non-zero incomes from wages, pensions and agricultural activities (WG, PE, and AG). Individuals who declare zero income (NO) are excluded from these plots. Right-hand side scatter plots contain only winners. Left-hand side plots contain taxpayer units who have not won in the lottery.

Fig. 6 Annual Income and Electronic Consumption - Taxpayers with SB spouses



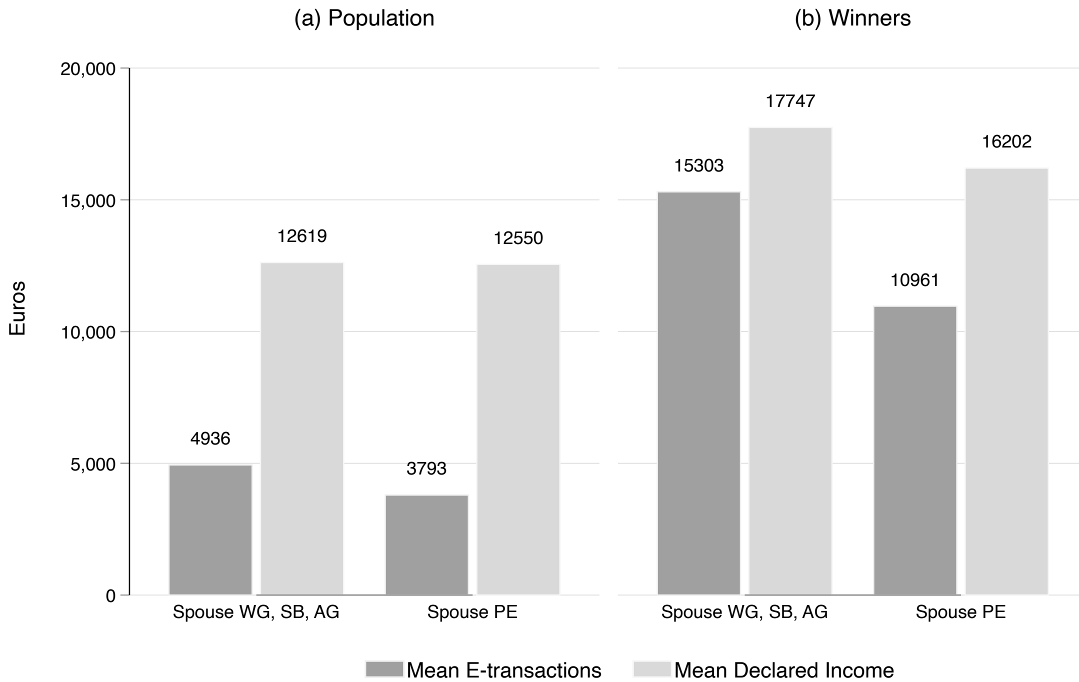
Notes: The figure compares the mean annual income and mean annual electronic consumption of the taxpayer population, figure (a) on the left-hand side and for winners, figure (b) on the right-hand side. The left-hand side columns of each figure include individuals who have a spouse with primary income 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. 7 Annual Income and Electronic Consumption - Taxpayers with WG spouses



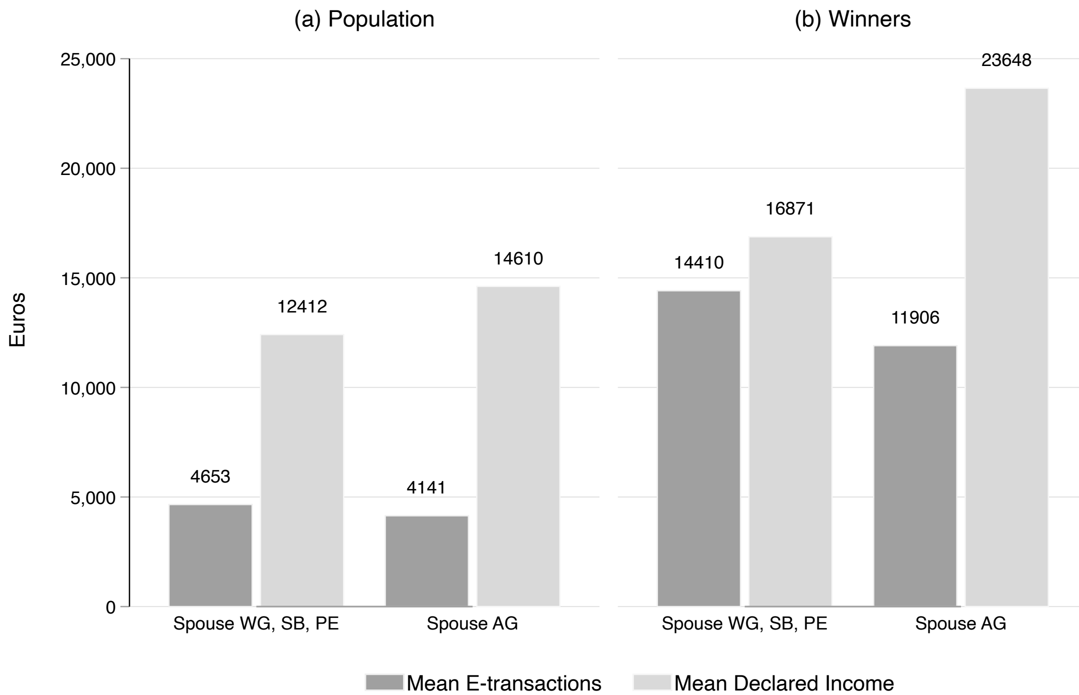
Notes: The figure compares the mean annual income and mean annual electronic consumption of the taxpayer population, figure (a) on the left-hand side and for winners, figure (b) on the right-hand side. The left-hand side columns of each figure include individuals who have a spouse with primary income 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. 8 Annual Income and Electronic Consumption - Taxpayers with PE spouses



Notes: The figure compares the mean annual income and mean annual electronic consumption 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.

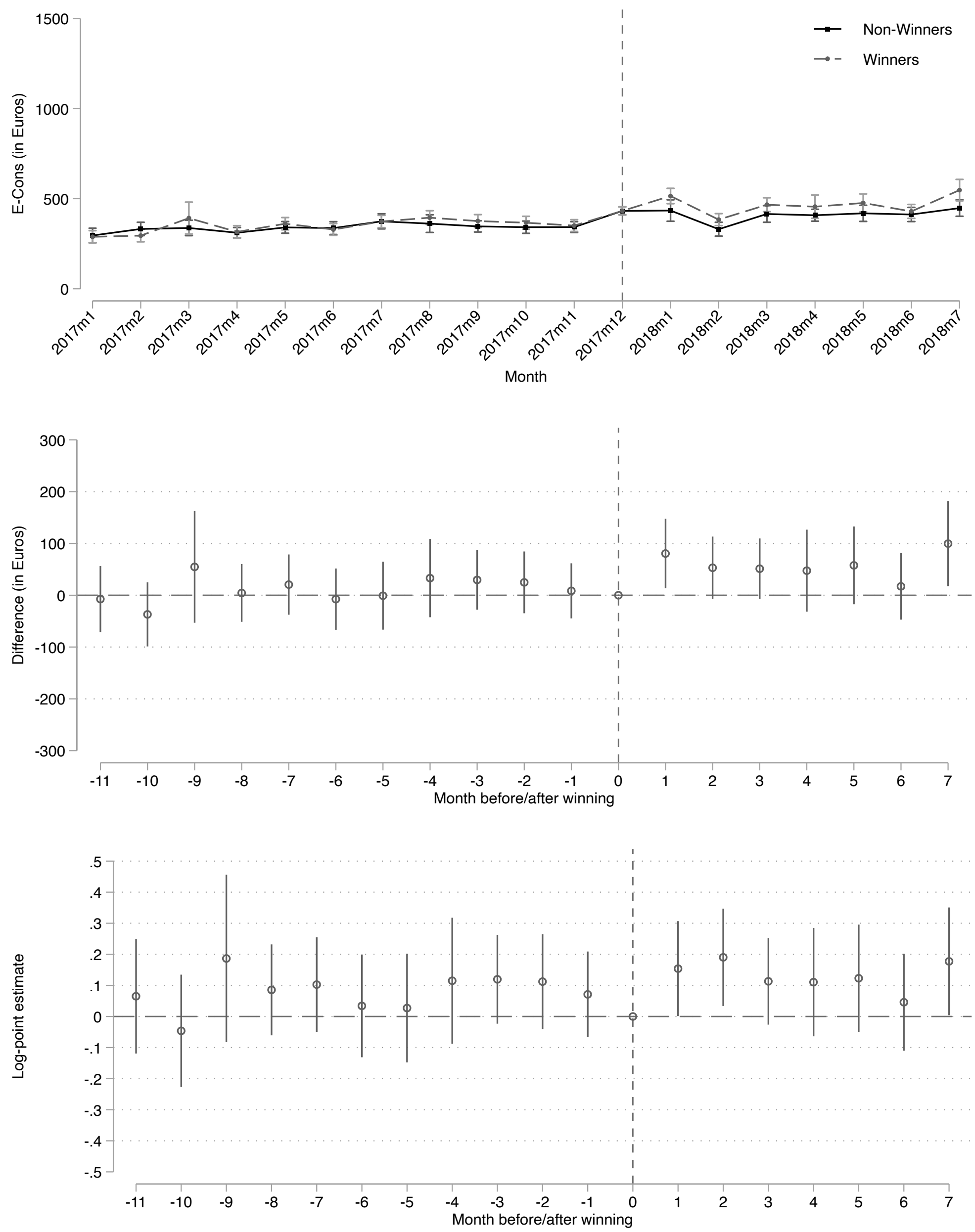
Fig. 9 Annual Income and Electronic Consumption - Taxpayers with AG spouses



Notes: The figure compares the mean annual income and mean annual electronic consumption 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-consumption as shown in Fig. 4. NO income category and single filings are excluded from the sample.

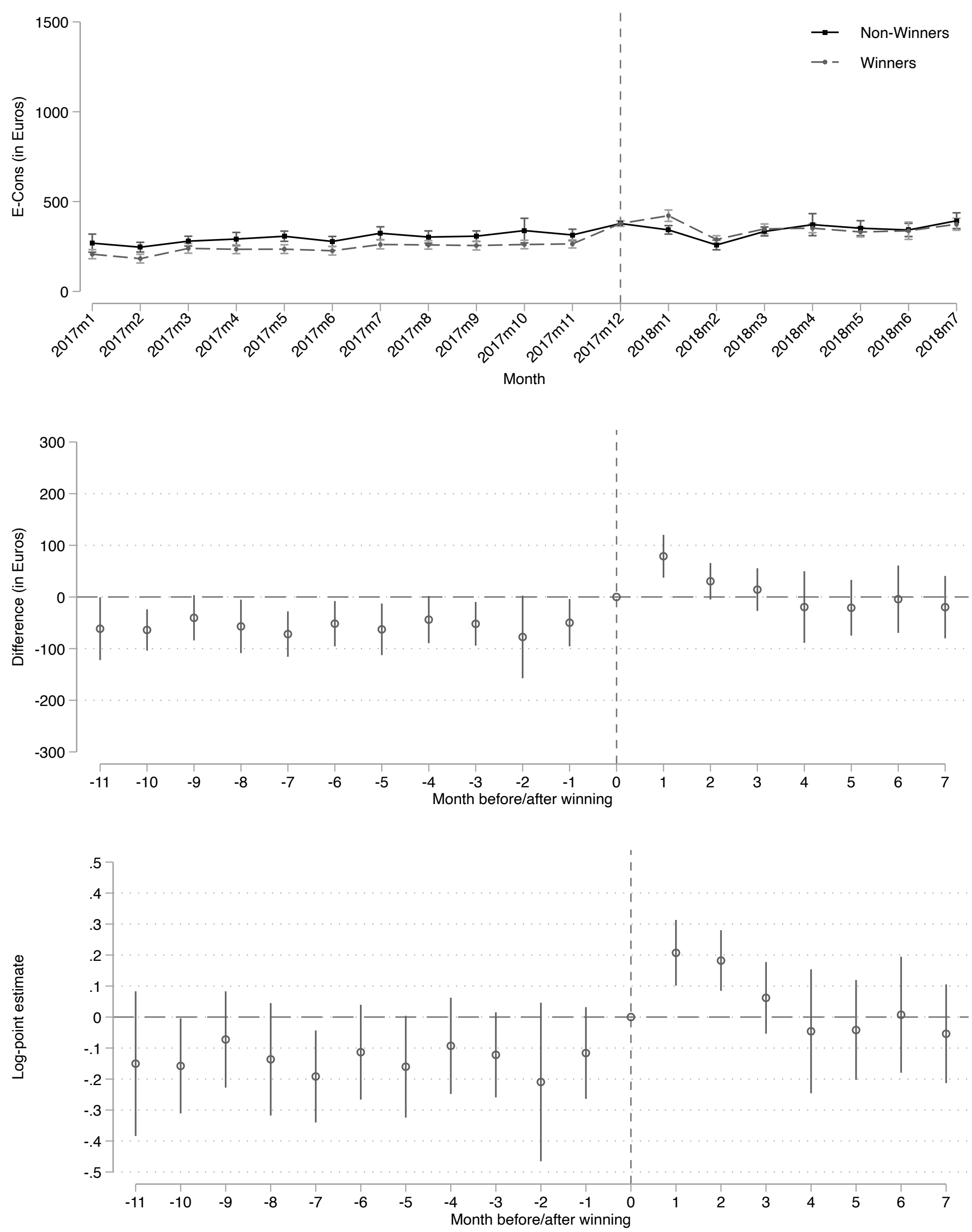
A.2 Event Studies

Fig. 10 First Income Quantile



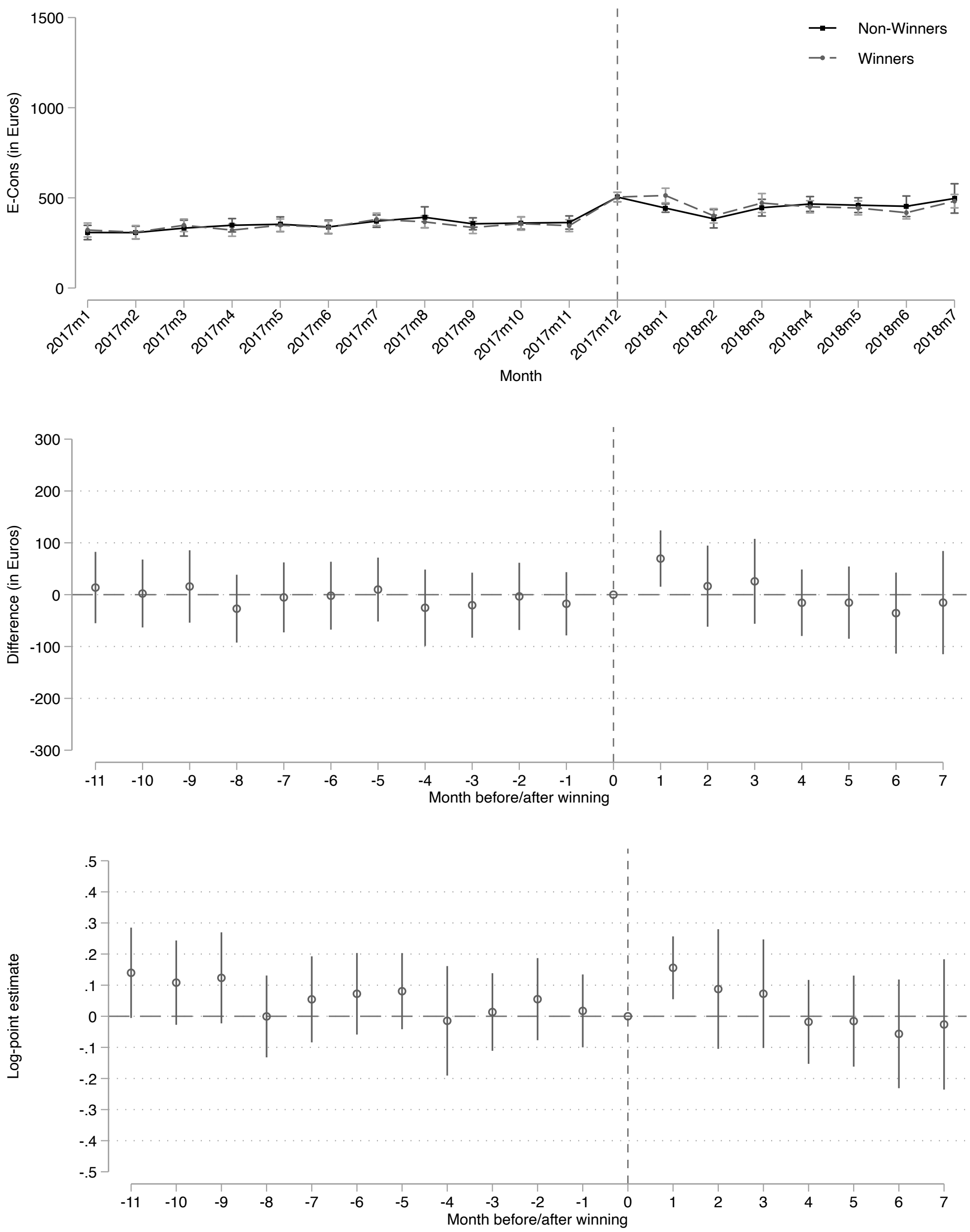
Notes: The figure presents the evolution of electronic consumption from January 2017 to July 2018 (top graph), monthly differences in euros (middle graph), and log-point differences (bottom graph) between winners and non-winners, who belong in the first income quantile. The line at time 0 indicates December 2017, when the superdraw took place. The estimates are obtained from fitted values in a linear regression (for the top and middle graphs) and with Poisson pseudo maximum likelihood (bottom graph). Estimates are shown in Column (1) of Table 7 for the former and in Table 8 for the latter. Confidence intervals are drawn at the 99% level. Robust standard errors are clustered at the individual level.

Fig. 11 Second Income Quantile



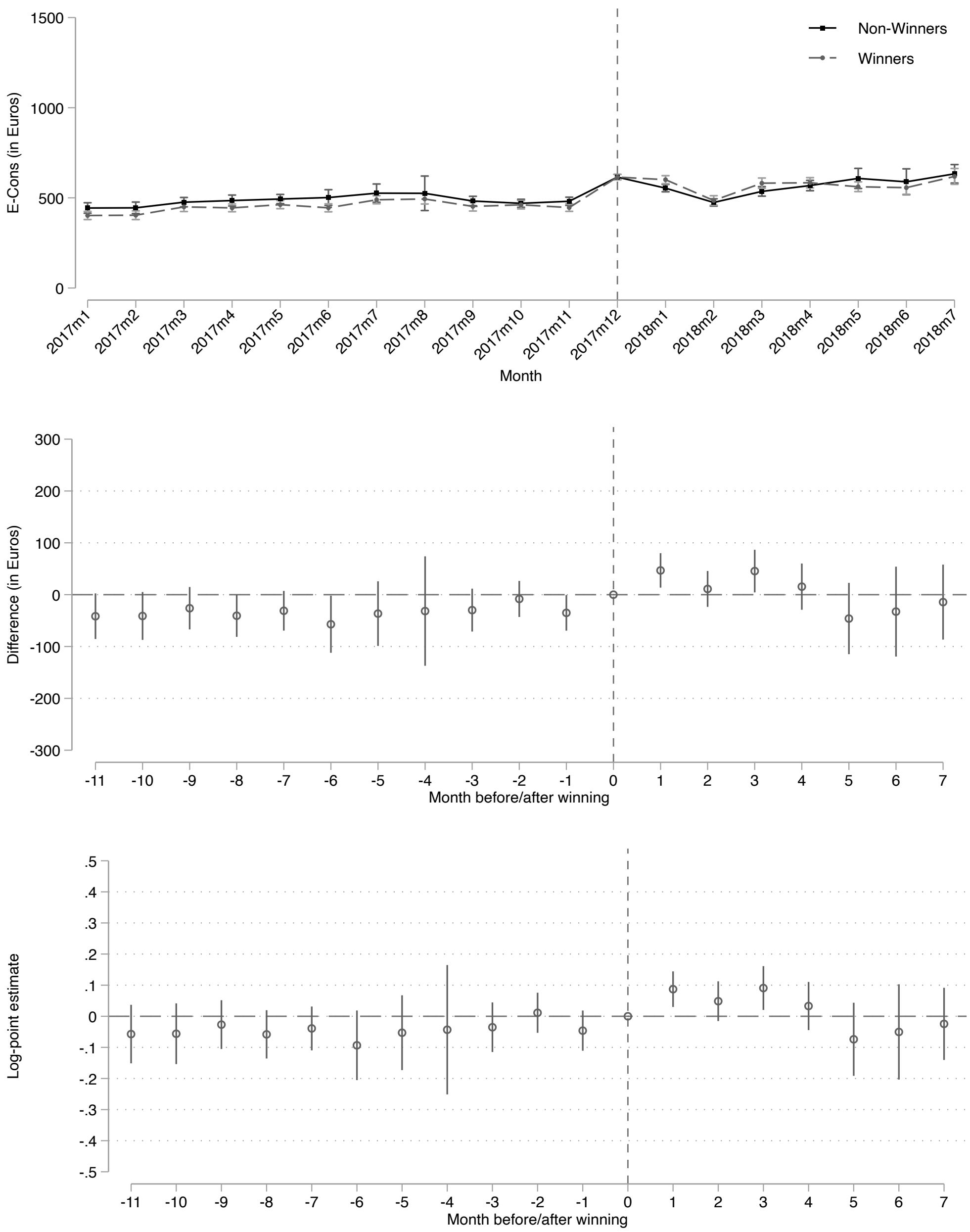
Notes: The figure presents the evolution of electronic consumption from January 2017 to July 2018 (top graph), monthly differences in euros (middle graph), and log-point differences (bottom graph) between winners and non-winners, who belong in the second income quantile. The line at time 0 indicates December 2017, when the superdraw took place. The estimates are obtained from fitted values in a linear regression (for the top and middle graphs) and with Poisson pseudo maximum likelihood (bottom graph). Estimates are shown in Column (2) of Table 7 for the former and in Table 8 for the latter. Confidence intervals are drawn at the 99% level. Robust standard errors are clustered at the individual level.

Fig. 12 Third Income Quantile



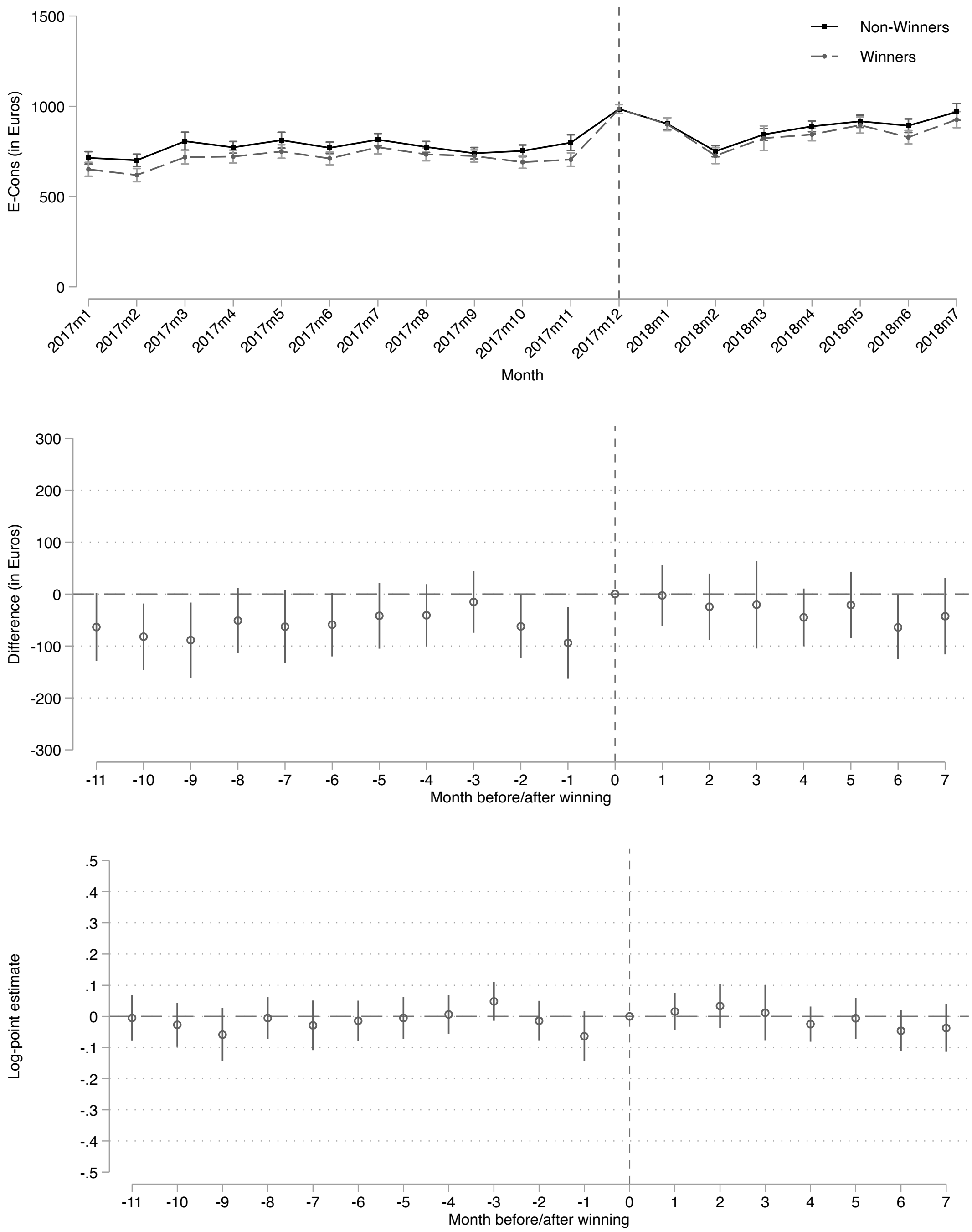
Notes: The figure presents the evolution of electronic consumption from January 2017 to July 2018 (top graph), monthly differences in euros (middle graph), and log-point differences (bottom graph) between winners and non-winners, who belong in the third income quantile. The line at time 0 indicates December 2017, when the superdraw took place. The estimates are obtained from fitted values in a linear regression (for the top and middle graphs) and with Poisson pseudo maximum likelihood (bottom graph). Estimates are shown in Column (3) of Table 7 for the former and in Table 8 for the latter. Confidence intervals are drawn at the 99% level. Robust standard errors are clustered at the individual level.

Fig. 13 Fourth Income Quantile



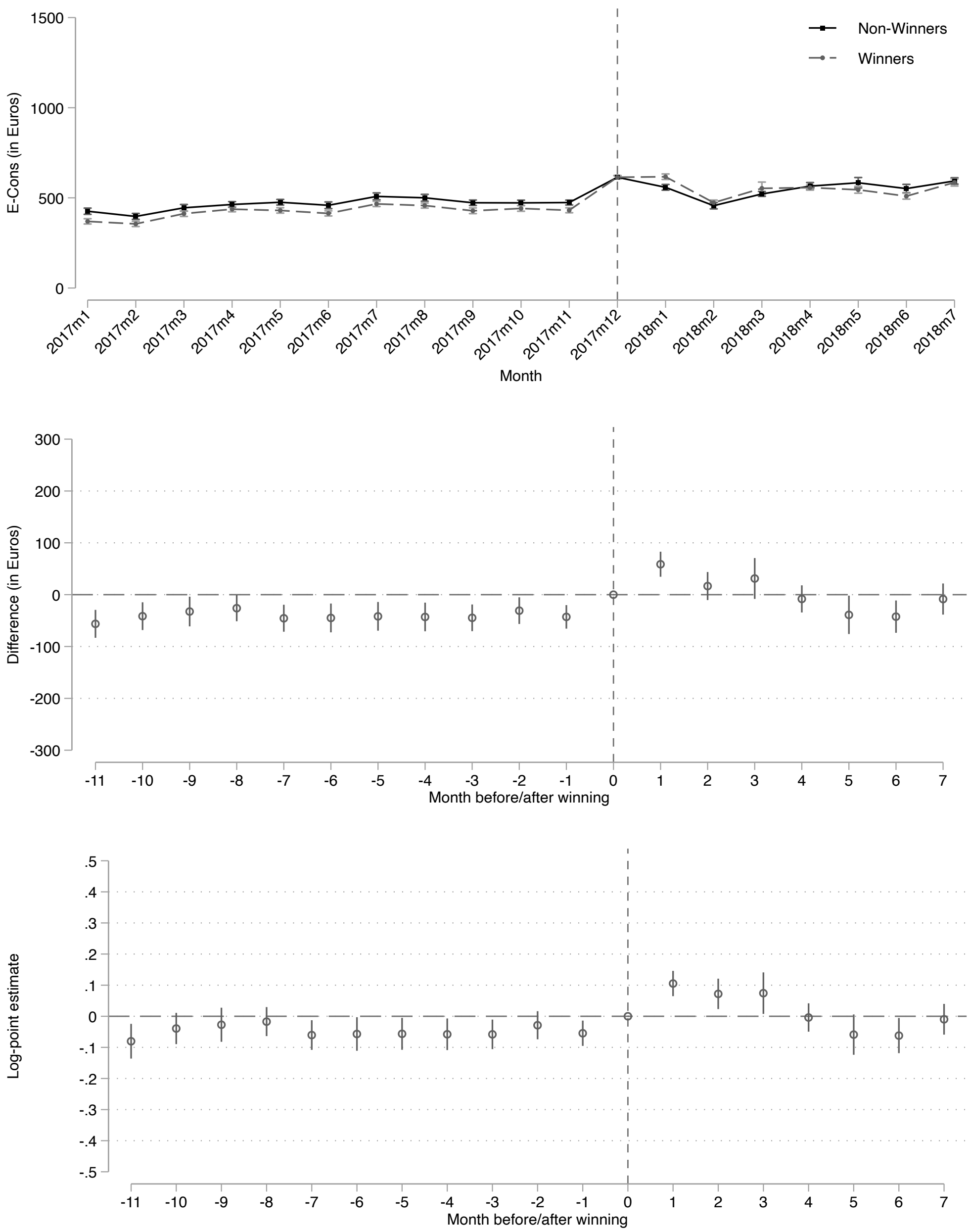
Notes: The figure presents the evolution of electronic consumption from January 2017 to July 2018 (top graph), monthly differences in euros (middle graph), and log-point differences (bottom graph) between winners and non-winners, who belong in the fourth income quantile. The line at time 0 indicates December 2017, when the superdraw took place. The estimates are obtained from fitted values in a linear regression (for the top and middle graphs) and with Poisson pseudo maximum likelihood (bottom graph). Estimates are shown in Column (4) of Table 7 for the former and in Table 8 for the latter. Confidence intervals are drawn at the 99% level. Robust standard errors are clustered at the individual level.

Fig. 14 Fifth Income Quantile



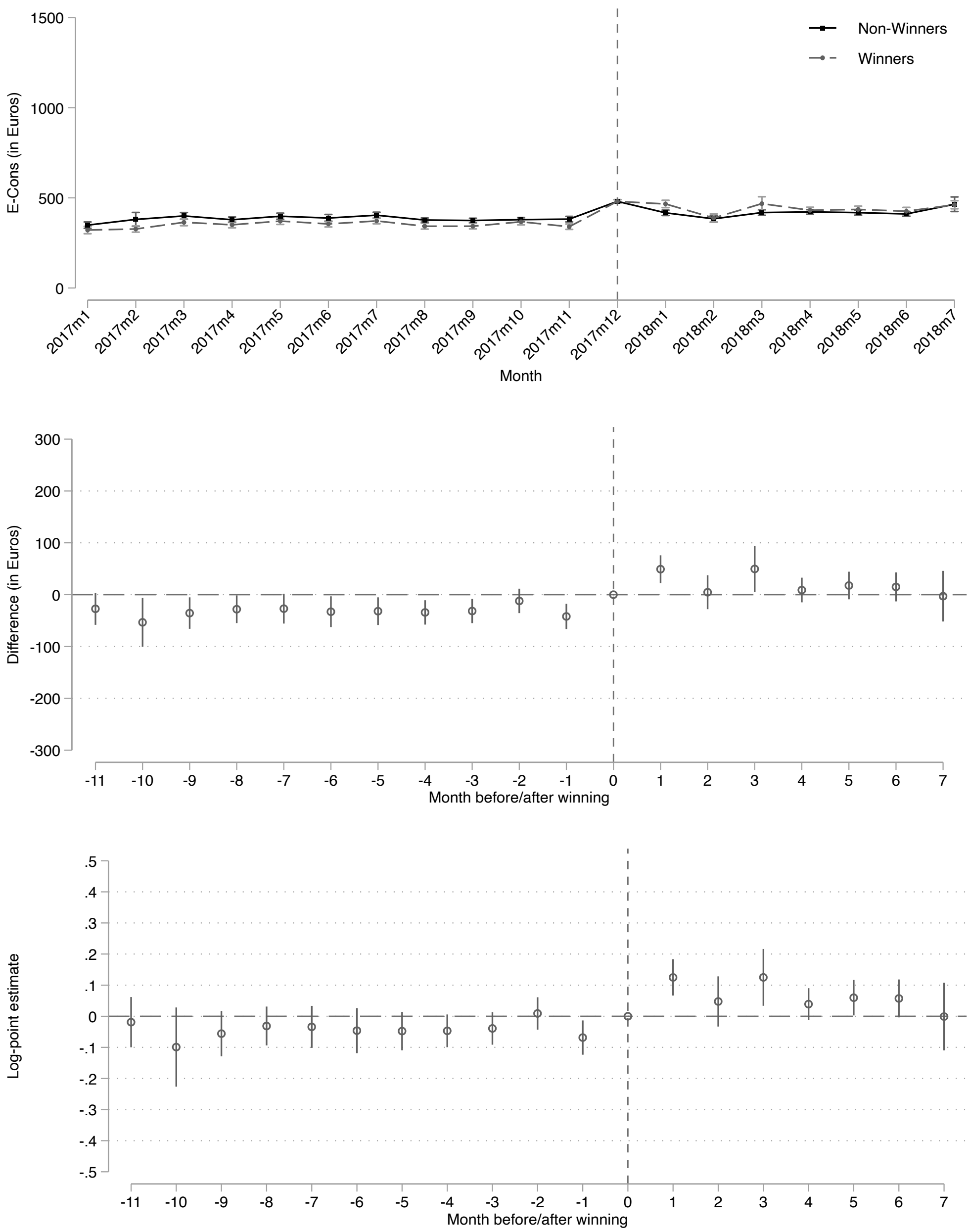
Notes: The figure presents the evolution of electronic consumption from January 2017 to July 2018 (top graph), monthly differences in euros (middle graph), and log-point differences (bottom graph) between winners and non-winners, who belong in the fifth income quantile. The line at time 0 indicates December 2017, when the superdraw took place. The estimates are obtained from fitted values in a linear regression (for the top and middle graphs) and with Poisson pseudo maximum likelihood (bottom graph). Estimates are shown in Column (5) of Table 7 for the former and in Table 8 for the latter. Confidence intervals are drawn at the 99% level. Robust standard errors are clustered at the individual level.

Fig. 15 Wage Earners



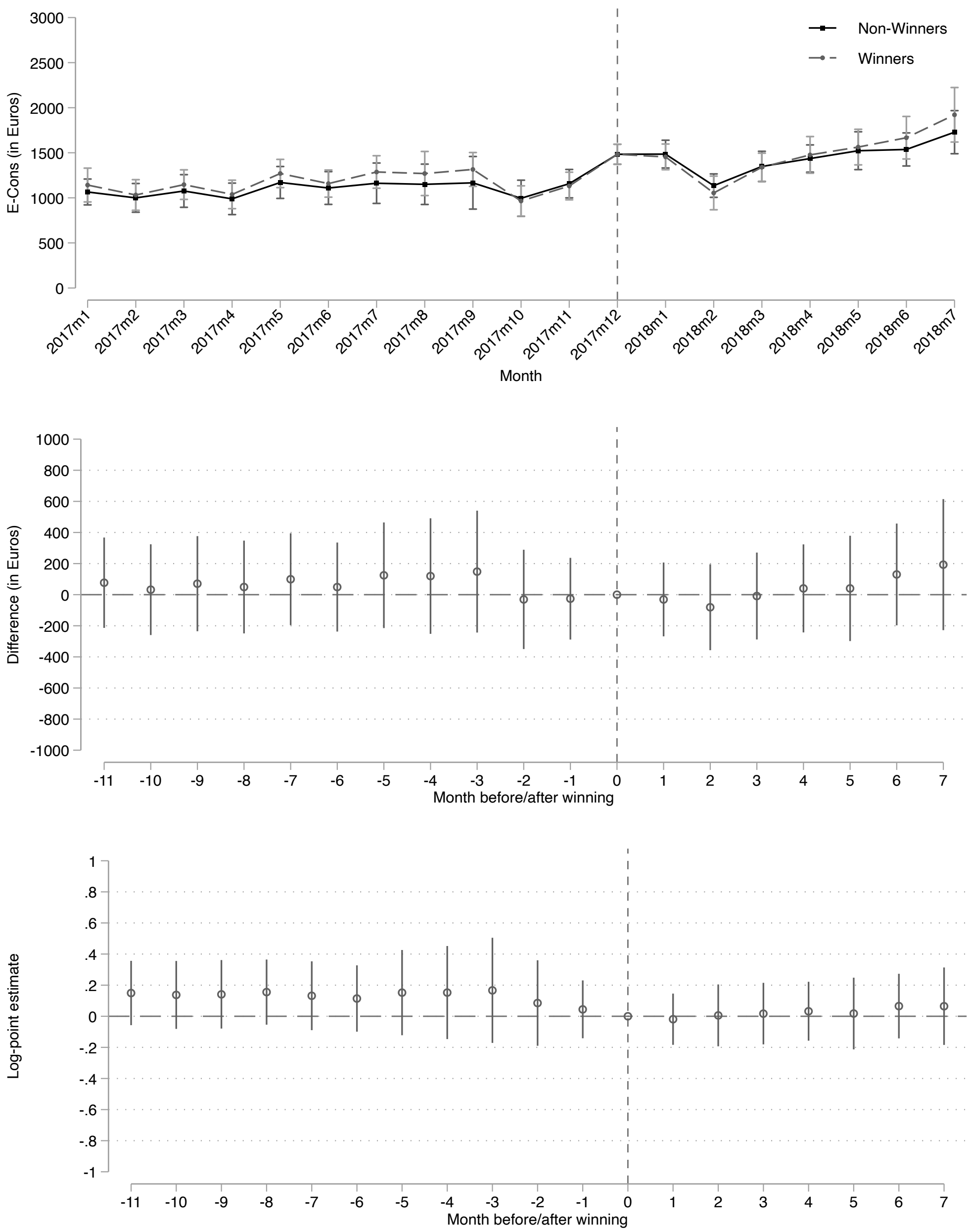
Notes: The figure presents the evolution of electronic consumption from January 2017 to July 2018 (top graph), monthly differences in euros (middle graph), and log-point differences (bottom graph) between winners and non-winners, who are wage-earners (WG). The line at time 0 indicates December 2017, when the superdraw took place. The estimates are obtained from fitted values in a linear regression (for the top and middle graphs) and with Poisson pseudo maximum likelihood (bottom graph). Estimates are shown in Column (1) of Table 9 for the former and in Table 10 for the latter. Confidence intervals are drawn at the 99% level. Robust standard errors are clustered at the individual level.

Fig. 16 Pensioners



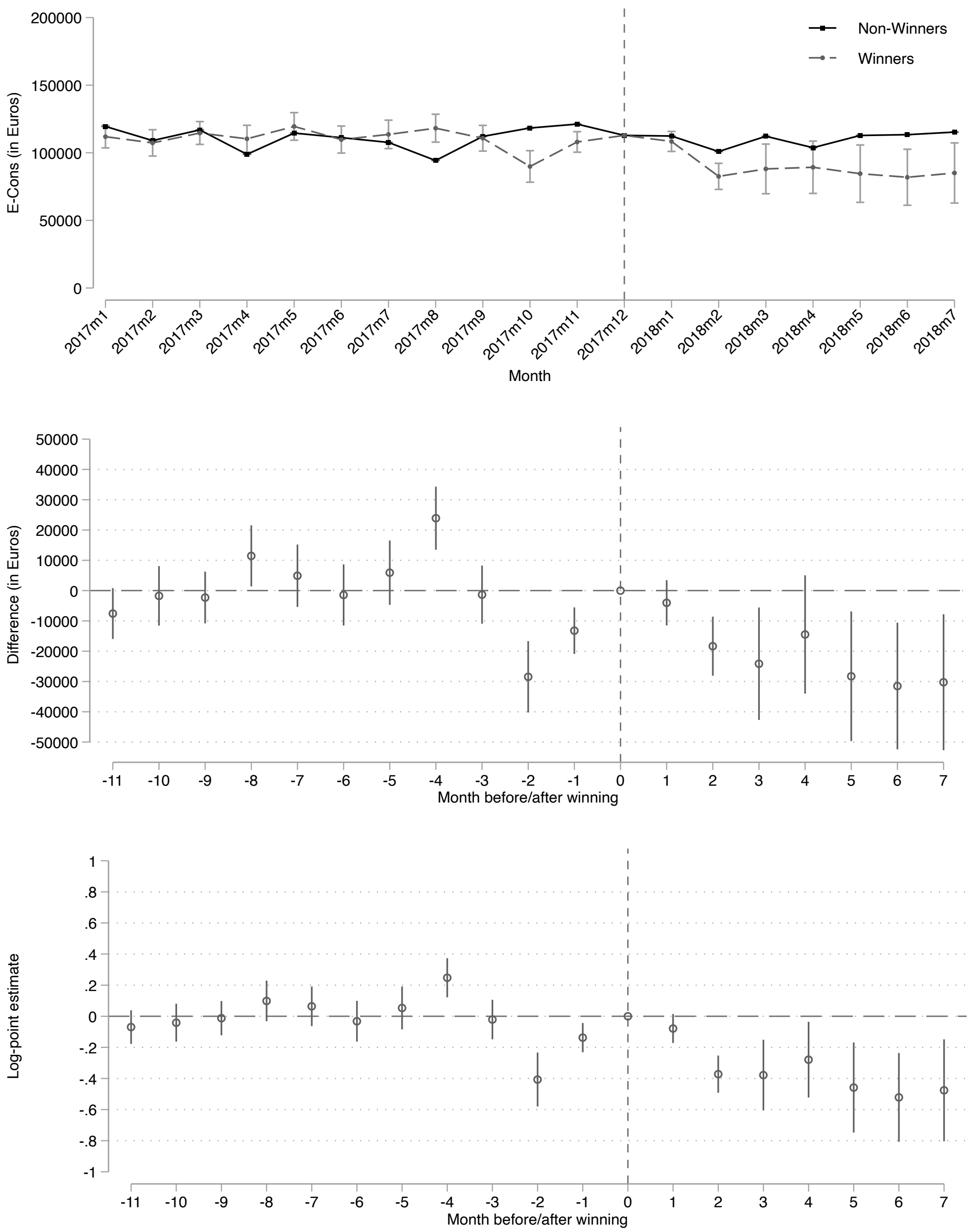
Notes: The figure presents the evolution of electronic consumption from January 2017 to July 2018 (top graph), monthly differences in euros (middle graph), and log-point differences (bottom graph) between winners and non-winners, who are pensioners (PE). The line at time 0 indicates December 2017, when the superdraw took place. The estimates are obtained from fitted values in a linear regression (for the top and middle graphs) and with Poisson pseudo maximum likelihood (bottom graph). Estimates are shown in Column (2) of Table 9 for the former and in Table 10 for the latter. Confidence intervals are drawn at the 99% level. Robust standard errors are clustered at the individual level.

Fig. 17 Self-Employed (without High Electronic Consumption)



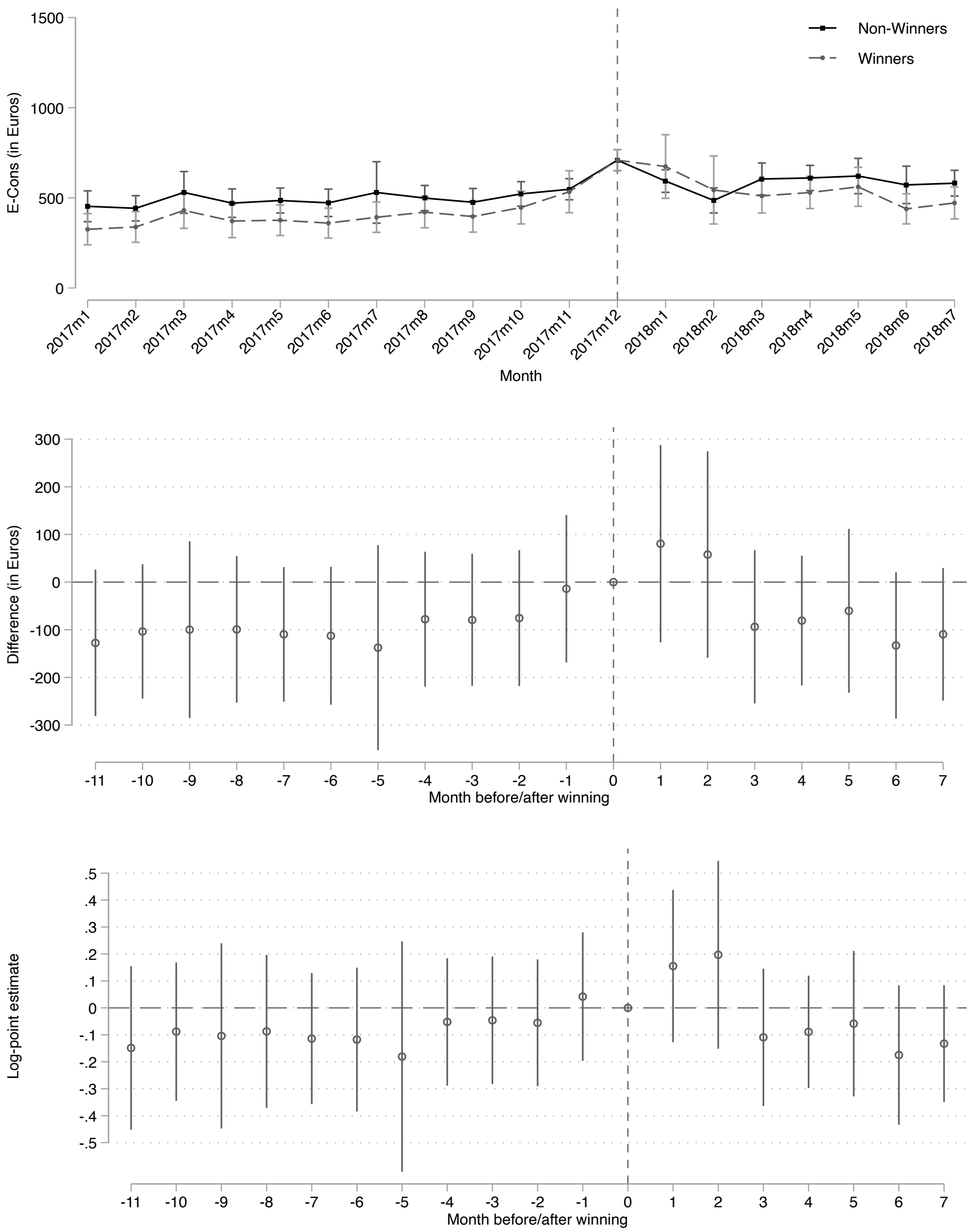
Notes: The figure presents the evolution of electronic consumption from January 2017 to July 2018 (top graph), monthly differences in euros (middle graph), and log-point differences (bottom graph) between winners and non-winners, who are self-employed (SB), but excluding high consumption individuals as explained in Section 5. The line at time 0 indicates December 2017, when the superdraw took place. The estimates are obtained from fitted values in a linear regression (for the top and middle graphs) and with Poisson pseudo maximum likelihood (bottom graph). Estimates are shown in Column (5) of Table 9 for the former and in Table 10 for the latter. Confidence intervals are drawn at the 99% level. Robust standard errors are clustered at the individual level.

Fig. 18 Self-Employed (with High Electronic Consumption)



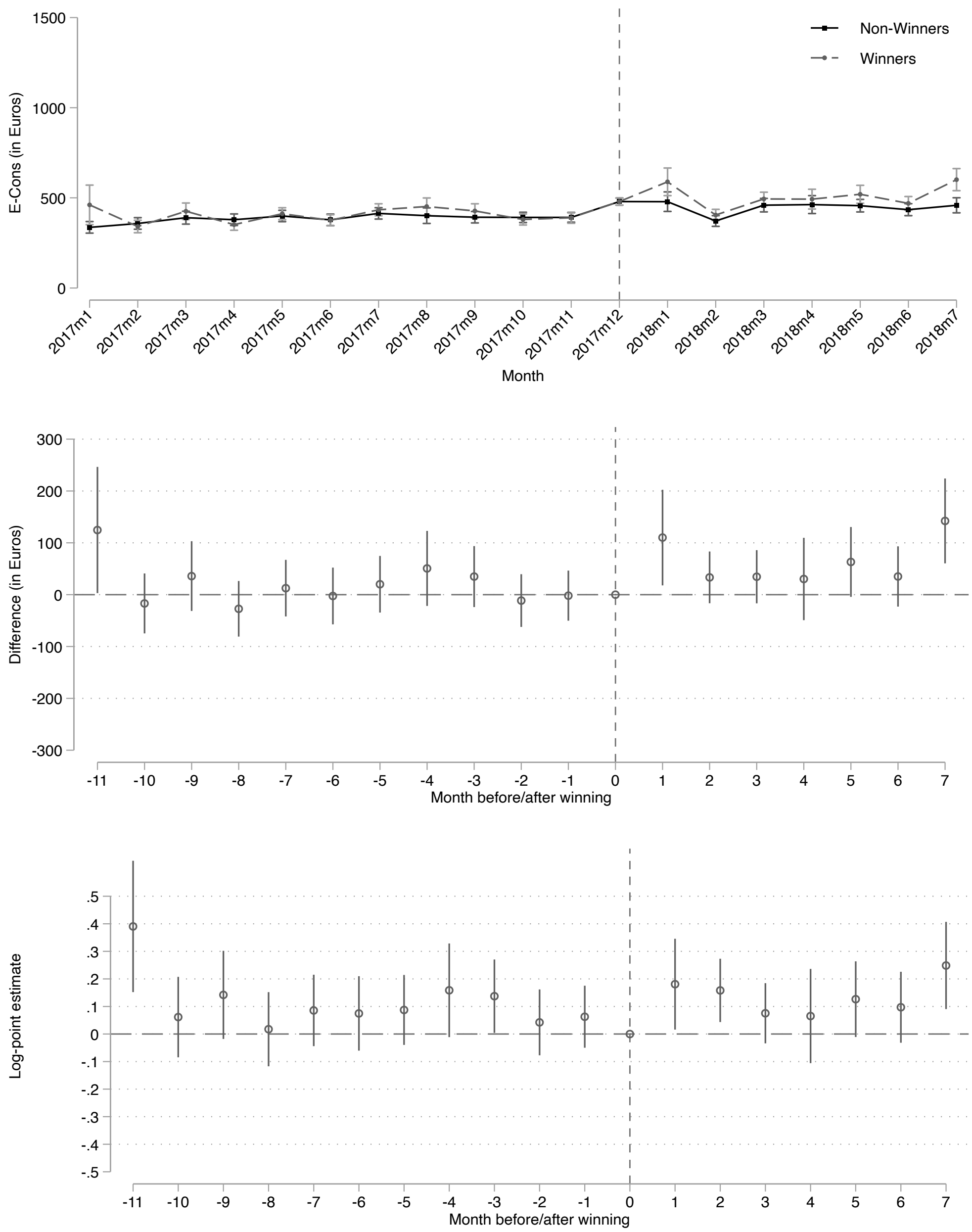
Notes: The figure presents the evolution of electronic consumption from January 2017 to July 2018 (top graph), monthly differences in euros (middle graph), and log-point differences (bottom graph) between winners and non-winners, who are self-employed (SB) with high consumption, as explained in Section 5. The line at time 0 indicates December 2017, when the superdraw took place. The estimates are obtained from fitted values in a linear regression (for the top and middle graphs) and with Poisson pseudo maximum likelihood (bottom graph). Estimates are shown in Column (6) of Table 9 for the former and in Table 10 for the latter. Confidence intervals are drawn at the 99% level. Robust standard errors are clustered at the individual level.

Fig. 19 Agriculture Income



Notes: The figure presents the evolution of electronic consumption from January 2017 to July 2018 (top graph), monthly differences in euros (middle graph), and log-point differences (bottom graph) between winners and non-winners, who have agricultural income (AG). The line at time 0 indicates December 2017, when the superdraw took place. The estimates are obtained from fitted values in a linear regression (for the top and middle graphs) and with Poisson pseudo maximum likelihood (bottom graph). Estimates are shown in Column (3) of Table 9 for the former and in Table 10 for the latter. Confidence intervals are drawn at the 99% level. Robust standard errors are clustered at the individual level.

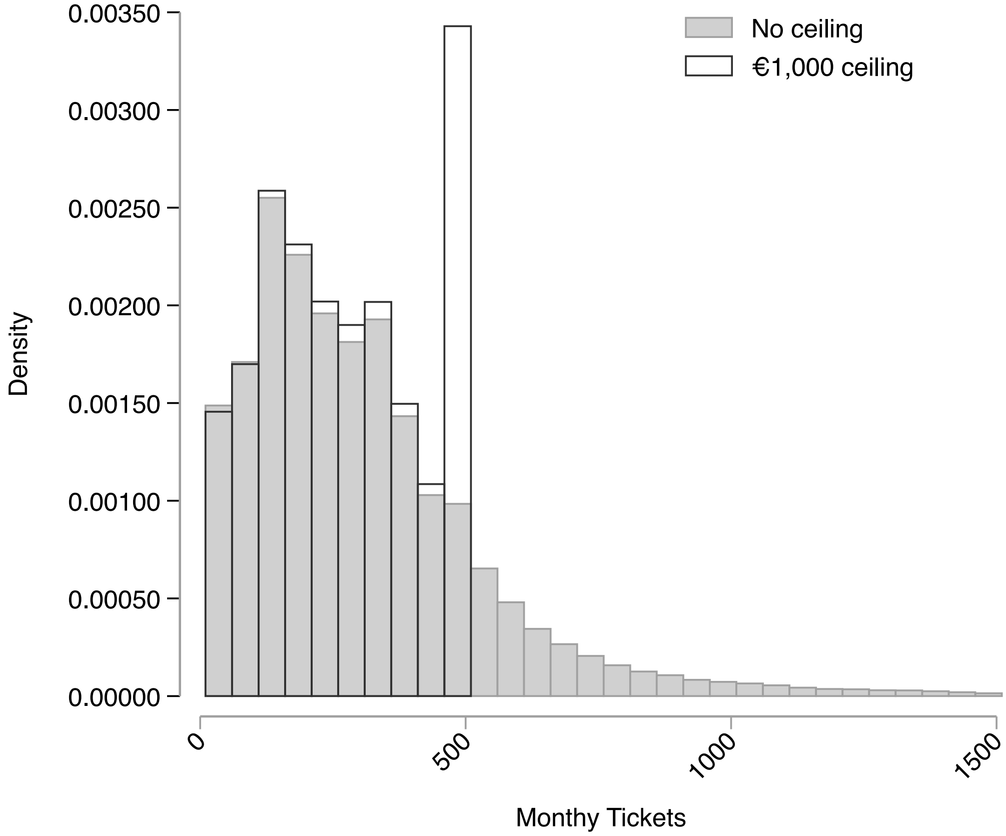
Fig. 20 Zero Income



Notes: The figure presents the evolution of electronic consumption from January 2017 to July 2018 (top graph), monthly differences in euros (middle graph), and log-point differences (bottom graph) between winners and non-winners, who declare zero income (NO). The line at time 0 indicates December 2017, when the superdraw took place. The estimates are obtained from fitted values in a linear regression (for the top and middle graphs) and with Poisson pseudo maximum likelihood (bottom graph). Estimates are shown in Column (4) of Table 9 for the former and in Table 10 for the latter. Confidence intervals are drawn at the 99% level. Robust standard errors are clustered at the individual level.

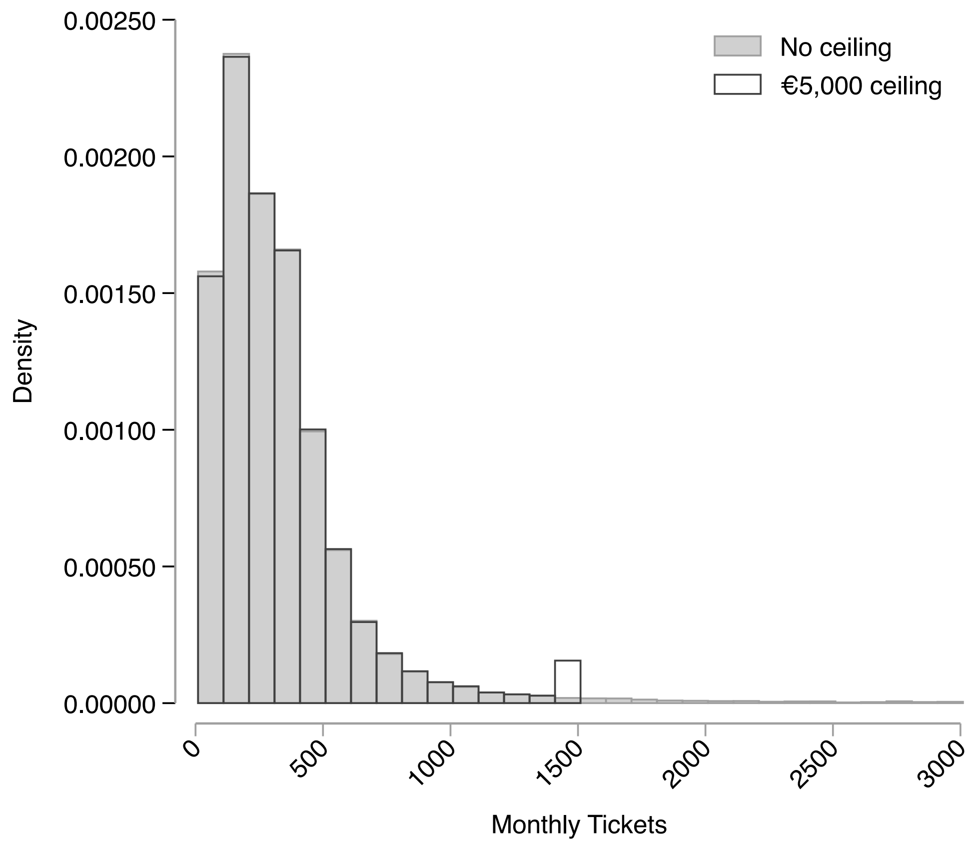
A.3 Simulations

Fig. 21 Winners' Distribution of Tickets in Simulations



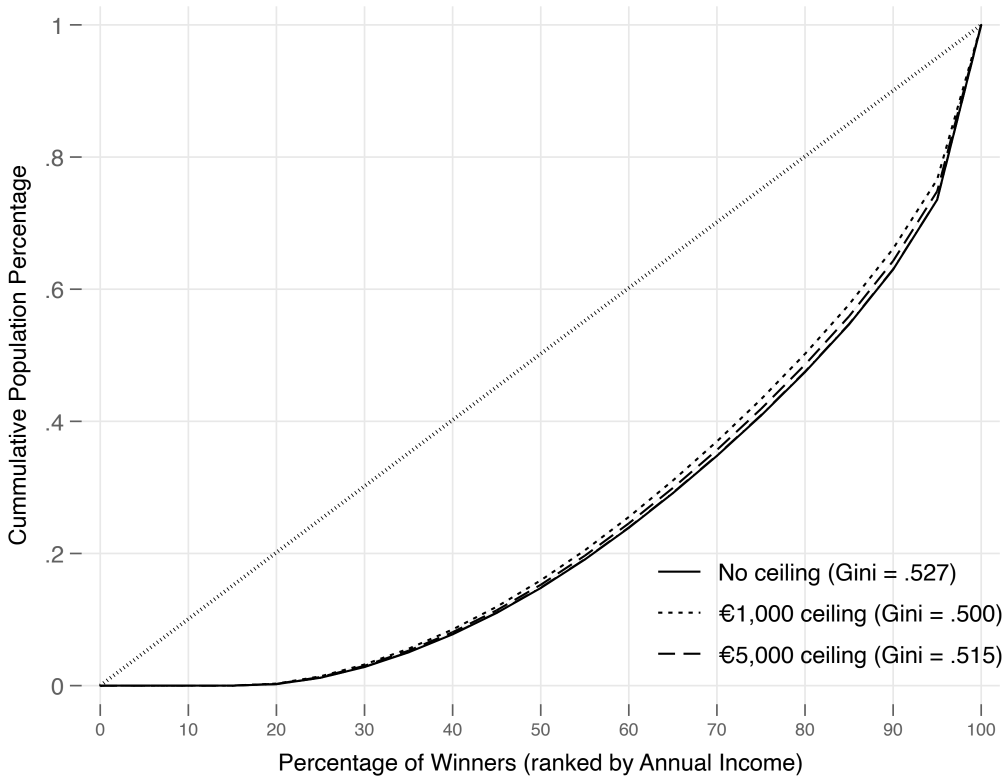
Notes: The figure compares the distribution of winners' tickets in simulations with and without the € 1,000 ceiling. The ceiling translates to a maximum number of 467 monthly tickets per individual. Both simulations contain 1,200 iterations of the lottery (100 for each month in 2017), drawing 1,000 winners in each iteration. Both distributions contain 1.2 million winners. The "no ceiling" distribution is a simulation of the lottery assigning tickets using the ticket-awarding mechanism in Table 13. The "€ 1,000 ceiling" distribution retains introduces a maximum ceiling in monthly tickets. For electronic consumption 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. 22 Winners' Distribution of Tickets in Simulations



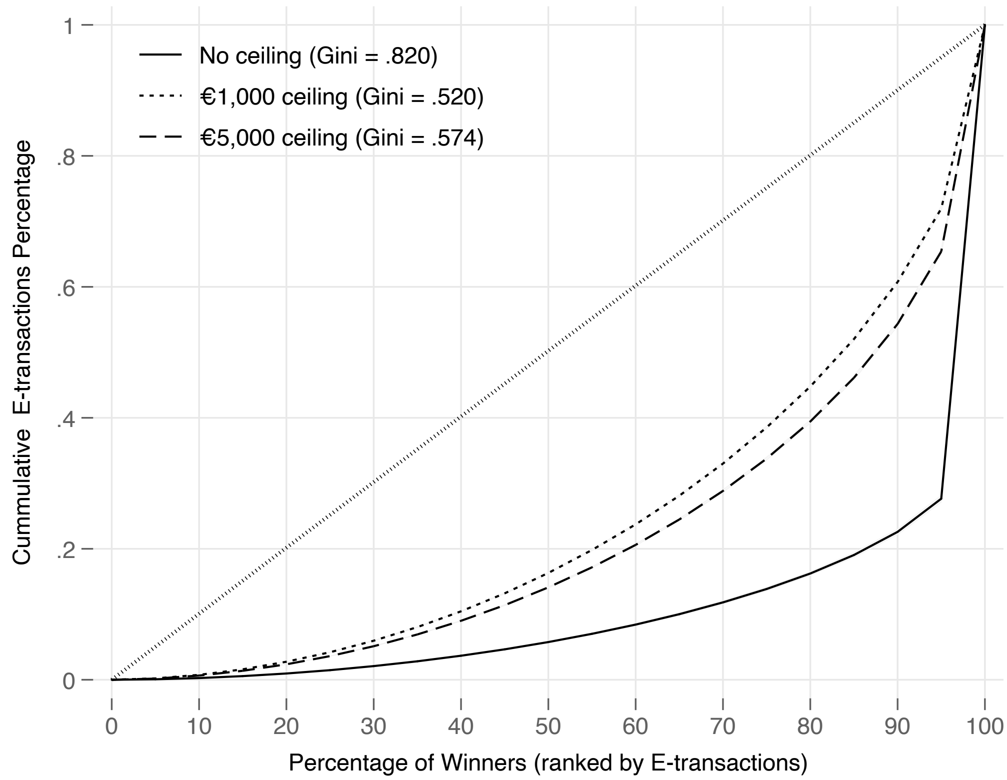
Notes: The figure compares the distribution of winners' tickets in simulations with and without the € 5,000 ceiling. The ceiling translates to a maximum number of 1,467 monthly tickets per individual. Both simulations contain 1,200 iterations of the lottery (100 for each month in 2017), drawing 1,000 winners in each iteration. Both distributions contain 1.2 million winners. The "no ceiling" distribution is a simulation of the lottery assigning tickets using the ticket-awarding mechanism in Table 13. The "€ 5,000 ceiling" distribution retains introduces a maximum ceiling in monthly tickets. For electronic consumption beyond € 5,000 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.

Fig. 23 Income Distribution with Ticket Ceilings



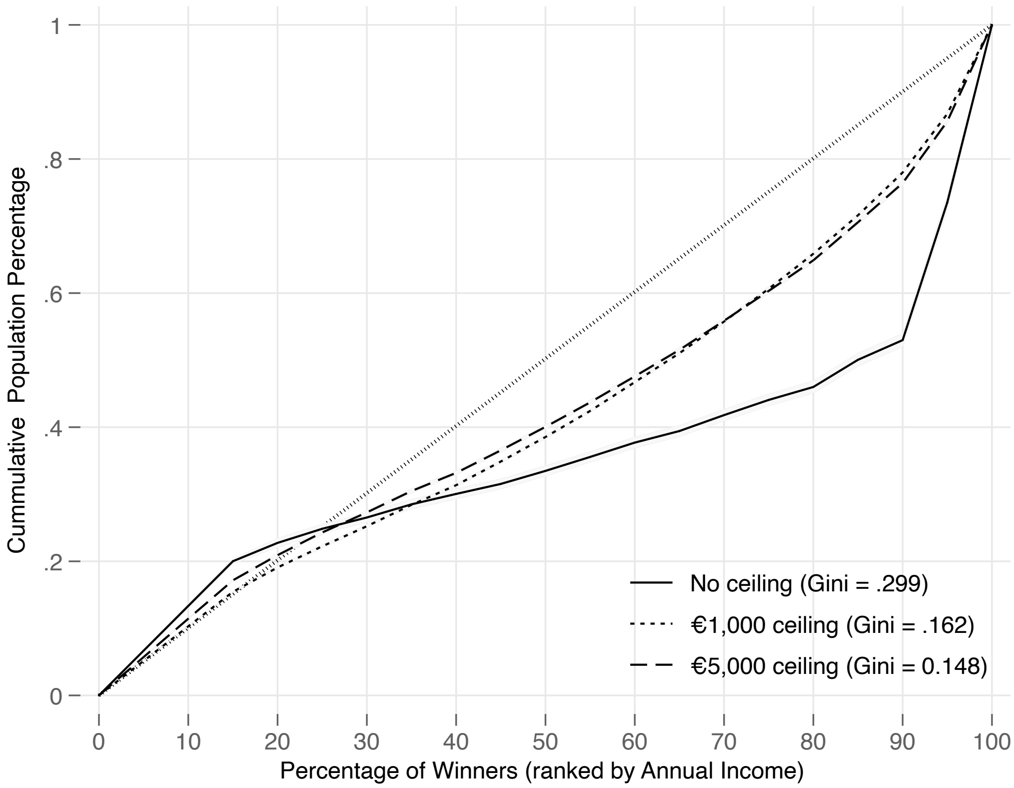
Notes: The figure plots of cumulative income distribution curves for 1.2 million winners in simulations of the lottery. The x-axis represents the population percentiles of winners, ranked by their declared annual income in 2017. The y-axis shows the percentage of individuals who have won the lottery 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 ticket-awarding mechanism in Table 13. The € 1,000 ceiling introduces a maximum ceiling in monthly tickets. For electronic consumption beyond € 1,000 per month no more tickets are awarded to individuals. Similarly, the € 5,000 curve introduces a ceiling at the € 5,000 monthly electronic consumption level.

Fig. 24 Electronic Consumption Distribution with Ticket Ceilings



Notes: The figure plots of cumulative electronic consumption distribution curves for 1.2 million winners in simulations of the lottery. The x-axis represents the population percentiles of winners, ranked by electronic consumption 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 ticket-awarding mechanism in Table 13. The € 1,000 ceiling introduces a maximum ceiling in monthly tickets. For electronic consumption beyond € 1,000 per month no more tickets are awarded to individuals. Similarly, the € 5,000 curve introduces a ceiling at the € 5,000 monthly electronic consumption level.

Fig. 25 Electronic Consumption Distribution (ranked by Income) with Ticket Ceilings



Notes: The figure plots electronic consumption distribution curves for 1.2 million winners in 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 winners' income percentage equals the winners' e-consumption percentage. The "no ceiling" curve is a simulation of the lottery assigning tickets using the ticket-awarding mechanism in Table 13. The € 1,000 ceiling introduces a maximum ceiling in monthly tickets. For electronic consumption beyond € 1,000 per month no more tickets are awarded to individuals. Similarly, the € 5,000 curve introduces a ceiling at the € 5,000 monthly electronic consumption level.

B Tables

B.1 Summary Statistics and Regressions

Table 3 Sample 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	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	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 winners and non-winners 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. The right-hand side columns present the frequencies and percentages of single and joint-filing tax units in each primary income source category. Joint-filing units 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 4 Summary Statistics - Winners and Reconstructed Taxpayer Population

	Winners			*Population*		
	Obs.	Income	E-Cons	Obs.	Income	E-Cons
by Primary Income Category:						
SB	988	20,753	181,520	266,317	12,120	11,420
Self-Employed income	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 income and the mean electronic consumption Z_i in 2017 (nominal € values) winners and the reconstructed taxpayer population. They are presented by primary income source as has been declared in their tax returns: from wages (WG), self-employed (SB), agricultural income (AG), pensions (PE). Additional categories indicates zero-declared income (NO) and no filing. Standard deviations are in parenthesis. The 'Winners' sample includes all individual winners from 2017 draws. The 'Population' sample is a reconstructed sample of the taxpayer population (see Section 3).

Table 5 Non-parametric Estimates, by 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-consumption						
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 income category. Results on the top table use annual income as independent variable and at the bottom, annual electronic consumption. The NO and No Filing categories in columns (5) and (6) do not record results for annual income regressions, since no income was declared. Robust standard errors (clustered at the individual level, depending on the number in each income category) in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 6 Estimates - Spouse's Income Source

	(1) E-Cons	(2) E-Cons	(3) E-Cons	(4) E-Cons	(5) E-Cons	(6) E-Cons
SB Spouse (primary source)	2,006***		1,778***	1,781***	1,791***	1,747***
	(317)		(320)	(256)	(268)	(256)
SB Spouse (any SB income)		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 taxpayers with SB spouses. The sample is restricted to 2,406,971 individuals in the population who filed jointly, shown 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.

B.2 Event Study Estimates

Table 7 Event Studies by Income Quantile (Linear Fixed Effects Estimation)

	1st Quantile	2nd Quantile	3rd Quantile	4th Quantile	5th Quantile
	(1)	(2)	(3)	(4)	(5)
	E-cons	E-cons	E-cons	E-cons	E-cons
Winners Interaction ×					
January 2017	-7.381	-61.46**	13.79	-41.55*	-63.45*
	(32.47)	(31.00)	(35.08)	(22.43)	(33.47)
February 2017	-36.97	-63.79***	2.256	-41.05*	-82.02**
	(31.58)	(20.45)	(33.43)	(23.55)	(32.58)
March 2017	54.74	-40.35*	15.81	-26.13	-88.64**
	(55.04)	(22.30)	(35.62)	(20.88)	(36.87)
April 2017	4.401	-56.95**	-27.00	-40.63*	-51.01
	(28.44)	(26.40)	(33.38)	(20.75)	(31.99)
May 2017	20.53	-71.82***	-5.216	-31.05	-62.76*
	(29.69)	(22.42)	(34.43)	(19.53)	(35.79)
June 2017	-7.578	-51.69**	-1.961	-57.02**	-58.89*
	(30.19)	(22.37)	(33.41)	(28.10)	(31.20)
July 2017	-0.902	-62.70**	9.870	-36.52	-41.87
	(33.46)	(25.47)	(31.44)	(31.80)	(32.29)
August 2017	33.09	-43.83*	-25.26	-31.57	-40.87
	(38.53)	(23.28)	(37.61)	(53.84)	(30.55)
September 2017	29.56	-52.01**	-20.28	-29.76	-15.12
	(29.30)	(21.55)	(31.97)	(21.09)	(30.30)
October 2017	24.83	-77.39*	-3.307	-8.225	-62.29**
	(30.42)	(40.77)	(33.07)	(17.75)	(31.10)
November 2017	8.454	-49.84**	-17.55	-35.08**	-93.95***
	(27.13)	(23.34)	(31.13)	(17.50)	(35.26)
January 2018	80.56**	78.86***	69.68**	46.78***	-2.676
	(34.20)	(21.16)	(27.71)	(16.95)	(29.84)
February 2018	53.09*	30.45*	16.44	11.07	-24.39
	(30.69)	(17.99)	(39.91)	(17.65)	(32.60)
March 2018	51.26*	14.35	25.80	45.50**	-20.45
	(29.78)	(21.07)	(41.79)	(20.97)	(43.03)
April 2018	47.46	-19.50	-15.55	15.48	-44.89
	(40.37)	(35.43)	(32.73)	(22.73)	(28.35)
May 2018	57.69	-20.90	-15.31	-46.03	-21.07
	(38.30)	(27.53)	(35.55)	(35.07)	(32.66)
June 2018	17.15	-4.339	-35.54	-32.65	-64.07**
	(32.79)	(33.24)	(39.83)	(44.23)	(31.30)
July 2018	99.66**	-19.70	-15.25	-14.32	-42.75
	(41.91)	(30.80)	(50.79)	(36.90)	(37.43)
Constant	432.5***	378.0***	504.5***	615.1***	985.1***
	(11.38)	(7.522)	(13.47)	(7.691)	(12.57)
Individual FE	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes
Number of obs.	195373	194974	203277	212024	229192
Number of Individuals	10284	10263	10700	11160	12065

Notes: The table presents monthly event study estimates from Regression 4 for each of the income quantiles. A linear regression is used. Estimates present differences in monthly electronic consumption between winners and non-winners. Results are plotted in the top and middle graphs of Figures 10 to 14 for each quantile, respectively. The regressions use inverse probability weights generated from the propensity scores as shown in Section 5. Robust standard errors are clustered at the individual level.

Table 8 Event Studies by Income Quantile (Poisson Maximum Likelihood Estimation)

	1st Quantile (1) E-cons	2nd Quantile (2) E-cons	3rd Quantile (3) E-cons	4th Quantile (4) E-cons	5th Quantile (5) E-cons
Winners Interaction ×					
January 2017	0.0652	-0.150	0.140*	-0.0572	-0.00521
	(0.0941)	(0.119)	(0.0741)	(0.0480)	(0.0374)
February 2017	-0.0462	-0.158**	0.108	-0.0562	-0.0268
	(0.0922)	(0.0781)	(0.0692)	(0.0498)	(0.0361)
March 2017	0.187	-0.0724	0.124*	-0.0268	-0.0587
	(0.137)	(0.0793)	(0.0747)	(0.0400)	(0.0438)
April 2017	0.0856	-0.136	-0.000520	-0.0582	-0.00520
	(0.0746)	(0.0925)	(0.0671)	(0.0396)	(0.0340)
May 2017	0.103	-0.192**	0.0544	-0.0391	-0.0288
	(0.0776)	(0.0757)	(0.0706)	(0.0360)	(0.0407)
June 2017	0.0341	-0.113	0.0724	-0.0933	-0.0142
	(0.0843)	(0.0780)	(0.0669)	(0.0571)	(0.0330)
July 2017	0.0272	-0.160*	0.0807	-0.0528	-0.00512
	(0.0893)	(0.0837)	(0.0624)	(0.0613)	(0.0341)
August 2017	0.115	-0.0929	-0.0146	-0.0433	0.00643
	(0.103)	(0.0792)	(0.0898)	(0.106)	(0.0316)
September 2017	0.120	-0.122*	0.0137	-0.0352	0.0482
	(0.0729)	(0.0700)	(0.0636)	(0.0407)	(0.0317)
October 2017	0.112	-0.210	0.0550	0.0113	-0.0142
	(0.0780)	(0.131)	(0.0674)	(0.0329)	(0.0328)
November 2017	0.0712	-0.116	0.0173	-0.0462	-0.0637
	(0.0703)	(0.0755)	(0.0598)	(0.0329)	(0.0408)
January 2018	0.154**	0.207***	0.156***	0.0871***	0.0154
	(0.0779)	(0.0541)	(0.0515)	(0.0292)	(0.0306)
February 2018	0.190**	0.182***	0.0876	0.0485	0.0333
	(0.0799)	(0.0499)	(0.0981)	(0.0326)	(0.0356)
March 2018	0.113	0.0619	0.0725	0.0908**	0.0115
	(0.0711)	(0.0589)	(0.0891)	(0.0359)	(0.0457)
April 2018	0.111	-0.0459	-0.0180	0.0330	-0.0248
	(0.0890)	(0.102)	(0.0688)	(0.0395)	(0.0288)
May 2018	0.123	-0.0418	-0.0155	-0.0740	-0.00605
	(0.0880)	(0.0822)	(0.0747)	(0.0599)	(0.0335)
June 2018	0.0459	0.00750	-0.0566	-0.0503	-0.0460
	(0.0796)	(0.0955)	(0.0891)	(0.0782)	(0.0334)
July 2018	0.178**	-0.0540	-0.0263	-0.0245	-0.0375
	(0.0884)	(0.0812)	(0.107)	(0.0592)	(0.0388)
Constant	6.934***	6.631***	6.720***	6.841***	7.243***
	(0.0259)	(0.0224)	(0.0265)	(0.0137)	(0.0124)
Individual FE	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes
Number of obs.	195373	194974	203277	212024	229192
Number of Individuals	10284	10263	10700	11160	12065

Notes: The table presents monthly event study estimates from Regression 4 for each of the income quantiles. A Poisson pseudo maximum likelihood regression is used. Estimates present log-point differences in monthly electronic consumption between winners and non-winners. Results are plotted in the bottom graphs of Figures 15 to 20 for each occupation category, respectively. The regressions use inverse probability weights generated from the propensity scores as shown in Section 5. Robust standard errors are clustered at the individual level.

Table 9 Event Studies by Occupation Category (Linear Fixed Effects Estimation)

	WG	PE	AG	NO	SB w/o high cons	SB high cons
	(1)	(2)	(3)	(4)	(5)	(6)
	E-cons	E-cons	E-cons	E-cons	E-cons	E-cons
Winners Interaction ×						
January 2017	-56.25***	-27.24*	-127.5	124.7**	76.96	-7569.1*
	(13.71)	(15.79)	(78.38)	(62.09)	(148.1)	(4244.0)
February 2017	-41.46***	-53.34**	-103.5	-16.89	32.20	-1733.3
	(13.66)	(23.97)	(71.93)	(29.49)	(148.7)	(4968.4)
March 2017	-32.50**	-35.52**	-99.61	35.85	70.65	-2282.1
	(14.59)	(15.49)	(94.67)	(34.31)	(155.9)	(4315.4)
April 2017	-26.09**	-28.07**	-99.23	-27.26	49.25	11472.6**
	(12.82)	(13.63)	(78.38)	(27.35)	(152.1)	(5095.2)
May 2017	-45.48***	-26.92*	-109.5	12.57	99.27	4911.6
	(13.27)	(14.70)	(72.02)	(27.84)	(150.6)	(5204.6)
June 2017	-44.87***	-32.82**	-112.6	-2.478	49.06	-1427.8
	(14.19)	(15.13)	(73.80)	(27.91)	(145.9)	(5099.7)
July 2017	-41.79***	-31.83**	-137.5	20.14	124.7	5916.3
	(14.13)	(13.61)	(109.7)	(27.84)	(173.2)	(5366.0)
August 2017	-42.95***	-34.22***	-77.85	50.69	119.7	23914.2***
	(14.09)	(11.98)	(72.24)	(36.91)	(189.5)	(5260.8)
September 2017	-44.59***	-31.53***	-79.24	34.93	148.4	-1338.3
	(13.12)	(11.90)	(70.67)	(30.00)	(199.8)	(4847.1)
October 2017	-30.83**	-12.03	-75.47	-11.38	-30.36	-28453.1***
	(13.12)	(12.03)	(72.67)	(25.94)	(162.8)	(5956.7)
November 2017	-42.84***	-41.94***	-13.85	-1.893	-25.79	-13213.5***
	(11.56)	(12.38)	(78.91)	(24.65)	(133.9)	(3880.7)
January 2018	58.74***	49.23***	80.59	110.2**	-30.67	-4019.5
	(12.33)	(13.63)	(105.5)	(47.04)	(121.0)	(3769.0)
February 2018	16.64	4.753	57.88	33.37	-80.87	-18346.7***
	(13.80)	(16.72)	(110.4)	(25.45)	(140.8)	(4929.7)
March 2018	31.25	49.69**	-93.83	34.60	-8.540	-24142.8**
	(20.11)	(22.73)	(81.99)	(26.17)	(142.5)	(9382.0)
April 2018	-8.158	8.977	-80.76	30.26	40.48	-14480.0
	(13.33)	(12.14)	(69.44)	(40.53)	(144.3)	(9870.8)
May 2018	-38.91**	17.70	-60.16	63.17*	40.38	-28282.3***
	(18.85)	(13.58)	(87.61)	(34.37)	(172.7)	(10816.3)
June 2018	-42.35***	15.01	-132.9*	35.10	130.2	-31509.0***
	(15.88)	(14.28)	(78.26)	(29.64)	(166.9)	(10559.9)
July 2018	-8.456	-2.901	-109.5	142.3***	193.2	-30240.4***
	(15.31)	(24.89)	(70.95)	(41.74)	(214.9)	(11344.9)
Constant	614.3***	480.5***	709.0***	479.4***	1483.5***	112850.0***
	(4.642)	(4.774)	(29.71)	(10.11)	(56.60)	(0.00126)
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Number of obs.	479763	252658	52649	201206	48640	2359
Number of Individuals	25255	13299	2771	10591	2560	129

Notes: The table presents monthly event study estimates from Regression 4 for each of the occupation categories. A linear regression is used. Estimates present differences in monthly electronic consumption between winners and non-winners. Results are plotted in the top and middle graphs of Figures 15 to 20 for each occupation category, respectively. The regressions use inverse probability weights generated from the propensity scores as shown in Section 5. Robust standard errors are clustered at the individual level.

Table 10 Event Studies by Occupation Category (Poisson Maximum Likelihood Estimation)

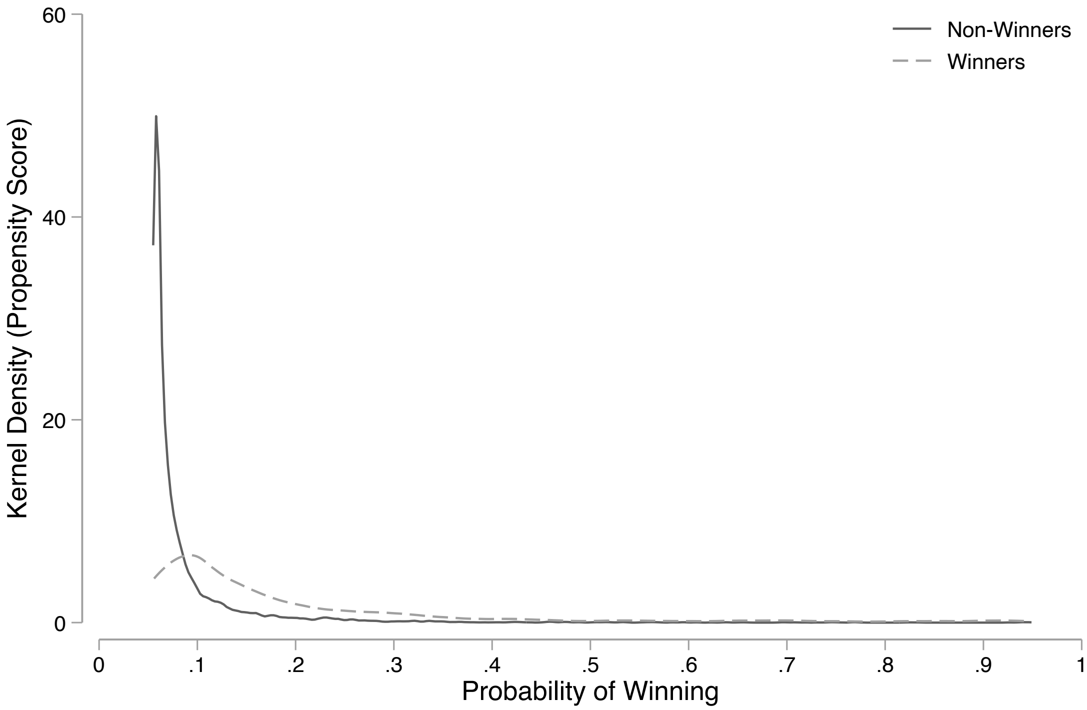
	WG	PE	AG	NO	SB w/o high cons	SB high cons
	(1)	(2)	(3)	(4)	(5)	(6)
	E-cons	E-cons	E-cons	E-cons	E-cons	E-cons
Winners Interaction ×						
January 2017	-0.0802***	-0.0189	-0.149	0.390***	0.150	-0.0694
	(0.0285)	(0.0412)	(0.155)	(0.122)	(0.105)	(0.0552)
February 2017	-0.0393	-0.0990	-0.0881	0.0615	0.137	-0.0413
	(0.0256)	(0.0650)	(0.131)	(0.0745)	(0.111)	(0.0623)
March 2017	-0.0273	-0.0558	-0.104	0.142*	0.141	-0.0122
	(0.0279)	(0.0372)	(0.175)	(0.0816)	(0.112)	(0.0563)
April 2017	-0.0171	-0.0312	-0.0873	0.0174	0.155	0.0987
	(0.0238)	(0.0318)	(0.145)	(0.0687)	(0.107)	(0.0666)
May 2017	-0.0604**	-0.0341	-0.114	0.0856	0.132	0.0642
	(0.0243)	(0.0345)	(0.124)	(0.0661)	(0.113)	(0.0650)
June 2017	-0.0570**	-0.0461	-0.118	0.0747	0.114	-0.0320
	(0.0275)	(0.0369)	(0.136)	(0.0689)	(0.109)	(0.0668)
July 2017	-0.0565**	-0.0477	-0.181	0.0874	0.152	0.0533
	(0.0262)	(0.0314)	(0.218)	(0.0648)	(0.140)	(0.0704)
August 2017	-0.0579**	-0.0468*	-0.0520	0.159*	0.152	0.248***
	(0.0259)	(0.0269)	(0.121)	(0.0867)	(0.153)	(0.0642)
September 2017	-0.0582**	-0.0390	-0.0462	0.137**	0.167	-0.0211
	(0.0242)	(0.0267)	(0.121)	(0.0679)	(0.172)	(0.0648)
October 2017	-0.0288	0.00913	-0.0553	0.0423	0.0853	-0.407***
	(0.0230)	(0.0265)	(0.120)	(0.0610)	(0.140)	(0.0884)
November 2017	-0.0546***	-0.0685**	0.0419	0.0627	0.0444	-0.137***
	(0.0207)	(0.0281)	(0.122)	(0.0574)	(0.0949)	(0.0480)
January 2018	0.105***	0.125***	0.155	0.181**	-0.0190	-0.0786*
	(0.0207)	(0.0298)	(0.144)	(0.0841)	(0.0840)	(0.0475)
February 2018	0.0720***	0.0476	0.197	0.158***	0.00579	-0.372***
	(0.0249)	(0.0411)	(0.178)	(0.0586)	(0.101)	(0.0610)
March 2018	0.0744**	0.125***	-0.109	0.0753	0.0171	-0.378***
	(0.0340)	(0.0466)	(0.130)	(0.0557)	(0.101)	(0.116)
April 2018	-0.00394	0.0392	-0.0888	0.0652	0.0323	-0.279**
	(0.0233)	(0.0261)	(0.106)	(0.0872)	(0.0966)	(0.124)
May 2018	-0.0590*	0.0597**	-0.0586	0.126*	0.0177	-0.458***
	(0.0331)	(0.0291)	(0.138)	(0.0700)	(0.118)	(0.148)
June 2018	-0.0621**	0.0574*	-0.175	0.0971	0.0655	-0.521***
	(0.0288)	(0.0311)	(0.132)	(0.0657)	(0.106)	(0.146)
July 2018	-0.00964	-0.000902	-0.132	0.249***	0.0648	-0.476***
	(0.0252)	(0.0554)	(0.110)	(0.0807)	(0.127)	(0.167)
Constant	6.848***	6.591***	7.190***	6.955***	7.991***	11.63***
	(0.00731)	(0.0100)	(0.0456)	(0.0205)	(0.0363)	(1.10e-08)
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Number of obs.	479763	252658	52649	201206	48640	2359
Number of Individuals	25255	13299	2771	10591	2560	129

Notes: The table presents monthly event study estimates from Regression 4 for each of the occupation categories. A Poisson pseudo maximum likelihood regression is used. Estimates present log-point differences in monthly electronic consumption between winners and non-winners. Results are plotted in the bottom graph of Figures 10 to 14 for each quantile, respectively. The regressions use inverse probability weights generated from the propensity scores as shown in Section 5. Robust standard errors are clustered at the individual level.

C Propensity Score Matching Tables and Figures

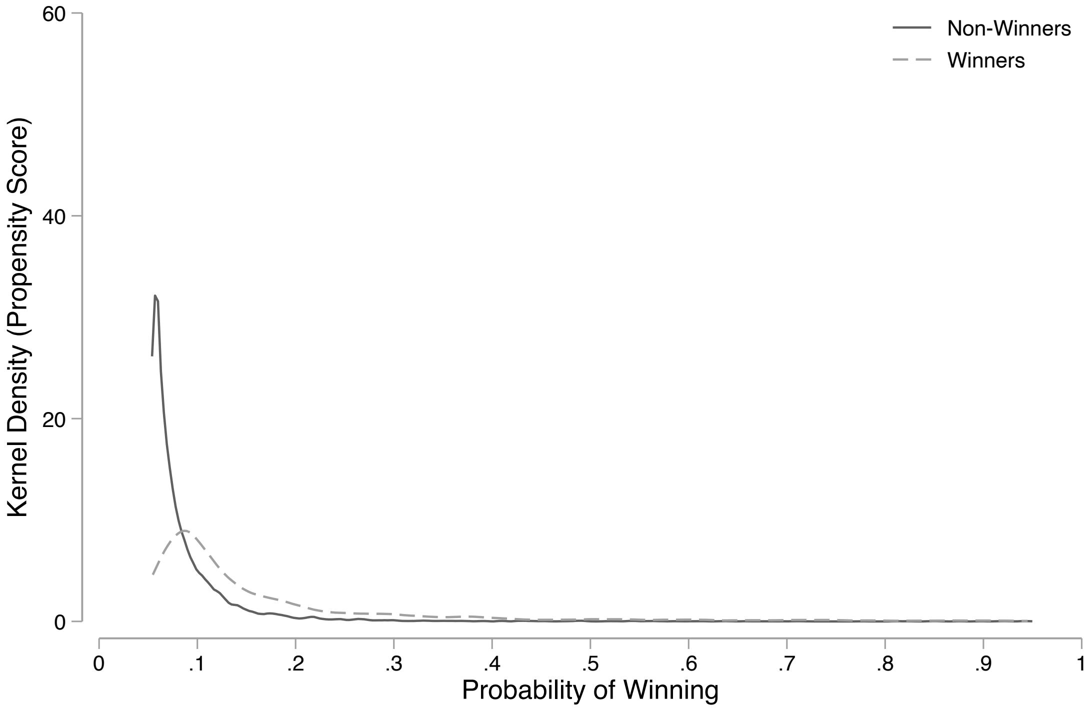
C.1 Propensity Score Matching Figures

Fig. 26 Propensity Score - First Quantile



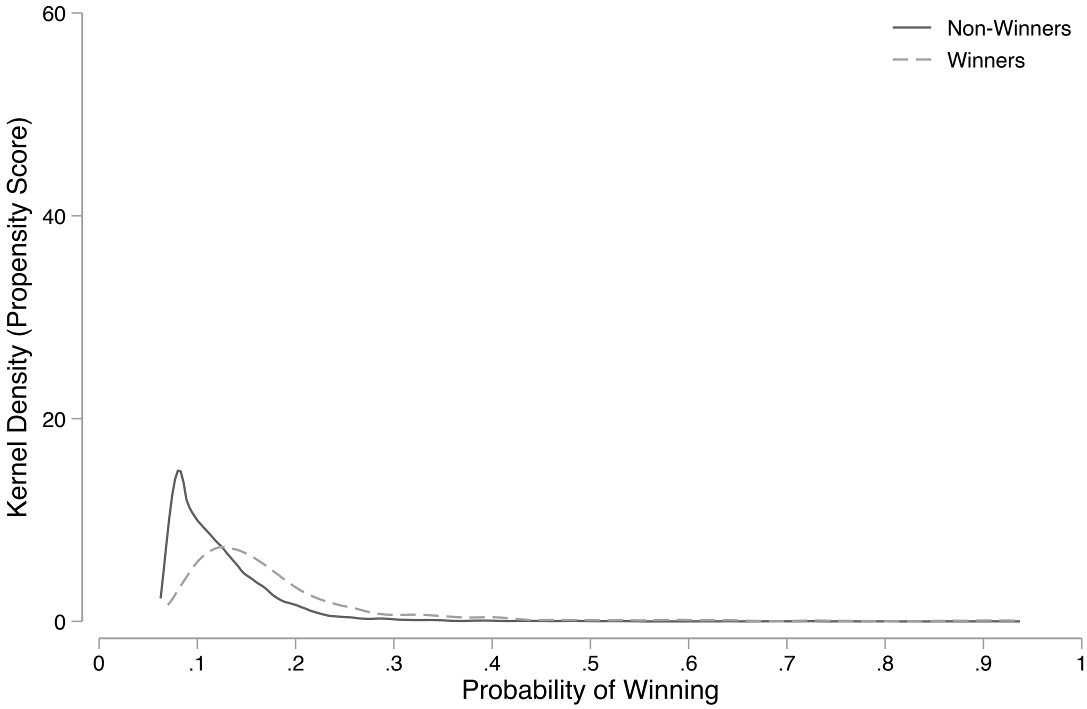
Notes: The graph plots kernel density functions of the propensity scores generated by Equation 3 for winners (dashed line) and non-winners (solid line). Taxpayers in the first quantile ranked by household income are included. The samples exclude SB and NO income categories. The propensity scores indicate the probability of individuals in winning in the superdraw based on the tickets generated in the months of January to September 2017 (which corresponded to the lottery draws). Both winners and non-winners are winsorised at 5%.

Fig. 27 Propensity Score - Second Quantile



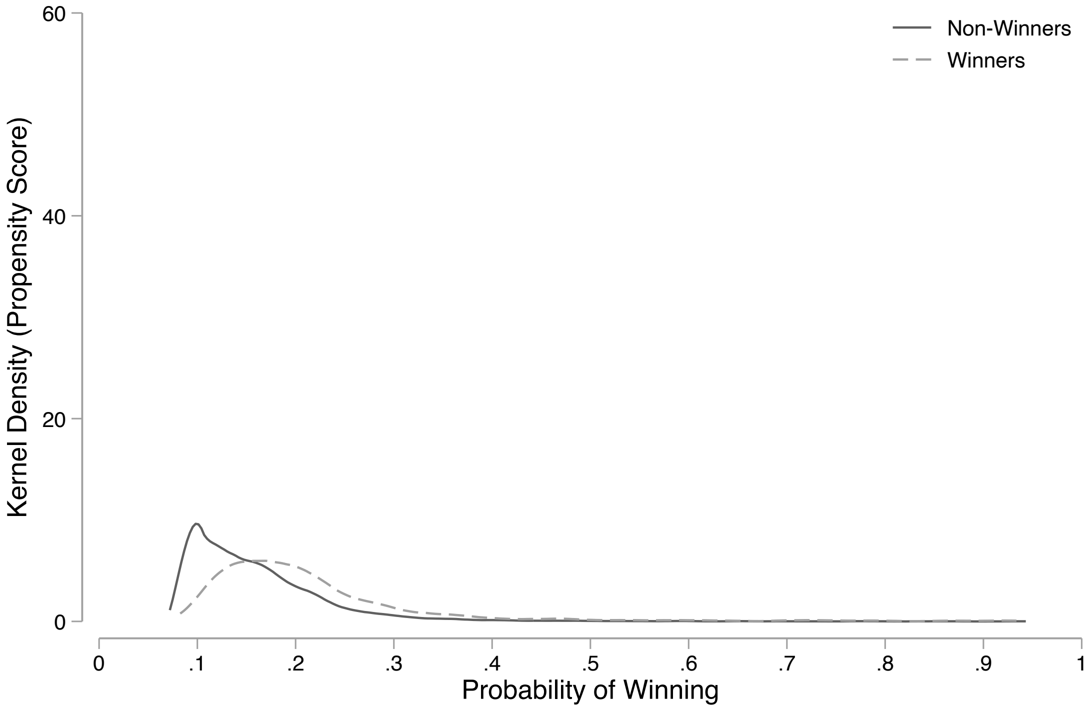
Notes: The graph plots kernel density functions of the propensity scores generated by Equation 3 for winners (dashed line) and non-winners (solid line). Taxpayers in the second quantile ranked by household income are included. The samples exclude SB and NO income categories. The propensity scores indicate the probability of individuals in winning in the superdraw based on the tickets generated in the months of January to September 2017 (which corresponded to the lottery draws). Both winners and non-winners are winsorised at 5%.

Fig. 28 Propensity Score - Third Quantile



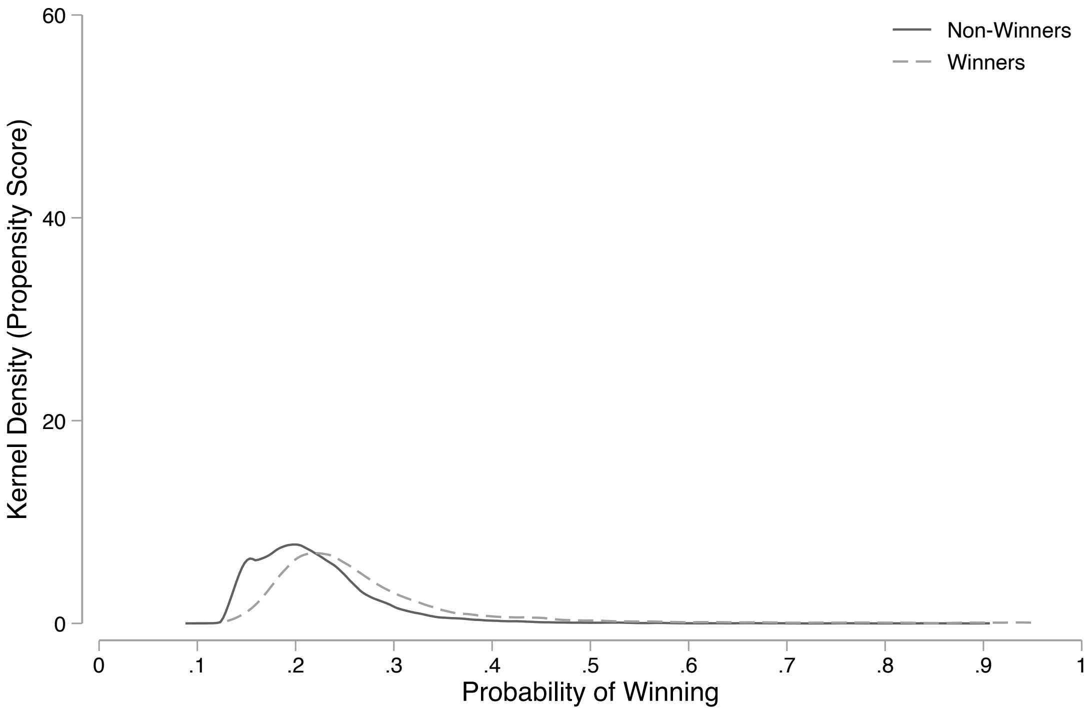
Notes: The graph plots kernel density functions of the propensity scores generated by Equation 3 for winners (dashed line) and non-winners (solid line). Taxpayers in the third quantile ranked by household income are included. The samples exclude SB and NO income categories. The propensity scores indicate the probability of individuals in winning in the superdraw based on the tickets generated in the months of January to September 2017 (which corresponded to the lottery draws). Both winners and non-winners are winsorised at 5%.

Fig. 29 Propensity Score - Fourth Quantile



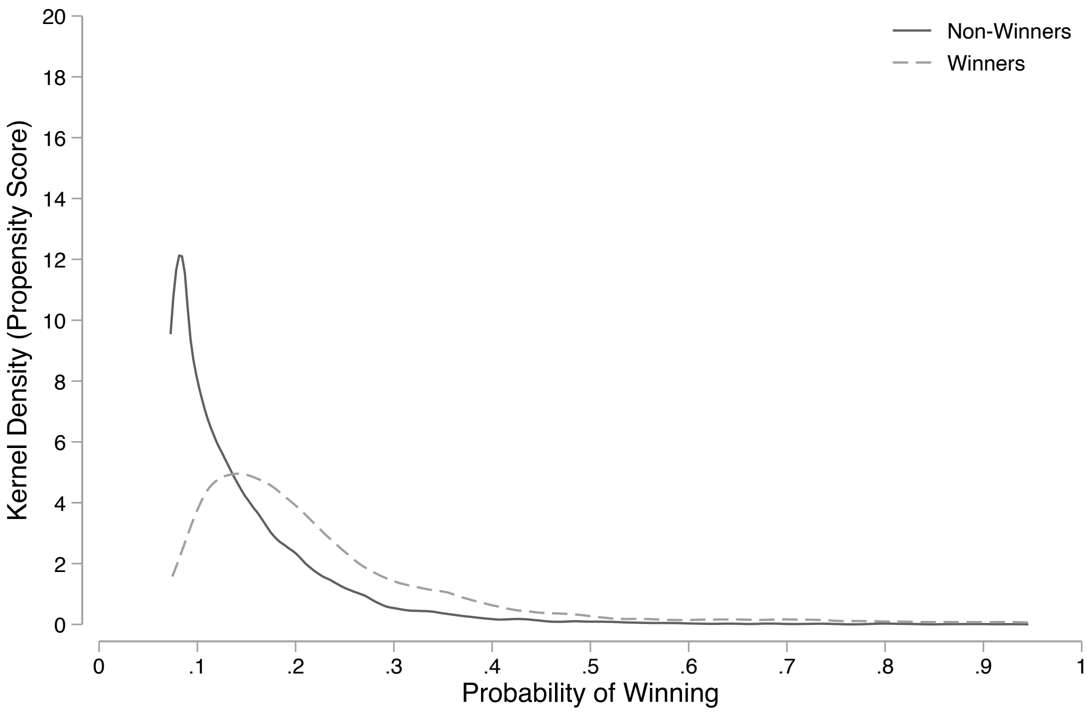
Notes: The graph plots kernel density functions of the propensity scores generated by Equation 3 for winners (dashed line) and non-winners (solid line). Taxpayers in the fourth quantile ranked by household income are included. The samples exclude SB and NO income categories. The propensity scores indicate the probability of individuals in winning in the superdraw based on the tickets generated in the months of January to September 2017 (which corresponded to the lottery draws). Both winners and non-winners are winsorised at 5%.

Fig. 30 Propensity Score - Fifth Quantile



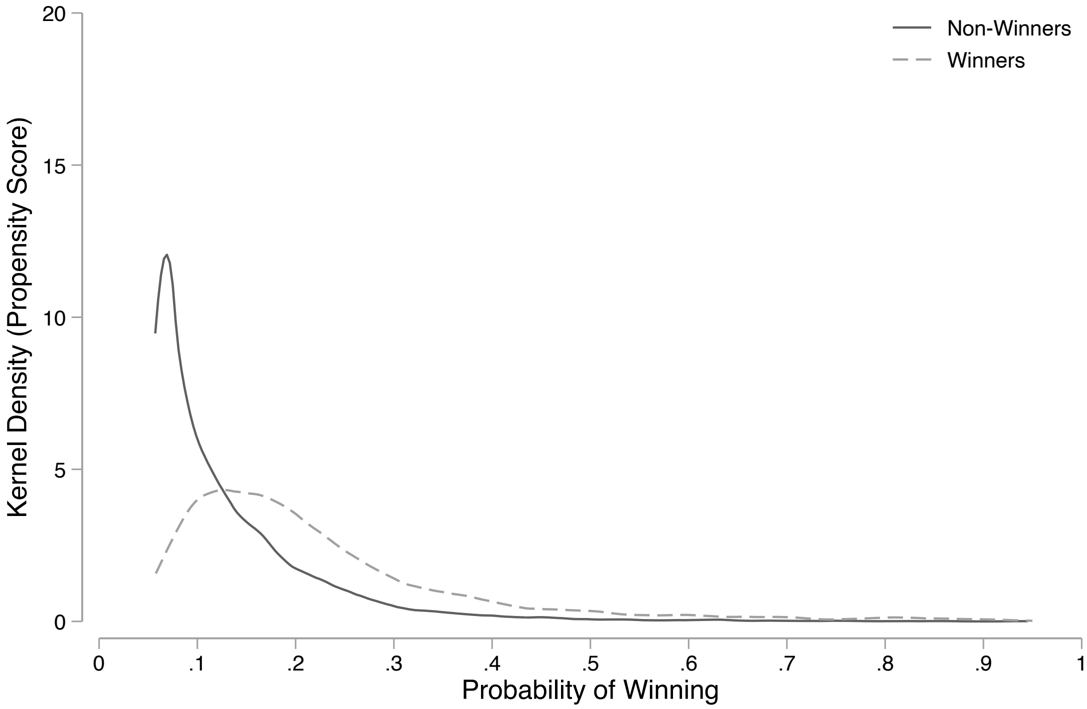
Notes: The graph plots kernel density functions of the propensity scores generated by Equation 3 for winners (dashed line) and non-winners (solid line). Taxpayers in the fifth quantile ranked by household income are included. The samples exclude SB and NO income categories. The propensity scores indicate the probability of individuals in winning in the superdraw based on the tickets generated in the months of January to September 2017 (which corresponded to the lottery draws). Both winners and non-winners are winsorised at 5%.

Fig. 31 Propensity Score - Wage-Earners



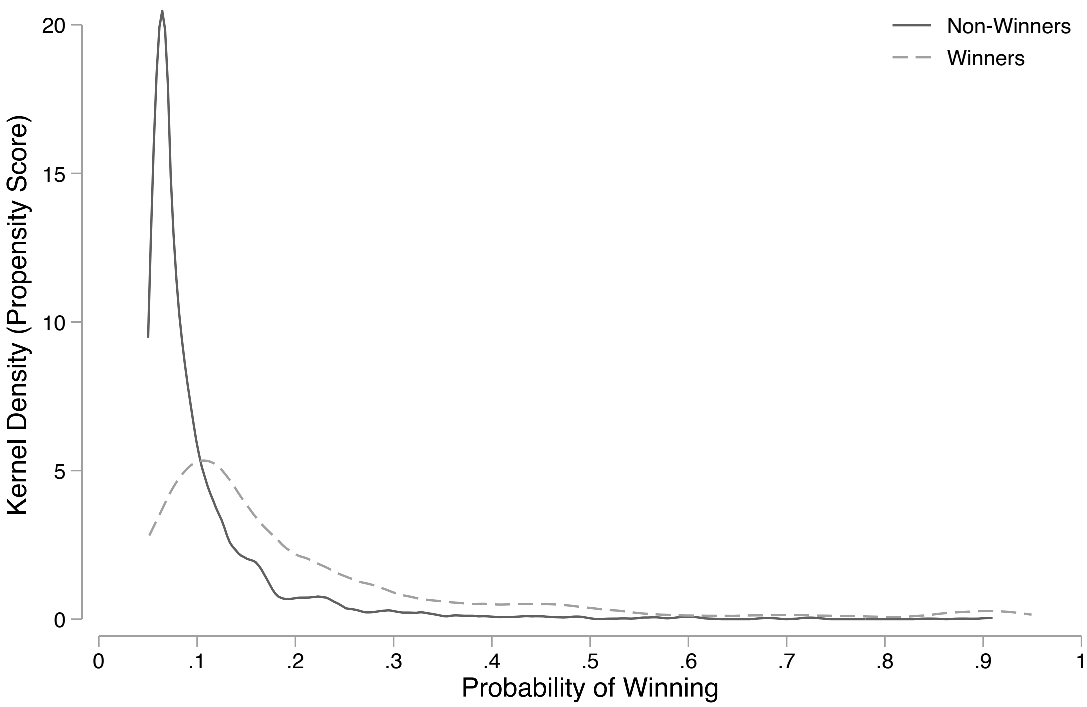
Notes: The graph plots kernel density functions of the propensity scores generated by Equation 3 for winners (dashed line) and non-winners (solid line). Wage-earners only are matched. The propensity scores indicate the probability of individuals in winning in the superdraw based on the tickets generated in the months of January to September 2017 (which corresponded to the lottery draws). Both winners and non-winners are winsorised at 5%.

Fig. 32 Propensity Score - Pensioners



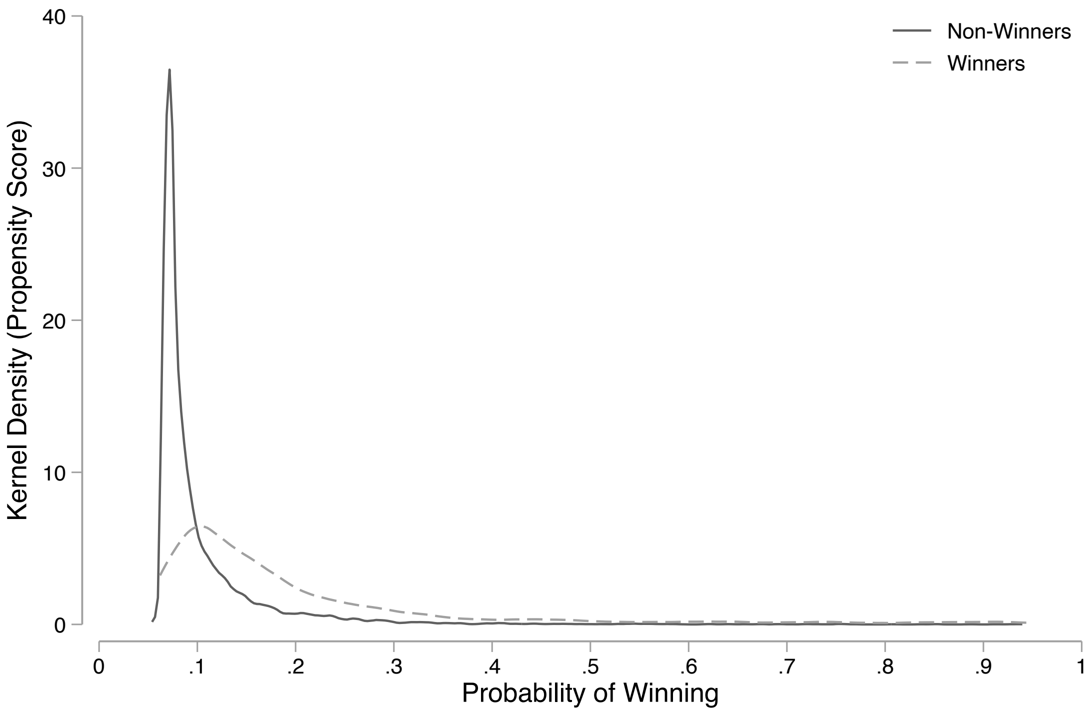
Notes: The graph plots kernel density functions of the propensity scores generated by Equation 3 for winners (dashed line) and non-winners (solid line). Pensioners only are matched. The propensity scores indicate the probability of individuals in winning in the superdraw based on the tickets generated in the months of January to September 2017 (which corresponded to the lottery draws). Both winners and non-winners are winsorised at 5%.

Fig. 33 Propensity Score - Agriculture



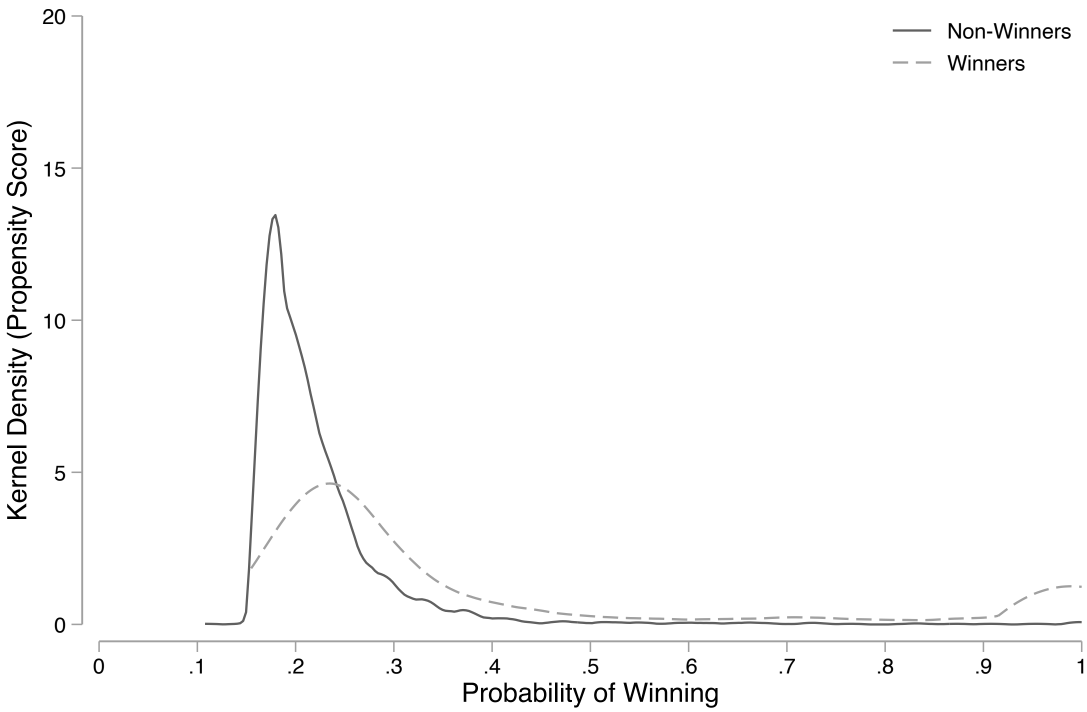
Notes: The graph plots kernel density functions of the propensity scores generated by Equation 3 for winners (dashed line) and non-winners (solid line). Agriculture income only are matched. The propensity scores indicate the probability of individuals in winning in the superdraw based on the tickets generated in the months of January to September 2017 (which corresponded to the lottery draws). Both winners and non-winners are winsorised at 5%.

Fig. 34 Propensity Score - Zero income



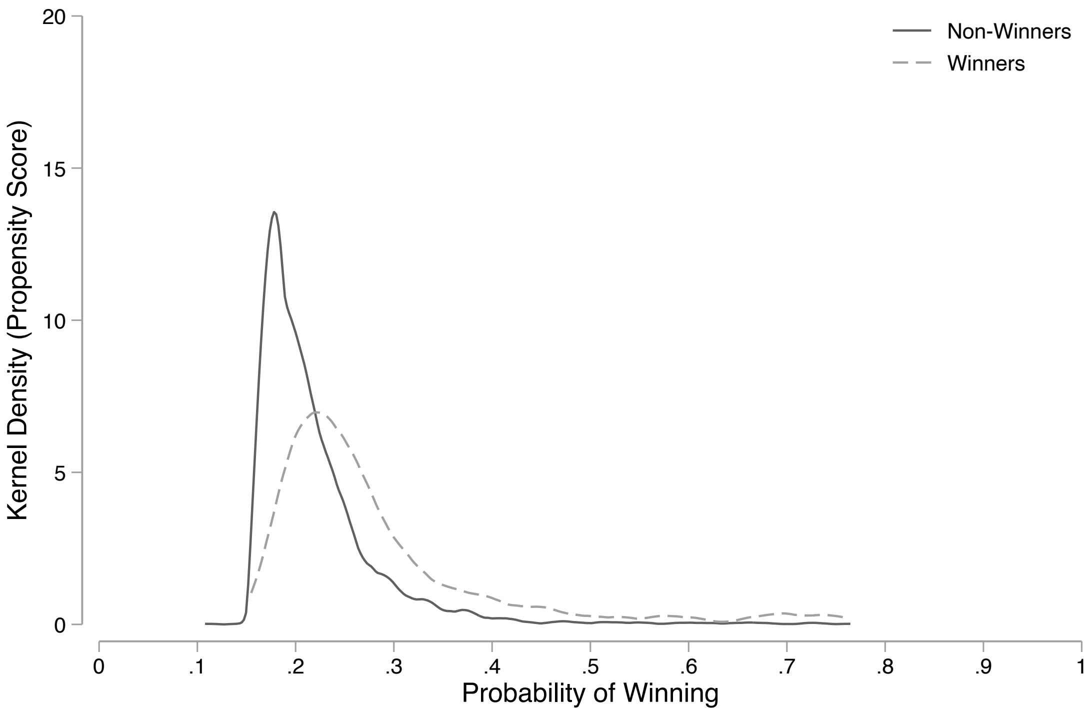
Notes: The graph plots kernel density functions of the propensity scores generated by Equation 3 for winners (dashed line) and non-winners (solid line). Zero income (NO category) only are matched. The propensity scores indicate the probability of individuals in winning in the superdraw based on the tickets generated in the months of January to September 2017 (which corresponded to the lottery draws). Both winners and non-winners are winsorised at 5%.

Fig. 35 Propensity Score - Self-Employed (All)



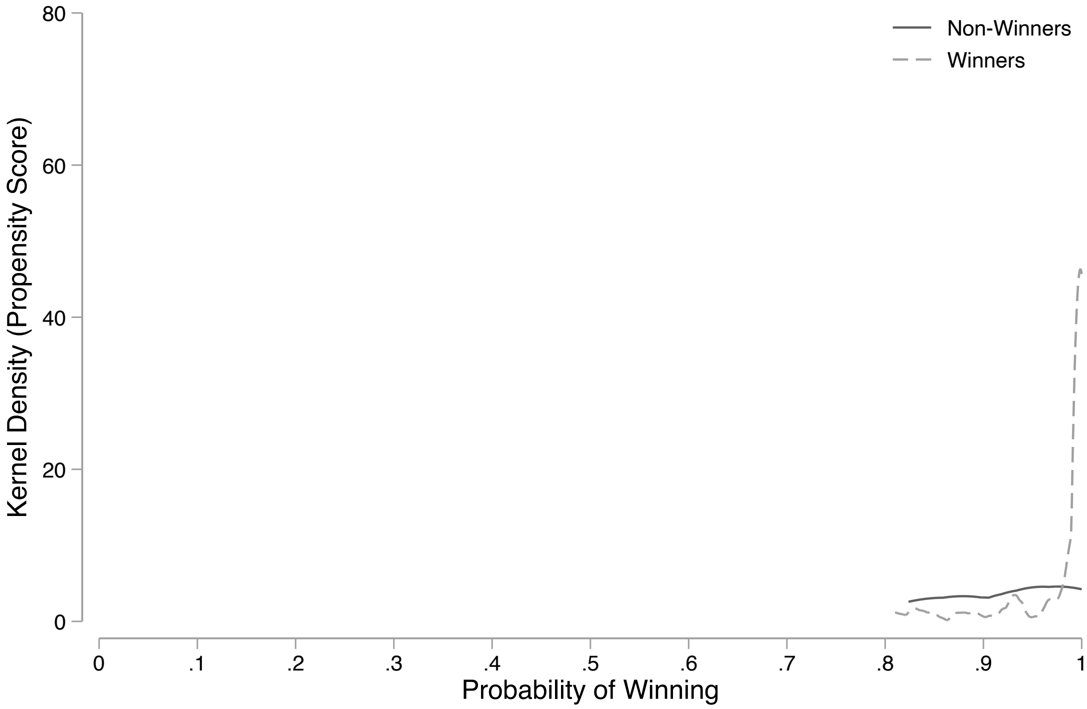
Notes: The graph plots kernel density functions of the propensity scores generated by Equation 3 for winners (dashed line) and non-winners (solid line). Self-employed only are matched (entire sample). The propensity scores indicate the probability of individuals in winning in the superdraw based on the tickets generated in the months of January to September 2017 (which corresponded to the lottery draws).

Fig. 36 Propensity Score - Self-Employed (without High E-Consumption)



Notes: The graph plots kernel density functions of the propensity scores generated by Equation 3 for winners (dashed line) and non-winners (solid line). Self-employed only are matched (from 0.8 propensity score matching and lower). This excludes the highest 20% of self-employed individuals, both for winners and non-winners. The sample is truncated on the lower side at 5%. The propensity scores indicate the probability of individuals in winning in the superdraw based on the tickets generated in the months of January to September 2017 (which corresponded to the lottery draws).

Fig. 37 Propensity Score - Self-Employed (High E-Consumption)



Notes: The graph plots kernel density functions of the propensity scores generated by Equation 3 for winners (dashed line) and non-winners (solid line). Self-employed only are matched (from 0.8 propensity score matching and higher), corresponding to SB taxpayers with high consumption both for winners and non-winners. The propensity scores indicate the probability of individuals in winning in the superdraw based on the tickets generated in the months of January to September 2017 (which corresponded to the lottery draws).

C.2 Propensity Score Matching Tables

Table 11 Logistic Regression - Probability of Winning by Income Quantile

	1st Quantile (1) P(W=1)	2nd Quantile (2) P(W=1)	3rd Quantile (3) P(W=1)	4th Quantile (4) P(W=1)	5th Quantile (5) P(W=1)
January 2017	0.0008074***	0.0005565***	-0.0001914	0.0003780**	0.0001560
	(0.0002046)	(0.0001767)	(0.0002104)	(0.0001535)	(0.0001011)
February 2017	0.0003480*	0.0008430***	0.0001972	0.0004650***	0.0002808**
	(0.0002023)	(0.0002378)	(0.0001958)	(0.0001363)	(0.0001179)
March 2017	-0.0000238	0.0004860**	0.0003292*	0.0003185***	0.0003122***
	(0.0001914)	(0.0002380)	(0.0001903)	(0.0001212)	(0.0000928)
April 2017	0.0006890***	0.0003445*	0.0011533***	0.0008519***	0.0003169***
	(0.0002109)	(0.0002085)	(0.0002208)	(0.0001596)	(0.0001146)
May 2017	0.0005256***	0.0002735	0.0005250**	0.0001659	0.0003100***
	(0.0001710)	(0.0002430)	(0.0002159)	(0.0001268)	(0.0000933)
June 2017	0.0005256**	0.0003677	0.0000300	0.0005593***	0.0002866***
	(0.0002128)	(0.0002497)	(0.0002307)	(0.0001608)	(0.0001069)
July 2017	0.0004123**	0.0006417**	0.0001613	0.0001907	0.0002787***
	(0.0001903)	(0.0002576)	(0.0002175)	(0.0001479)	(0.0000963)
August 2017	0.0000332	0.0004797**	0.0006224***	-0.0001677*	0.0000547
	(0.0001820)	(0.0002128)	(0.0001904)	(0.0000983)	(0.0001121)
September 2017	0.0002983**	0.0004078***	0.0005707***	0.0004412***	-0.0001664
	(0.0001504)	(0.0001562)	(0.0002031)	(0.0001441)	(0.0001155)
Constant	-2.8221821***	-2.8642454***	-2.5310466***	-2.3491349***	-1.8079439***
	(0.0459354)	(0.0482135)	(0.0465215)	(0.0453989)	(0.0391753)
Number of Individuals	10359	10315	10747	11213	12191

Notes: The table presents estimates from the logistic regression in Equation 3 for each household income quantile. Taxpayers from SB and NO categories are excluded, as well as, taxpayers from other income categories who exhibited zero consumption in 2017 (no lottery participation). The results are used to generate the propensity scores of winning the lottery. The months used correspond to the months that generated the tickets for the superdraw, from January to September 2017. Winning was a rare event, hence to ensure convergence of the maximum-likelihood function, a Firth logistic regression is used for these estimates. The positive values indicate the percentage increase in the probability of winning of one extra ticket in each of the months. The regression produces propensity scores, which are plotted for each quantile in Figures 26 to 30.

Table 12 Logistic Regression - Probability of Winning by Occupation Category

	WG (1) $P(W=1)$	PE (2) $P(W=1)$	AG (3) $P(W=1)$	NO (4) $P(W=1)$	SB (5) $P(W=1)$
January 2017	0.0006191***	0.0002919**	0.0002418	0.0000446*	-0.0000634
	(0.0001258)	(0.0001442)	(0.0002684)	(0.0000264)	(0.0001179)
February 2017	0.0003792***	0.0008837***	0.0008282***	0.0004129***	0.0001008
	(0.0001218)	(0.0001718)	(0.0002963)	(0.0001564)	(0.0001070)
March 2017	0.0006011***	0.0003440**	-0.0002930	-0.0002520	0.0002909**
	(0.0001106)	(0.0001408)	(0.0002174)	(0.0001689)	(0.0001179)
April 2017	0.0002238*	0.0005810***	0.0003674*	0.0010939***	0.0005584***
	(0.0001189)	(0.0001744)	(0.0002099)	(0.0001785)	(0.0001405)
May 2017	0.0004430***	0.0003082**	0.0004271*	0.0005056***	0.0001264
	(0.0001144)	(0.0001402)	(0.0002522)	(0.0001520)	(0.0000808)
June 2017	0.0002798**	0.0003965**	0.0003959	0.0007636***	0.0000199
	(0.0001222)	(0.0001768)	(0.0002541)	(0.0001842)	(0.0000893)
July 2017	0.0005468***	0.0008127***	0.0002938	0.0003013*	-0.0000076
	(0.0001208)	(0.0002052)	(0.0002159)	(0.0001611)	(0.0000913)
August 2017	0.0004917***	0.0007107***	0.0006596**	-0.0003071***	0.0000830
	(0.0001167)	(0.0001992)	(0.0002648)	(0.0000949)	(0.0000662)
September 2017	0.0002109*	0.0005264***	0.0006502**	0.0004137***	-0.0000895
	(0.0001094)	(0.0001796)	(0.0002610)	(0.0001309)	(0.0000883)
Constant	-2.5455238***	-2.8028727***	-2.8012180***	-2.6084126***	-1.5915826***
	(0.0307634)	(0.0462220)	(0.0916161)	(0.0424092)	(0.0630651)
Number of Individuals	25334	13330	2793	10674	2694

Notes: The table presents estimates from the logistic regression in Equation 3 for each income category. Taxpayers with zero consumption in 2017 were excluded (no lottery participation). The results are used to generate the propensity scores of winning the lottery. The months used correspond to the months that generated the tickets for the superdraw, from January to September 2017. Winning was a rare event, hence to ensure convergence of the maximum-likelihood function, a Firth logistic regression is used for these estimates. The positive values indicate the percentage increase in the probability of winning of one extra ticket in each of the months. The regression produces propensity scores, which are plotted for each occupation category in Figures 34 to 37.

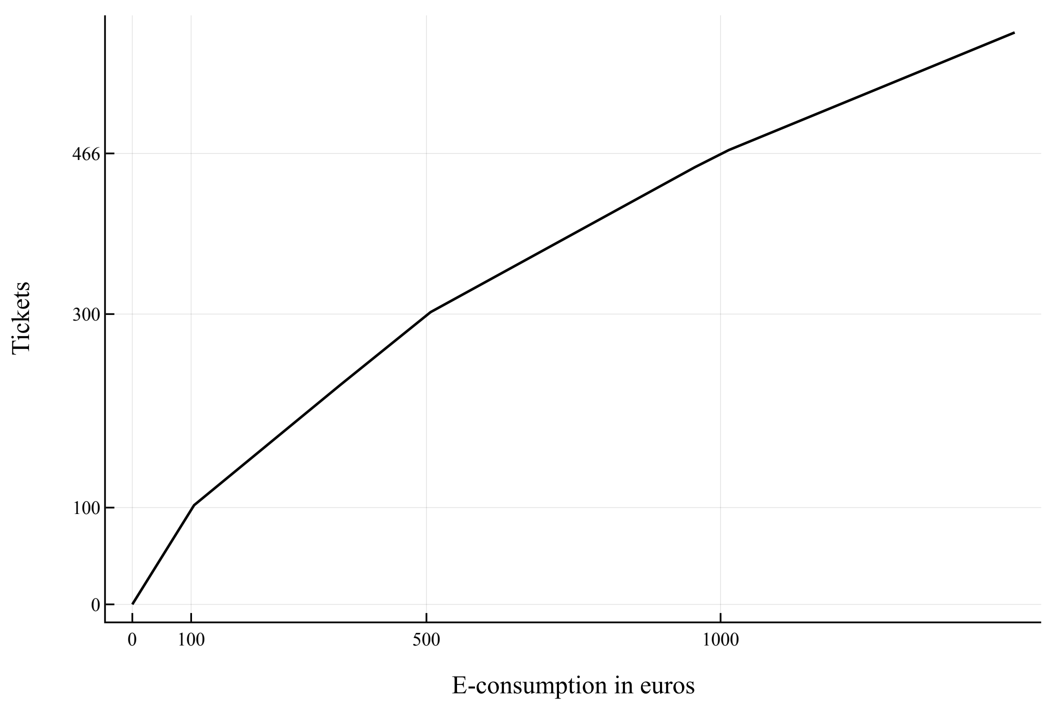
D Tax Lottery Information

Table 13 Ticket-Awarding Mechanism

Monthly E-Consumption	Ticket Conversion	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

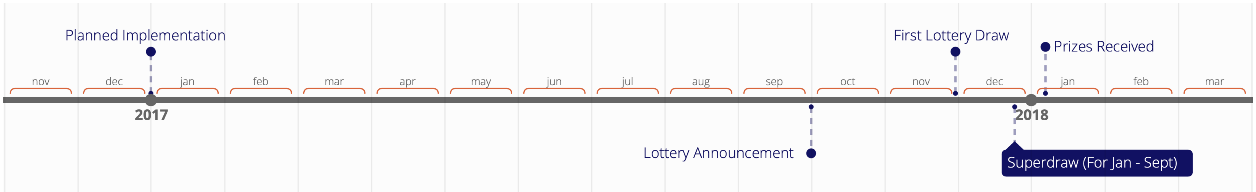
Notes: The table shows the ticket-awarding mechanism (TAM) used to convert the monthly electronic consumption of individuals into lottery tickets. At €1-100, tickets correspond at 1 for every €1. At €101-500, tickets correspond at 1 for every €2. At €501-1,000, tickets correspond at 1 for every €3. For over 1,000, tickets correspond at 1 for every €4. There was no upper limit in tickets.

Fig. 38 Euro-to-Ticket Scale



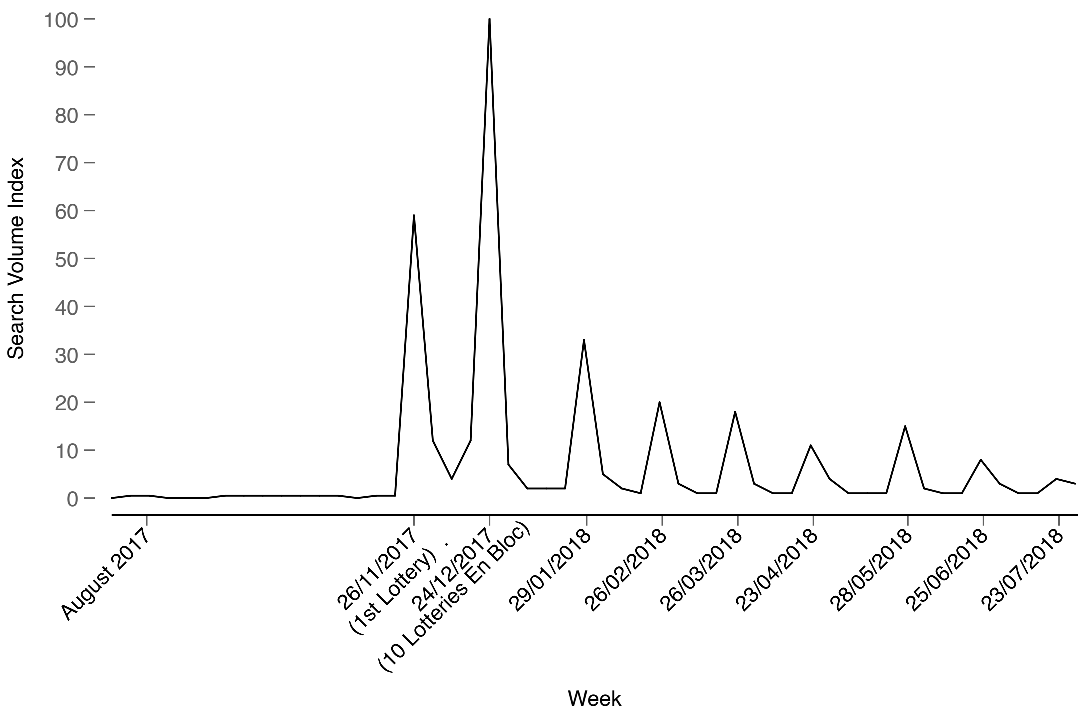
Notes: The graph illustrates the scale used to convert the aggregate level of monthly electronic consumption to eligible tickets in the lottery. Banks send the aggregate level of electronic consumption completed by each individual and this is converted to ticket using the following scale. At €1-100, 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.

Fig. 39 Superdraw Timeline



Notes: The figure shows an indicative timeline of the superdraw that took place on Christmas Eve 2017. The planned implementation was in January 2017. The lottery announcement took place in October 2017 with the first draw taking place at the end of November 2017 for payments completed in October. The superdraw took place on the 24th of December 2017, for payments corresponding to months of January to September 2017. Prizes were handed out directly to the individuals' bank accounts in early January 2017.

Fig. 40 "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.