CESifo WORKING PAPERS

7117
2018

June 2018

Combatting Tax Evasion: Evidence from Comparing Commercial and Business Tax Registry

Collen Lediga, Nadine Riedel, Kristina Strohmaier

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editors: Clemens Fuest, Oliver Falck, Jasmin Gröschl
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 7117*
Category 1: Public Finance

Combatting Tax Evasion: Evidence from Comparing Commercial and Business Tax Registry

Abstract

In 2008 and 2014, the South African Revenue Service (SARS) did snapshot synchronizations of its business tax registry with the country's commercial register in an attempt to identify firms that are non-compliant with their obligation to register with SARS for business tax purposes. We analyse these interventions drawing on SARS's business tax registry and the population of business tax returns between 2009 and 2014. Several findings emerge. First, in both years, the comparisons resulted in the identification of around 300,000 non-compliant taxpayers, providing prima facie evidence of significant extensive-margin tax evasion. The interventions significantly raised South African business tax revenues in the following years despite the fact that the identified 'extensive-margin evaders' exhibit a lower propensity to submit tax returns and, conditional on return submission, report less income than comparable entities that voluntarily registered with SARS. The analysis further suggests that the observed gap in reported taxable income relates to underlying differences in firm size and corporate profitability rather than intensive-margin tax evasion. In line with 'missing middle theories', extensive-margin evaders that submit tax returns are, moreover, found to exhibit increased sales and asset growth after their forced registration with SARS.

JEL-Codes: H200, H700.

Keywords: tax evasion, less developed countries, tax administration.

Collen Lediga
University of Bochum / Germany
Sedumetsane.Lediga@rub.de

Nadine Riedel
University of Bochum / Germany
nadine.riedel@rub.de

Kristina Strohmaier
University of Tübingen / Germany
k.strohmaier@uni-tuebingen.de

1 Introduction

Tax evasion is considered to be a major obstacle to the economic prosperity of emerging and developing economies (see, e.g., Besley and Persson, 2013). While reliable quantifications of evasion activities are hard to come by, there is a strong notion that the small number of formal taxpayers in many less developed economies is a major contributor to countries' tax gaps (see, e.g., Besley and Persson, 2014, Waseem, 2018a).

Recent years have thus seen continuous efforts by governments and tax administrations to raise taxpayer registration rates. As conventional strategies such as field audits and taxpayer information campaigns tend to be costly and limited in their outreach, several tax authorities recently turned to identifying extensive-margin evaders based on information available within the public administration sphere, namely to expanding the "interconnectivity [of tax registries] with other government data" (GIZ, 2013, see also Russell, 2010; Hoekstra et al., 2013). To date, rigorous empirical evaluations of such efforts are, to the best of our knowledge, still missing though.

The aim of this study is to help close this gap. Our testing ground is South Africa, where the South African Revenue Service (SARS) conducted two snapshot synchronisations between its business tax registry and the country's commercial register at the Companies and Intellectual Property Commission (CIPC) in 2008 and 2014. The empirical analysis in the paper draws on SARS's business tax registry and the population of business tax returns between 2009 and 2014 to assess the fiscal and behavioural consequences of these two interventions.

Several findings emerge. Firstly, semi-formality - i.e. registration with the commercial registry but not with the tax authorities - is found to be a quantitatively important phenomenon. While all firms in CIPC's commercial register are liable for business taxation and have to register with SARS to be compliant with the South African tax law, both registry synchronisations resulted in the identification of around $300,000$ taxpayers that were registered with CIPC only. Non-compliance rates are particularly high for young firms and for entities in dense economic areas.

Secondly, the behaviour of the identified 'extensive-margin evaders' after their forced registration with SARS significantly differs from firms, which voluntarily registered with SARS for business tax purposes.[1] Comparing unconditional means between groups and estimating models that account for observed firm characteristics in a flexible non-parametric way by using (coarsened) exact matching, we show that extensive-margin evaders have a significantly reduced propensity to comply with their obligation

[1] In the following, firms identified as non-compliant in the two commercial registry comparisons will be referred to as 'extensive-margin evaders' or 'semi-formal'/'partially-formal' firms.

to submit business tax returns in years in which they are deemed to be active. Quantitatively, submission rates are around 50% smaller than those of comparable control entities. Conditional on submitting a tax return, extensive-margin evaders, on top, report significantly less taxable income and tax payable than control entities, again falling short by around 50% in quantitative terms.

This income and tax gap may 1) root in size differences between the two groups of businesses, 2) reflect differences in firm productivity and profitability or 3) differences in intensive-margin tax evasion behaviour.[2] Our analysis suggests that firm size explains a significant part of the observed income and tax gap, reflecting that extensive-margin evaders are, on average, significantly smaller than control entities. This matches the perception that firm size and shadow economy operations are negatively correlated, as smaller entities can more easily stay below public authorities' radar.

To test whether extensive-margin evaders engage in more evasion at the intensive margin than firms that voluntarily registered with SARS, we turn to information from taxpayer audits and compare (the propensity for) tax adjustments in the auditing process between extensive-margin evaders and control entities in empirical models that account for systematic audit selection based on criteria such as firm size and reported profitability relative to an industry average. The results indicate no significant difference in the propensity of treatment and control firms to face tax adjustments, or in the size of these adjustments, suggesting that the observed gap in reported taxable income between extensive-margin evaders and control firms relates to size and profitability differences rather than heterogeneity in intensive-margin evasion behaviour.

This evidence is corroborated when studying the responsiveness of firms' taxable income reporting to increases in the marginal tax rate. While income responses may reflect adjustments in evasion/avoidance behaviour or changes in corporate real activity, the prior literature on taxable income elasticities in less developed economies largely assigns identified effects to intensive-margin evasion (Bachas and Soto, 2015; Waseem, 2018a). In this study, we draw on prior work in Lediga et al. (2018) and identify the corporate taxable income elasticity in South Africa based on taxpayer bunching at kinks in the corporate tax schedule for small business corporations (see Saez, 2010; Chetty et al., 2011). The findings suggest significant taxpayer bunching at the kink points which translates into sizable estimates for the elasticity of taxable income. Taxpayer responses are, however, quantitatively comparable between extensive-margin evaders and firms that voluntarily registered with SARS, providing indirect evidence that in-

[2]Specifically, taxable income of firm i at time t can be written as $INCOME_{it} = \frac{INCOME_{it}}{SIZE_{it}} \cdot SIZE_{it}$, with $SIZE_{it}$ depicting the businesses' real activity and $\frac{INCOME_{it}}{SIZE_{it}}$ capturing income reported per activity unit, with the latter being determined by the underlying firm productivity and profitability respectively as well as intensive-margin tax compliance.

tensive margin evasion does not contribute to the observed income gap across groups.

Taken together, our results suggest that the assessed synchronizations of business tax and commercial registry are effective elements of government strategies to combat tax evasion. In South Africa, the comparisons drew a significant number of firms into the tax net. While many of the identified extensive-margin evaders failed to submit tax returns, those that did show no elevated intensive-margin evasion. Moreover, extensive-margin evaders that do start submitting tax returns experience a catch-up growth in sales and assets in the years after their SARS registration. This is consistent with the theoretical predictions of "missing middle theories", which suggest that firms in the shadow economy stay inefficiently small to remain below tax authorities' radar. When drawn into the tax net, sales and production activities of these firms increase.

Overall, the identified 'semi-formal' entities contributed 2.2 billion rand of additional business tax revenue to South Africa's budget between 2009 to 2014 (about 190 million US dollars). This corresponds to about 0.25% of the overall business tax revenue during that time period, or 1.36% of the business tax revenue in 2014. These numbers are, moreover, lower bounds to the true revenue effects, as business tax payments in out-of-sample years remain unconsidered, as do potential effects on other tax bases (e.g. the value added tax, VAT).[3] Given that the administrative costs of synchronizing tax and commercial registry are small relative to these revenue gains, the sketched interventions turn out to be highly effective instruments from a cost-benefit perspective.

Our study contributes to several strands of the literature. First, we provide direct evidence for significant extensive-margin tax evasion in an emerging economy. Quantifying tax evasion activities has proved difficult in the past, especially at the extensive margin and in less developed economies. Reliable estimates are largely restricted to tax gap studies and evidence from randomized audit experiments (see, e.g., Kleven et al., 2011), which are available for a small number of countries only and are restricted to evasion at the intensive margin.[4] Shadow economy estimates based on macro-indicators such as electricity consumption and cash demand account for extensive-margin evasion (see Schneider, 2012), but were criticised in the literature for building on various non-trivial assumptions (see, e.g., Fuest and Riedel, 2009).

Our study furthermore extends the literature by rectifying the commonly assumed dichotomy that firms either operate in the shadow economy or are fully compliant. In turn, we find that a significant fraction of entities are semi-formal, being registered with some government registries (in our case, the South African commercial register)

[3]The exchange rate between the US dollar and South African rand was 11.55 at the end of our sample period on 31 December, 2014.

[4]A number of related studies moreover use indirect methods to discover tax evasion activities at the intensive margin (see, e.g., Best et al., 2015 or Waseem, 2018b).

but not with others (the South African tax authorities).

Moreover, we are the first to provide evidence on the fiscal and behavioural consequences of comparing tax registries to other government data. The study thus contributes to recent efforts to move tax analysis beyond the emphasis on optimal tax rates and bases to consider aspects of taxation such as administration and tax compliance (see Slemrod and Gillitzer, 2013). Various determinants of compliance behaviour have been identified in the literature, including traditional deterrence instruments, moral appeals, and third-party reporting (see e.g. Mascagni, 2017 for a survey). Especially third-party information has been shown to serve as a powerful enforcement device to raise tax compliance rates (see e.g. Kleven et al., 2011; Pomeranz, 2015).[5] Our study adds to this literature by stressing that taxpayer information available within the government sphere may equally help to enforce tax obligations. Two further particularities differentiate our paper from prior work: First, the existing empirical literature on individual and corporate tax compliance is largely set in developed economies. Studies for developing countries are rare, with Africa being especially under-represented (see Mascagni, 2017). Second, existing work largely focuses on the determinants of intensive-margin tax compliance. Not much is known, in turn, about the effectiveness of instruments that address non-registration of taxpayers with the tax authorities (which is a pressing problem in many less developed countries, see above). Two exceptions are Chetty et al. (2011) and DeGiorgi et al. (2015) who study the role of social recognition and field audits in driving taxpayer registration rates in Bangladesh.[6]

In addition, our study relates to recent work that quantifies the elasticity of taxable income in developing countries based on taxpayer bunching at kink points or notches in the tax schedule (see, e.g., Kleven and Waseem, 2013; Bachas and Soto, 2015; Waseem, 2018a; Lediga et al. (2018)). Authors commonly derive sizeable elasticity estimates that exceed the ones found in developed economies. Our study adds to this literature by showing that these intensive-margin tax elasticities do not differ substantially between firms that voluntarily registered with the tax authorities and extensive-margin evaders that are drawn into the tax net (and start submitting tax returns).

Last but not least, we add to the literature on the shape of the firm size distribution in lower income countries, which stresses that many less developed economies are characterised by a 'missing middle', with much of aggregate corporate activity taking place in either small or large entities, while medium-sized entities are under-represented. This pattern is commonly interpreted to reflect that firms remain small to be able to

[5] Carrillo et al. (2017) show that the effectiveness of third-party reporting in reducing evasion behaviour is limited in environments where constraints on information and enforcement are high.

[6] In related papers, Gangl et al. (2015) and Brockmeyer et al. (2016) assess the determinants of registered taxpayers' decision to file tax returns.

stay in the shadow economy and avoid the tax and regulatory burden of the official sector (see, e.g., Dharmapala et al., 2011). In line with this notion, we find that firms that are pushed into the tax net (and start submitting tax returns) show higher sales and asset growth in the years after their forced registration with SARS.

The rest of the paper is structured as follows: Section 2 presents theoretical considerations. Section 3 discusses the institutional setting and data for the empirical analysis. Sections 4 and 5 present information on the estimation strategy and the estimation results. Section 6 concludes.

2 Theoretical Considerations

The aim of this analysis is to study the fiscal and behavioural consequences of two snapshot synchronizations of SARS's business tax registry with the commercial registry at CIPC. Our empirical analysis is guided by two questions: First, we will assess whether partial formality is a quantitatively relevant phenomenon, and commercial registry comparisons help draw a relevant number of firms into the tax net. Second, we investigate whether the identified extensive-margin evaders systematically differ from their 'fully formal' counterparts in terms of observed characteristics and behaviour after being identified as non-compliant and being registered with the tax authorities.

From a theoretical perspective, firms' decision to operate under 'full formality' (i.e. registering with the commercial and tax registry), 'partial formality' (i.e. registering with the commercial registry only) or 'full informality' (i.e. registering with neither the commercial nor tax registry) depends on the benefits and costs related to the commercial and tax register entry. On the business side, registering with the commercial registry offers several benefits: limited liability of owners, facilitated access to external capital and the enabling and facilitation of transactions with other formal businesses, among others.[7] The business benefits expand for firms that additionally register with the tax registry, as signing up with both registries allows to compete for government tenders, helps ensure firms' eligibility for special government business grants and might further facilitate business with other formal sector firms.[8] Registration with SARS also rids owners of psychic cost and fine payments related to non-compliance with their ob-

[7]Procurement policies, for example, regulate that only CIPC registered firms can compete for business contracts of certain size.

[8]Firms registered with CIPC and SARS also receive a score card indicating their conformity with the Black Economic Empowerment Act, which aims for inclusive economic growth and redressing of previous inequalities in the country. The score is important for the assignment of government tenders and for the eligibility of government grants. Furthermore, note that SARS registration is a strict requirement for doing business with the government and parastatals. The private sector requirements are in turn relaxed, as many firms focus on their business partners being registered with CIPC.

ligation to register for tax purposes (see, e.g., Dulleck et al., 2016).

Tax register entry, however, also comes with tax costs and compliance burdens. Informal or partially formal firms are subject to these costs only if their non-compliance is detected by the authorities; in case of detection, additional fine payments may be levied. Consequently, firms' incentive to register with the tax authorities correlates negatively with (expected) tax costs and positively with potential fine payments and firms' (perceived) probability that non-registration will be detected.

Our empirical analysis will compare 'fully formal' firms with 'partially formal' entities in terms of their characteristics and behaviour after SARS registration. The focus of the analysis is on testing for potential differences in entities' income reporting and tax payments. Such differences may emerge because of 1) discrepancies in the size of firms' real activity, 2) differences in firm profitability or 3) differences in the scope of intensive-margin tax evasion (see Footnote 2). In general, observed behavioural differences between the two groups may, moreover, stem from either systematic firm selection into partial and full formality or may reflect behavioural responses to being identified as an extensive-margin evader.

In terms of firm selection, the theoretical predictions on sorting based on size, profitability, or tax compliance characteristics are ambiguous. Increases in business profitability and size may, on the one hand, raise firms' incentives to select into formality, as large and profitable entities plausibly reap more business benefits from SARS registration (e.g. related to the ability to bid for government tenders). Firm profitability and size, on the other hand, also positively correlate with expected tax costs, which, in turn, lowers incentives to register with SARS and makes the overall prediction on the link between profitability, size and firms' propensity to register with SARS ambiguous. Analogously, extensive-margin evasion may positively or negatively correlate with the propensity for and scope of intensive-margin evasion after registration with the tax authorities. On the one hand, evasion activities of firm owners depend on their risk and social preferences, implying that extensive-margin evaders may also exhibit increased levels of intensive-margin evasion after their forced registration with SARS. On the other hand, firm owners have particularly large incentives to engage in extensive-margin evasion if they have little opportunity for intensive-margin evasion, which suggests a negative correlation between intensive- and extensive-margin evasion propensities and, once again, makes the overall prediction ambiguous.

In addition, next to sorting, the behaviour of semi-formal and fully formal firms after SARS registration may also differ because the former firms react to being identified as extensive-margin evaders: They might, for example, expect increased audit propensities and may reduce intensive-margin evasion in response. Overall, the theoretical predic-

tions on potential differences in the characteristics and behaviour of semi-formal and fully formal firms are ambiguous and hence remain an empirical question, which will be addressed in the following sections.

3 Institutional Background and Data

Testing Ground: South Africa

Our empirical testing ground is South Africa, which is an upper-middle-income economy with a gross domestic product (GDP) per capita of 5,744 US dollars in 2015. South Africa's overall tax-to-GDP ratio is high compared with other less developed countries (amounting to 29% in 2015 relative to a 19.1% average for the African continent), but it still falls short of developed-country levels (with the OECD average corresponding to 34%). Similar to other less developed economies, corporate taxes are an important revenue source in South Africa as indicated by a high corporate-tax-to-GDP ratio of 4.7% (relative to an OECD average of 2.7%).[9]

Firms in South Africa are subject to business taxes levied under the Income Tax Act 58 of 1962. The headline company tax rate is 28%, but small and medium enterprises benefit from special tax treatment. First, 'Small Business Corporations' (SBCs) are subject to a progressive corporate tax scheme with lower tax rates. To qualify as an SBC, a company must meet specific criteria, including that gross income must not exceed 20 million South African rand (14 million South African rand prior to 2013). Furthermore, firms can opt to be taxed as micro businesses, in which case they face special tax dispensation in the form of a turnover tax. To be eligible for turnover taxation, firms' gross income for the year of assessment must not exceed 1 million South African rand and total assets must not exceed 5 million South African rand. The scheme is, however, perceived as rather unattractive, and only very few firms opted to be taxed as micro businesses during our sample period.

SARS Interventions

As described above, our empirical analysis will assess two administrative interventions where SARS compared its business tax registry with South Africa's commercial register. The commercial registry is the most obvious public information source to cross-check firms' compliance with the tax law's registration requirements as, in South Africa, all businesses must register with the commercial registry at CIPC, unless owners want to operate their business in a personal capacity as sole proprietors or partner-

[9]Information on GDP per capita was obtained from the World Bank, and information on tax-to-GDP ratios from OECD statistics.

ships. Simultaneously, all firms are required by law to register with SARS for business tax purposes, irrespective of their size or income. All entities registered with CIPC are thus liable for business taxation and have to register with SARS to be compliant with the tax law (where registration is required within 21 business days after a firm becomes liable for taxation, i.e. starts doing business and earning income).[10]

The commercial registry comparisons were conducted in early 2008 and early 2014, and were organised internally, i.e. there was no prior media coverage. All firms in the commercial register that were identified as non-compliant with their obligation to sign up with SARS for business tax purposes were added to the business tax registry in April 2008 and February 2014. They received a letter informing them that they were registered with SARS for business tax purposes based on their commercial register entry and that they were required to submit tax returns for all tax years from their date of incorporation onwards and for subsequent years to come (subject to fines for late registration and late submission of tax returns).

Data for the Empirical Analysis

In the following, we will empirically assess these interventions drawing on rich taxpayer data obtained from SARS. First, we rely on SARS's current and historic business tax registry which includes information on firms' registration and deregistration dates with SARS, coupled with data on the registration date with CIPC, which allows us to identify whether firms are 'active' in a given tax year and liable for business taxation. The business tax registry, on top, includes further baseline data, most importantly on the tax office that is responsible for the taxpayer (which allows us to determine the geographic location of the firm) and on firms' industry.[11]

In addition, these data are augmented by information on the population of business tax returns for the tax years 2009 to 2014, comprising information on reported taxable income, firms' tax payments as well as sales, costs and assets. Finally, we make use of data from taxpayer audits conducted between 2011 and 2015 (covering tax returns for 2009 to 2014 and including, among others, information on adjustments to firms' reported taxable income and tax payable).

4 Empirical Analysis

The empirical analysis is split in two parts. In Section 4.1, we assess the quantitative importance of semi-formality and compare the baseline characteristics of the identified

[10]Note that sole proprietors and partnerships must not separately register their businesses with SARS, but owners are required to include business income in their personal tax return.

[11]The industry information is only available for firms that submitted a tax return at least once.

semi-formal entities to fully formal control firms. In Section 4.2, we move on to investigating the behaviour of these two sub-groups of firms after SARS registration and, among others, test for differences in the propensity to submit tax returns, in income reporting, firm growth and tax compliance. Note that we will, in the following, stick to our terminology in Section 2 and refer to firms identified as non-compliant with their SARS registration requirement as 'extensive-margin evaders' or 'semi-formal/partially formal' entities. Firms that voluntarily registered with SARS are labelled as 'fully-formal' entities (- despite the fact they may be less than fully compliant at the intensive margin).

4.1 Semi-Formality: Numbers and Firm Characteristics

As described above, SARS conducted two comparisons of its business tax registry and the South African commercial register, one in early 2008 and one in early 2014. Firms identified as non-compliant (listed in the commercial but not the tax registry) were added to the business tax registry in April 2008 and February 2014 respectively.[12] At both points in time, the efforts resulted in the identification of a large number of non-compliant firms, leading to spikes of 296,486 and 344,889 new business registrations in the respective registration months. See Figure 1.

To put these numbers into perspective, note that SARS registered 16,639 (177,732) new firms in other months (years) between January 2007 and December 2015 on average.[13] The evidence thus suggests that a non-negligible number of entities in South Africa opted for partial formality in our sample frame and registered with the commercial registry but not with the tax authorities. As all firms in the CIPC register are required by law to register for business tax purposes, this is *prima facie* evidence of tax evasion in a quantitatively important dimension.

In terms of firm age, the large majority of semi-formal companies identified in the 2008 commercial registry comparison had registered with CIPC between 2000 and 2008, while semi-formal companies identified in the 2014 commercial registry comparison had registered with CIPC after the 2008 comparison (between 2008 and 2014). Figure 2 depicts the identified non-compliers as a fraction of all CIPC business registrations in

[12]Note that we can identify the days in these months on which the forced registrations took place (characterized by spikes in the number of new taxpayer registrations). All 'spike day' registrations will, in the following, be assumed to be related to SARS's commercial registry comparisons. For the 2008 commercial register comparison, the spike days are April 1–5 and April 8–9; for the 2014 commercial registry comparison the spike days are February 2–4, February 6, and February 18.

[13]Note that there is a smaller spike in registration numbers in July 2009. This likely relates to changes in the tax treatment of venture capital shares in July 2009. Precisely, from that month onwards, investors were allowed to claim amounts incurred on acquiring venture capital shares as a deduction from their taxable income (see SARS's External Guide for Venture Capital Companies).

a given registration year. The figure suggests that especially young firms have a high propensity to remain semi-formal: Precisely, 30.3% (40.6%) of all firms that registered with CIPC in 2006 and 2007 (2012 and 2013) were identified as missing in SARS's business tax registry in the course of the 2008 (2014) commercial register comparison. These fractions decrease for businesses that were older at the time of the commercial register comparisons. For example, among firms aged four, only 12.7% and 15.6% were identified as non-compliant with the SARS-registration requirement in the 2008 and 2014 comparisons respectively.[14] Reiterating this evidence, the median time span between registry with CIPC and SARS amounts to 601 and 596 days for non-compliant firms identified in the commercial registry comparisons in 2008 and 2014, respectively, compared with a median registration gap of 9 days for other newly registered firms between 2007 and 2015. See Figure 3 for kernel densities of the registration gap.

On top, Table 1 shows that semi-formal firms have an elevated probability to locate in dense economic provinces. Specifications (1) and (2) test for a correlation between being identified as an extensive margin evader in the 2008 and 2014 commercial registry comparison respectively and firm location in Gauteng province, which has the highest population density in South Africa (675 people per square metre) and hosts the country's leading cities of Johannesburg and Pretoria.[15] The results indicate that semi-formal firms exhibit an increased propensity to locate in Gauteng (where the quantitative gap amounts to 1.6 and 2.1 percentage points or 3.0% and 4.4%, evaluated at the average propensity to locate in Gauteng). Specifications (3) to (6) reiterate this pattern by assessing firms' propensity to be located in Gauteng or KwaZulu Natal province ('High Dens1') and Gauteng, KwaZulu Natal, Mpumalanga, Limpopo or Western Cape province ('High Dens 2'), respectively, which are the provinces with the next highest population density in South Africa. This pattern may reflect that firm owners in more urbanised regions have preferences that favour tax evasion (e.g. low levels of risk aversion) relative to owners in rural areas, or that their (perceived) risk for detection of evasion activities is systematically lower. The correlation may also root in higher levels of market competition in urbanised areas, which have been linked to increased levels of tax evasion in the economic literature (see e.g. Branco and Villas-Boas, 2015, Baumann and Friehe, 2016).

[14]Note that we only observe firms that voluntarily registered with SARS or were identified as semi-formal in the wake of the commercial registry comparisons. The commercial register comparisons, however, were restricted to the current CIPC registry. Hence, they only capture firms that were active in the CIPC registry at the time of the comparisons. Firms that signed up and deregistered with CIPC before the comparisons are, in turn, not captured in our analysis if they did not register with SARS for business tax purposes.

[15]Note that, in the former (latter) case, the sample is restricted to firms that had registered with CIPC between 2001 and 2008 (2009 and 2014) as the large majority of extensive-margin evaders identified in the 2008 (2014) commercial registry comparisons had registered with CIPC within these time frames. See also our discussion in the next section.

4.2 Behaviour after Registration with SARS

Starting from these baseline results, we embark on analysing the behaviour of the identified semi-formal firms after their forced registration with SARS. In doing so, we account for different decision margins, including the choice to submit tax returns, the reporting of taxable income as well as intensive margin tax evasion.

Empirical Approach

Methodologically, we estimate models of the following form

$$DEP_{it} = \alpha_1 + \alpha_2 TRC_i + \epsilon_{it} \tag{1}$$

where DEP_{it} depicts the dimension of firm behaviour considered (e.g. the decision of firm i to submit a tax return in year t). TRC_i is a dummy variable that takes on the value 1 if firm i was identified as non-compliant with its obligation to register with SARS for business tax purposes in the commercial registry comparisons and ϵ_{it} is the error term. To assess whether behavioural differences between fully formal and semi-formal firms are driven by observed heterogeneity across the two groups, we extend the model by control variables subsumed in the vector X_{it}, yielding

$$DEP_{it} = \beta_1 + \beta_2 TRC_i + \gamma X_{it} + \mu_{it} \tag{2}$$

where X_{it} varies across specifications (see below for further details). To flexibly control for several forms of observed imbalance between semi-formal and fully formal firms - not only imbalance in means, but also imbalance in higher moments and interactions - we implement Coarsened Exact Matching (CEM, see Iacus et al., 2012), where 1) covariates are coarsened in an economic sense, 2) entities are exactly matched on the coarsened data to obtain the matching weights, and 3) the regression model in equation (2) is run on the uncoarsened data using the derived matching weights.[16] To account for different values of continuous variables within coarsened categories, all continuous variables are furthermore included as regressors in equation (2).

[16]This matching technique has a number of appealing statistical features, see Iacus et al. (2012).

Submission of Tax Returns

The first choice margin analysed is the decision to submit tax returns with SARS. Entities identified as non-compliant in the wake of the commercial registry comparisons received a letter informing them that they were registered for business tax purposes based on their commercial register entry and that they were required to submit tax returns for past tax years since their incorporation and for future years to come. Hence, all newly registered entities were required to submit one or more returns.

In the following, we will assess firms' decision to comply with this legal obligation. Tax return submission rates tend to be low in many less developed countries (see, e.g., Brockmeyer et al., 2016) and South Africa is no exception. To assess the propensity of semi-formal firms to submit tax returns relative to firms that voluntarily registered with SARS, we first construct the set of tax years in which firms were active based on their registration date with CIPC and SARS (yielding the first tax year in which firms were liable for taxation[17]) and, if applicable, their date of deregistration with SARS (deriving from that the last tax year in which firms were liable for taxation). The data is linked to our tax return information for the tax years 2009 to 2014 to determine whether firms complied with the tax law and submitted tax returns in the years in which they were active during that six-year period.

Table 2 presents results for the commercial register comparison in 2008. As the large majority of extensive-margin evaders identified in 2008 had registered with CIPC between 2001 and 2008 (see also Figure 2), the sample is restricted to the 1.55 million firms that signed up with CIPC during that time period.[18] The table reports the difference in return submission propensities separately for each tax year. Columns with uneven numbers show the unconditional difference in average return submission rates of extensive-margin evaders and control entities that voluntarily signed up with SARS; columns with even numbers show CEM results that account for firms' year of CIPC registration and geographic location as measured by taxpayer affiliation with one of the 45 South African tax offices.[19]

The findings suggest that extensive-margin evaders have a lower propensity to submit tax returns: For the tax year 2009, the quantitative gap amounts to 18.16 percentage points (or 56.1%, evaluated at the average tax return submission propensity of fully

[17]The first liability year is indicated in the data and refers to the earlier of the two dates.

[18]Note that, strictly speaking, the 1.55 million firms do not comprise all entities that registered with CIPC between 2001 and 2008 but only firms that registered with both CIPC and SARS, including firms that were identified as non-compliant in the 2008 commercial registry comparison. Firms that registered *and* deregistered with CIPC before 2008 are in turn not captured in the analysis if they failed to sign up with SARS for business tax purposes.

[19]Note that the observation numbers in the CEM models fall short of the unconditional mean comparisons. This is because observations in the registration year-area cells that do not include both semi-formal and fully formal firms are assigned a CEM-weight of zero and drop out of the regression.

formal firms (32.4%)), cf. Column (2). The relative size of this gap remains constant across tax years, while the general propensity to submit tax returns declines over time.[20] The size of the gap is, moreover, largely unaffected by conditioning on firm characteristics, implying that neither firm age nor geographic location (nor any interaction between the two) drives the results. Analogous findings are derived for the commercial register comparison in 2014, see Table 3. As extensive-margin evaders identified in the 2014 comparison had registered with CIPC between 2008 and 2014, the corresponding sample comprises the 1.42 million firms that registered with CIPC during that time period. Our tax return data for 2009 to 2014 then allow us to identify return submission rates for all tax years *prior* to the 2014 comparison in which the firm had been active according to the CIPC registry (while, for the 2008 comparison analysis, we observe data for tax years *after* the 2008 commercial register comparison in which firms had been active).[21] In quantitative terms, the return submission rates of semi-formal and fully formal firms diverge even further for the 2014 comparison (by up to 84.2%).[22] See also Figure 4, which graphically depicts the difference between return submission rates of semi-formal and fully formal firms drawn from CEM models and, furthermore, shows that the significance of our results remains unaffected when we allow for clustering of errors at the level of tax authority districts (- the corresponding 95% confidence bounds are depicted).

The results thus point to significantly lower tax return submission propensities of extensive-margin evaders relative to fully formal control entities. One remaining concern might be that the results pick up differences in the propensity of semi-formal and fully formal firms to deregister with government registries when they become inactive. Specifications (11) and (12) of Table 3 address this concern by restricting the sample to firms' initial registration year with CIPC, thus ensuring that all businesses in the sample were indeed active for at least a fraction of that tax year. The results resemble our prior findings. While firms that voluntarily registered with SARS have a probability of 38.3% to submit a tax return for the year of their initial registration, this propensity is 27.19 percentage points (or 71%) lower for extensive-margin evaders.[23]

[20]The absolute gap in return submission propensities continuously declines over time, reaching 10.12 percentage points in 2014. This pattern is driven by a general decline in submission propensities for both, extensive-margin evaders and fully formal firms, potentially reflecting that more and more firms close down when they grow older and a fraction of them fails to deregister with SARS. The relative size of the gap, in turn, remains approximately constant.

[21]Extensive-margin evaders identified in the commercial registry comparisons were required to submit tax returns for all years since their registration with CIPC, including tax years before their forced registration with SARS, in which they had been active according to CIPC.

[22]We do not report results for the tax year 2009, as only 39 extensive-margin evaders identified in the 2014 commercial registry comparison were required to submit a return in that year.

[23]Note that this part of the analysis is limited to the 2014 trade register comparison. For the 2008 intervention, we lack tax return information for years prior to 2008 when the firms used in this part

Reporting of Taxable Income and Tax Payments Due

Next, we turn to analysing the behaviour of firms that do submit tax returns. Table 4 tests for differences in the propensity of extensive-margin evaders, identified in 2008, and fully formal control firms to submit positive taxable income, conditional on tax return submission. The sample is again restricted to entities that registered with CIPC between 2001 and 2008 and columns with uneven (even) numbers show differences in unconditional means between the two groups of firms (CEM estimates, which account for firm age, host area, and firms' two-digit industry).[24] The results indicate that only a minority of the South African firms reports positive profits.[25] In the tax year 2009, 17.0% of the fully formal firms submitted a tax return with positive taxable income, with this propensity being about 4.42 percentage points or 26.0% smaller for the group of extensive-margin evaders. In line with intuition, the probability for reporting positive taxable income then increases in subsequent tax years (potentially reflecting that non-profitable entities drop out of the market, or move into the shadow economy, and remaining firms become more mature and more likely to break even). The reporting gap between semi-formal and fully formal businesses simultaneously declines, cf. Specifications (3)-(12). Table 5, furthermore, reports analogous results for the 2014 commercial registry comparison[26]; Figure 5 presents a graphical depiction of the findings and shows that the significance of the sketched results remains unaltered when we allow for clustering of errors at the tax district level (- the corresponding 95% confidence intervals are presented).

Complementary, Tables 6 and 7 compare the *level* of reported taxable income between semi-formal firms and their fully formal counterparts, conditional on taxable income being positive. Table 6 analyses the 2008 trade register comparison and again, separately for each tax year, presents specifications that report unconditional mean differences of the log of taxable income between extensive-margin evaders and control firms as well as CEM estimates which account for firm age, host area, and firms' two-digit industry. Additionally, we estimate specifications which assess whether firm size differences contribute to the observed income gap by adding the log of firms' total assets as a CEM-category and control variable.[27] Two main insights emerge: First, we find

of the analysis had registered with CIPC.

[24]Information on firms' industry can only now be accounted for as it fails to be available for firms that do not submit tax returns.

[25]Note that South African tax law allows for indefinite loss carryforwards. Our taxable income variable accounts for carried forward losses: If firms e.g. earn positive income in a given tax year but fully offset it by losses from prior years, zero income is coded in our data.

[26]We again disregard the tax year 2009 for this part of the analysis as only 22 extensive-margin evaders identified in the 2014 commercial registry comparison submitted a tax return for 2009.

[27]For the tax year 2009, Specification (1) presents the unconditional mean difference, Specification (2) the CEM estimate for the full sample, Specification (3) the CEM estimate for the sample of firms with non-missing asset information and Specification (4) includes the log of firms' assets as an

that extensive-margin evaders report less taxable income than other firms. The unconditional mean difference amounts to around 80%. In CEM specifications that, among others, account for the fact that extensive-margin evaders are systematically younger than other firms, the taxable income gap drops to around 50%. When total assets are included, this difference is further reduced by around 20 percentage points or 40%, suggesting that firm size explains a significant fraction of the taxable income difference between extensive-margin evaders and fully formal control firms.[28] Qualitatively and quantitatively similar results emerge for the 2014 commercial registry comparison, see Table 7 and Figure 7. Lastly, wee find an analogous gap between the tax payments of semi-formal and fully formal firms (see the online appendix for details).

Summarizing, the findings so far suggest that extensive-margin evaders, on average, contribute less business tax revenues than other entities to South Africa's budget: they are less likely to submit tax returns and, conditional on tax return submission, report lower income and tax payments. A fraction of this difference is, moreover, found to relate to their smaller firm size. The remaining gap indicates that extensive-margin evaders also report less income per activity unit than their fully formal counterparts, which may either reflect lower levels of firm profitability or elevated intensive-margin tax evasion. In the following, we will assess the latter transmission channel and test whether intensive-margin evasion adds to explaining the identified income and tax gap between extensive-margin evaders and control firms.

Tax Evasion

As laid out in Section 2, it is theoretically unclear whether semi-formal firms engage in more or less intensive-margin evasion than other entities after their registration with SARS. On the one hand, they may be less compliant at the intensive margin due to preference structures that favour tax evasion (e.g. low levels of risk aversion). On the other hand, they may be more compliant at the intensive margin as firms with little opportunities to lower their tax burden at the intensive margin have particularly large incentives to avoid registration with SARS. Firms in the two groups may, moreover, differ in their intensive-margin evasion behaviour because of differences in the (perceived) propensity that intensive margin evasion is detected by the authorities (see Section 2 for details).

From an empirical point of view, identifying intensive-margin tax evasion is a challenging task as firms hide evasion activities from the authorities and the public (see

additional CEM-category and control variable.

[28]For 2013 and 2014, the compensating effect of firm size turns out to be particularly large (amounting to 74% and 76%, respectively), which may, however, relate to changes in the sample size as the coverage of total assets drops in these years due to changed reporting requirements.

our discussion in Section 1). In the following, we will pursue two strategies to assess whether intensive margin evasion differs between extensive-margin evaders and fully formal control firms. In the first approach, we compare the response in firms' taxable income reporting to changes in the marginal tax rate. While income adjustments to tax changes might reflect evasion/avoidance behaviour as well as adjustments in real activities, the prior literature suggests that response behaviour is largely dominated by tax evasion in less developed economies (e.g., Bachas and Soto, 2015; Waseem, 2018a). As a second and complementary strategy, we use information on income adjustments in the course of business taxpayer audits as a proxy for evasion activities and compare the propensity for and the level of audit-related income adjustments between the two firm groups.

For the first strategy, we estimate the elasticity of corporate taxable income in South Africa by exploiting kinks in the South African business tax schedule for SBCs and using the bunching approach developed by Saez (2010) and Chetty et al. (2011). To maximize power, we pool information from all tax years 2009-2014. The SBC schedule features three kinks, which are graphically depicted in Figure A1 in the online appendix, for the most recent tax year in our data (where marginal tax rates jump from 0% to 7% at around R70,000, from 7% to 21% at R365,000 and from 21% to 28% at R550,000); tax schedules for the other years are available at SARS's webpage.[29] In the following we will focus on bunching at the first and second thresholds as the third kink was introduced in later sample years only and we observe only a small number of firms with taxable income close to this kink. Theoretically, the introduction of kinks in the budget set establishes incentives for taxpayers within a certain income range to bunch at the kink point. By comparing the income density distribution with a counterfactual scenario without the kink, the excess mass of taxpayers can be used to determine the elasticity of taxable income with respect to the net of tax rate (see Saez, 2010 or Chetty et al., 2011 for details). To estimate the excess mass of taxpayers in the bunching area, we follow Chetty et al. (2011) and determine the counterfactual density by running a local polynomial regression on binned data, while excluding data bins within the bunching window. We choose the number of polynomials using the Bayesian information criterion and follow an endogenous data-driven procedure to determine the bunching window (Dekker et al., 2016), thereby allowing the window to be asymmetric. To account for changes in the location of the tax kink over time, the data is rescaled relative to the threshold when pooling the tax years, and a weighted average for the threshold value and the tax change at the kink points are used to calculate the elasticity. See Lediga et al. (2018) for further details on the estimation strategy.

[29] http://www.sars.gov.za/Tax-Rates/Income-Tax/Pages/Companies-Trusts-and-Small-Business-Corporations.aspx

The results are presented in Figures 8 and 9 and show sizable taxpayer bunching at the first kink in the SBC business tax schedule, which translates into large estimates for the elasticity of taxable income. Importantly, the elasticity estimate for the subgroup of extensive-margin evaders (identified in the 2008 or 2014 commercial registry comparison), 0.94 (left panel of Figure 8), does not exceed the one for the group of control firms that voluntarily registered with SARS, 0.78 (right panel of Figure 8).[30] Assuming that the observed tax responses correlate with firms' evasion behaviour, the evidence hence rejects the notion that the observed gap in reported income between extensive-margin evaders and firms that voluntarily registered with SARS is, in part, established by differences in intensive-margin tax evasion. A similar picture emerges at the second kink, where the elasticity estimates turn out to be smaller (see Lediga et al. (2018)), but do not substantially differ between extensive-margin evaders and firms that voluntarily registered with SARS (0.07 vs. 0.11).

Complementarily, we reassess the link between intensive-margin evasion and extensive-margin evasion by drawing on data from taxpayer audits. Specifically, we assess whether adjustments of taxable income in the course of taxpayer audits systematically differ between semi-formal and fully formal firms. The data include all major audits of business tax returns conducted by SARS between 2011 and 2015 (covering tax returns for the years 2009 to 2014).[31] In the following, we will restrict the analysis to tax returns with a positive taxable income. This leaves us with 61 audited tax returns from 60 extensive-margin evaders (all identified in the 2008 commercial registry comparison). Specification (1) in Table 8 reports the propensity of these extensive-margin evaders to observe an upward adjustment in their taxable income in the course of the taxpayer audit relative to other firms, conditioning on the year of the tax return, the firms' year of registration with CIPC, their two-digit industry and host region. In addition, the specification includes control variables for firm size, firm profitability (measured by the log of total assets and firms' initially reported taxable income relative to total assets) as well as firms' profitability relative to the average of its two-digit industry.[32]

[30]Note that b denotes the estimated excess mass of taxpayers, b_se is the corresponding standard error, and q is the number of polynomials used in the local polynomial regression to determine the counterfactual density at the kink (chosen based on the Bayesian information criterion). e is the estimated elasticity of taxable income.

[31]Note that SARS also conducts small-scale audits, where the risk engine identifies an evasion risk and issues a generic letter or email informing the company to revise the tax return and/or submit supporting documents. Once the documents are received, they are inspected by an official, who is not a tax auditor. These cases are considered to be assurance/authentication cases wherein no particular audit skill is required. We do not account for these so-called 'Small E audits' in our analysis.

[32]To avoid results driven by outliers, the profitability variables are winsorized at the 5% level. Furthermore, note that the relatively small number of taxpayer audits prevents us from implementing CEM models, as there are a limited number of 'control variable cells' that include both extensive margin evaders and fully-formal firms. We thus turn to OLS models that include the indicated variables as control regressors.

Including the respective control regressors accounts for potential effects of taxpayer characteristics on audit-related taxable income adjustments and systematic selection of firms for audits based on the observed taxpayer characteristics. The results suggest no significant difference in the propensity of semi-formal and fully formal firms to experience an increase in reported taxable income in the course of taxpayer audits. The same holds true when adjustments in the level of tax payments are used as a dependent variable (see Specification (2)). This - consistent with our prior findings - suggests that intensive-margin evasion does not contribute to the observed gap in reported taxable income and tax payable between extensive-margin evaders and fully formal firms - but that the gap is driven by size and profitability differences instead.

Sales and Asset Growth

Finally, we determine how extensive-margin evaders' firm sales and assets emerge after their forced registration with SARS. Specifically, our setting offers a unique opportunity to test the theoretical prediction of the missing middle literature that many firms in less developed economies remain inefficiently small in order to stay below the radar of the tax authorities. Testing this presumption has proved difficult in the past, as movements of firms between the shadow economy and the official sector are commonly not observed and such movements, on top, mark deliberate decisions that may correlate with firms' underlying sales and asset growth, thus preventing the estimation of a causal link. The commercial registry comparisons, in turn, exogenously forced semi-formal firms into formality and allow us to assess the emergence of their sales and asset growth relative to comparable control entities (conditional on tax return submission).

The results are presented in Table 9. Since we observe firm-level information from companies' tax returns in 2009 to 2014, we will assess the effect of the 2008 commercial registry comparison on sales and asset growth of the identified extensive-margin evaders in that time period relative to control entities. To mitigate the impact of outliers, we drop observations with sales and assets growth rates above 100% (where similar results emerge with different cut off choices). Specification (1) regresses firms' sales growth on a dummy variable for extensive-margin evaders. Specification (2) additionally conditions on differences in the geographic location, industry affiliation, firm age, and the lagged firm sales in a CEM approach. Both models suggest that extensive-margin evaders exhibit an elevated sales growth relative to control entities, with the quantitative gap amounting to 1.8 percentage points or 5.0% of a standard deviation in sales growth.[33]

Similar effects emerge for firm assets. Specification (3) shows that extensive-margin

[33]The average sales growth in our sample is 3.6%, with a standard deviation of 35.73.

evaders have a 0.8 percentage points higher asset growth rate than control entities, amounting to 2.3% of a standard deviation in asset growth.[34] In line with intuition, the increase in size after the forced SARS registration is, moreover, particularly pronounced for firms that also registered with SARS for VAT (see Specifications (4) to (7)), potentially reflecting that these entities reap more business benefits from formalizing their activities than entities not registered for VAT.[35]

5 Summary and Conclusion

Concluding, this study presented evidence on the fiscal and behavioural consequences of comparing business tax and commercial register to identify firms that failed to sign up with the tax authorities for business tax purposes. Using South Africa as a testing ground and relying on rich administrative micro data, we find that the comparisons resulted in the identification of a large number of non-compliant businesses. After their forced registration with SARS, the identified extensive-margin evaders are shown to exhibit a lower propensity than other entities to submit tax returns and, conditional on submitting a return, report less taxable income and lower tax payments than control firms. The identified income and tax gap is, moreover, found to relate to differences in firm size and profitability rather than differences in intensive-margin evasion behaviour. Consequently, while many extensive-margin evaders fail to be drawn into the tax net (as they never submit a tax return), others actually do become compliant. The latter group moreover, in line with 'missing-middle theories', exhibits increased sales and asset growth after their forced registration with SARS.

The implied revenue effects are moreover significant and dwarf the administrative costs of implementing the commercial registry comparisons, thus making the interventions highly successful from a cost-benefit perspective. Overall, the analysis suggests that synchronizing commercial registry and business tax register may be an effective element of government strategies to combat tax evasion behaviour.[36]

[34]The average asset growth in our sample amounts to 1.3%, with a standard deviation of 36.22.

[35]It is mandatory for a business in South Africa to register for VAT if the total value of taxable supplies made in any consecutive 12-month period exceeded R1 million. A business may also choose to register voluntarily for VAT if the value of taxable supplies made or to be made is less than R1 million, but has exceeded R50,000 in the previous 12-months period.

[36]After the end of our sample period, the South African revenue authority permanently synchronized the business tax registry with the CIPC register. From 2015 onwards, all firms that register with CIPC are automatically registered with SARS for business tax purposes.

References

Bachas, P. and Soto, M. (2015). Not (ch) your average tax system: corporate taxation under weak enforcement. *mimeo, World Bank Group*.

Baumann, F. and Friehe, T. (2016). Competitive pressure and corporate crime. *The BE Journal of Economic Analysis & Policy*, 16(2):647–687.

Besley, T. and Persson, T. (2013). Chapter 2 - taxation and development. In Auerbach, A. J., Chetty, R., Feldstein, M., and Saez, E., editors, *Handbook of public economics, vol. 5*, volume 5 of *Handbook of Public Economics*, pages 51 – 110. Elsevier.

Besley, T. and Persson, T. (2014). Why do developing countries tax so little? *Journal of Economic Perspectives*, 28(4):99–120.

Best, M. C., Brockmeyer, A., Kleven, H. J., Spinnewijn, J., and Waseem, M. (2015). Production versus revenue efficiency with limited tax capacity: Theory and evidence from pakistan. *Journal of Political Economy*, 123(6):1311–1355.

Branco, F. and Villas-Boas, J. M. (2015). Competitive vices. *Journal of Marketing Research*, 52(6):801–816.

Brockmeyer, A., Hernandez, M., Kettle, S., and Smith, S. (2016). Casting a wider tax net: Experimental evidence from costa rica. *World Bank - Policy Research Working Paper*, WPS7850.

Carrillo, P., Pomeranz, D., and Singhal, M. (2017). Dodging the taxman: Firm misreporting and limits to tax enforcement. *American Economic Journal: Applied Economics*, 9(2):144–64.

Chetty, R., Friedman, J. N., Olsen, T., and Pistaferri, L. (2011). Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from danish tax records *. *The Quarterly Journal of Economics*, 126(2):749–804.

DeGiorgi, G., Ploenzke, M., and Rahman, A. (2015). Small firms? formalization - the stick treatment. *The World Bank Group - Policy Research Working Paper*, (7318).

Dekker, V., Strohmaier, K., and Bosch, N. (2016). A data-driven procedure to determine the bunching window: An application to the netherlands. *Hohenheim Discussion Papers in Business, Economics and Social Sciences, Discussion Paper No. 05.2016.*

Dharmapala, D., Slemrod, J., and Wilson, J. D. (2011). Tax policy and the missing middle: Optimal tax remittance with firm-level administrative costs. *Journal of Public Economics*, 95(9-10):1036–1047.

Dulleck, U., Fooken, J., Newton, C., Ristl, A., Schaffner, M., and Torgler, B. (2016). Tax compliance and psychic costs: Behavioral experimental evidence using a physiological marker. *Journal of Public Economics*, 134:9 – 18.

Fuest, C. and Riedel, N. (2009). Tax evasion, tax avoidance and tax expenditures in developing countries: A review of the literature. *Report prepared for the UK Department for International Development (DFID)*, pages 1–69.

Gangl, K., Kirchler, E., Lorenz, C., and Torgler, B. (2015). Wealthy tax non-filers in a developing nation: The roles of taxpayer knowledge, perceived corruption and service orientation in pakistan. *CREMA - Working Paper*, (2015-08).

GIZ (2013). Regional itc workshop taxation in asean - identifying experiences and lessons learned in six countries. Author: Eckardt, U.

Hoekstra, W., Pereira, F., and Queijo, J. (2013). Unlocking tax-revenue collection in rapidly growing markets. *Public Sector Report, McKinsey&Company*.

Iacus, S. M., King, G., and Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political analysis*, 20(1):1–24.

Kleven, H. J., Knudsen, M. B., 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.

Kleven, H. J. and Waseem, M. (2013). Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from pakistan *. *The Quarterly Journal of Economics*, 128(2):669–723.

Lediga, C., Riedel, N., and Strohmaier, K. (2018). The elasticity of corporate taxable income - evidence from south africa. *University of Bochum - mimeo*.

Mascagni, G. (2017). From the lab to the field: A review of tax experiments. *Journal of Economic Surveys*, (forthcoming).

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

Russell, B. (2010). *Revenue administration: managing the shadow economy*. Number 2010-2014. International Monetary Fund.

Saez, E. (2010). Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy*, 2(3):180–212.

Schneider, F. (2012). The shadow economy and work in the shadow: what do we (not) know? *IZA Discussion Paper*, (6423).

Slemrod, J. and Gillitzer, C. (2013). *Tax Systems*. MIT Press.

Waseem, M. (2018a). Taxes, informality and income shifting: Evidence from a recent pakistani tax reform. *Journal of Public Economics*, 157:41–77.

Waseem, M. (2018b). Using movement of exemption cutoff to estimate tax evasion: Evidence from pakistan. *University of Manchester, mimeo*.

6 Figures and Tables

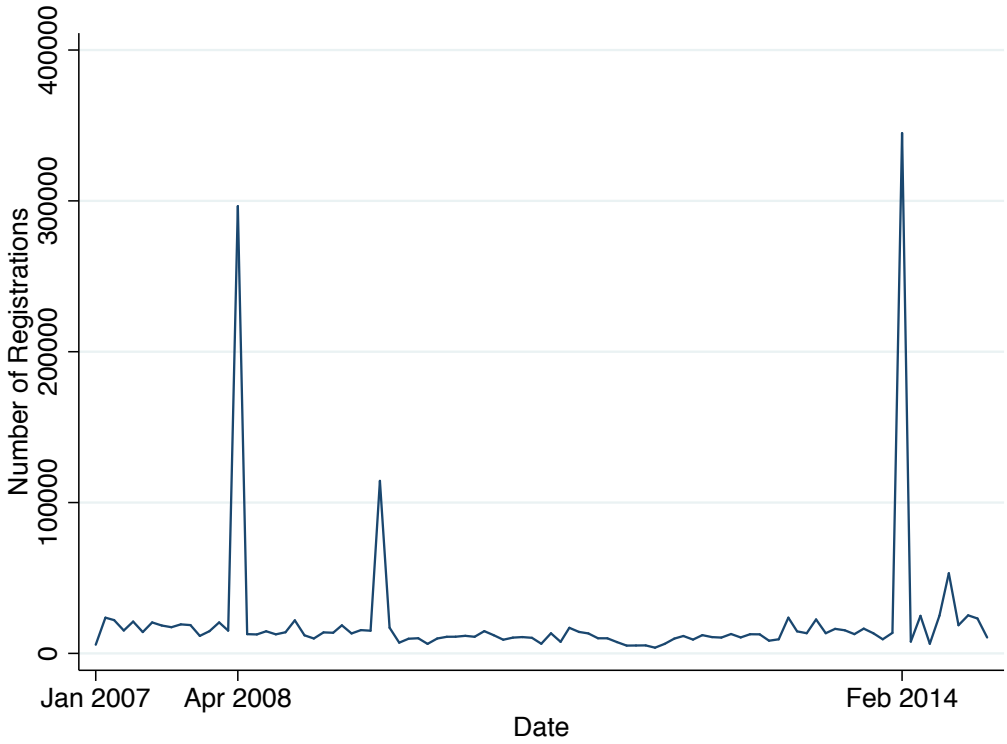
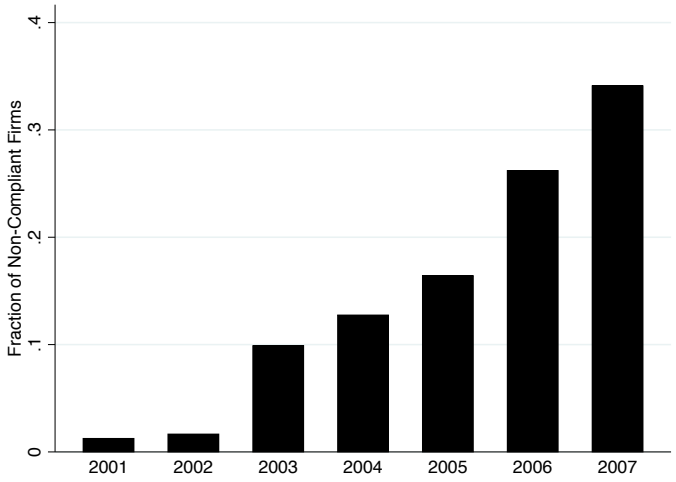
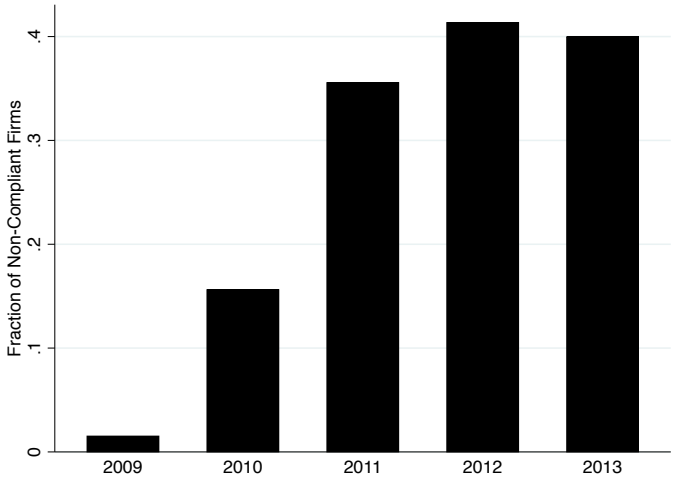


Figure 1: New Registrations for Business Tax Purposes with SARS per Month

Notes: The figure depicts the number of new firm registrations with SARS's business tax registry per registration month between January 2007 and December 2014.



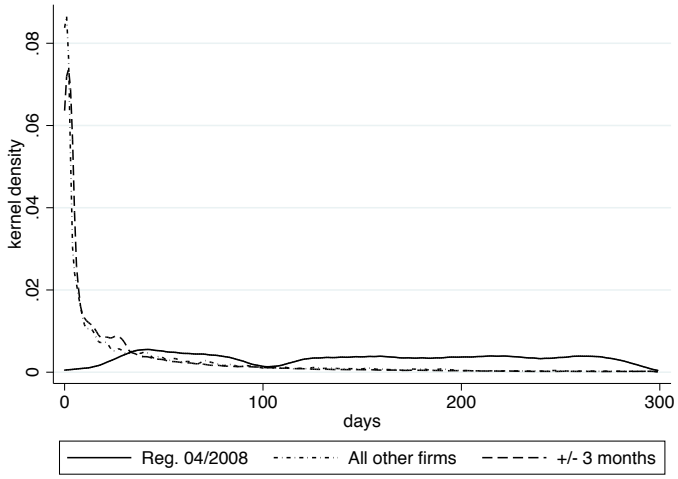
Commercial Registry Comparison 2008



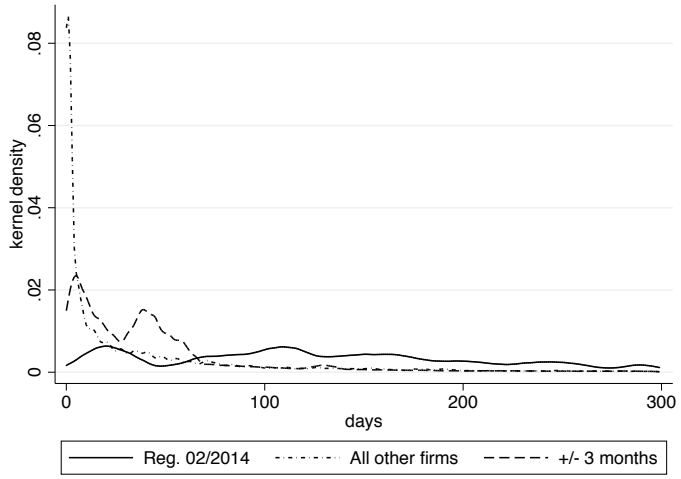
Commercial Registry Comparison 2014

Figure 2: Fraction of Non-Compliant Firms per Registration Year with CIPC

Notes: The figure depicts the number of firms that were identified as non-compliant with their SARS registration requirement in the 2008 commercial registry comparison (left panel) and the 2014 commercial registry comparison (right panel) as a fraction of all entities that had registered with CIPC in a given year. Note that our data includes only those firms that either voluntarily signed up with SARS or were identified as non-compliant in the commercial register comparisons, i.e. were active in the commercial register at the time of the comparisons. Firms that registered and deregistered with CIPC before the commercial registry comparisons are, in turn, not included if they failed to sign up with SARS for business tax purposes.



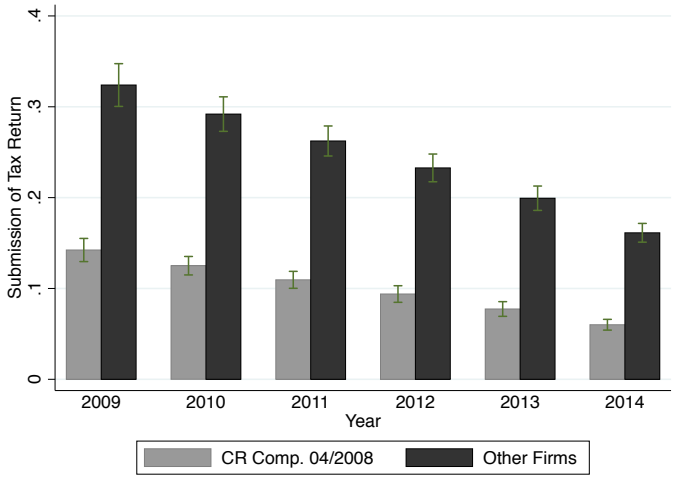
Commercial Registry Comparison 2008



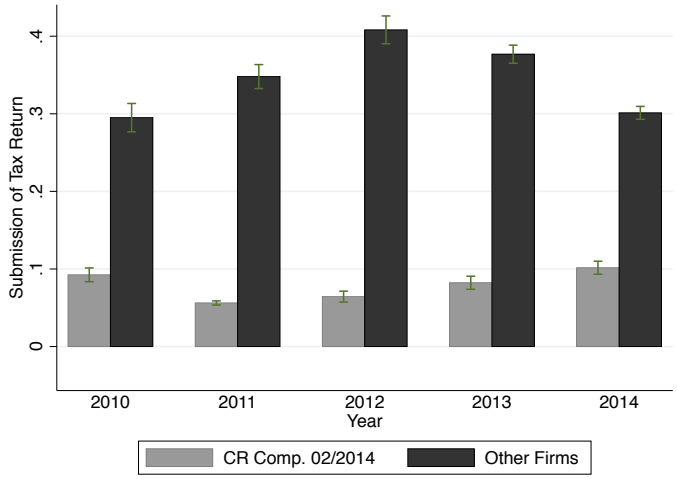
Commercial Registry Comparison 2014

Figure 3: Number of Days Between Firms' CIPC and SARS Registration

Notes: The figure depicts kernel density estimates for the number of days between a firm's registration with CIPC and with SARS for entities identified as non-compliant with their SARS registration requirement in the 2008 commercial register comparison (left) and the 2014 commercial register comparison (right), see the kernel density estimates depicted by the solid line. This is compared to kernel density estimates for all other firms that voluntarily registered with SARS (dotted line) and for all firms that voluntarily registered with SARS within a three-month-period around the dates of the commercial register comparisons in 2008 (left panel) and 2014 (right panel) respectively (dashed line).



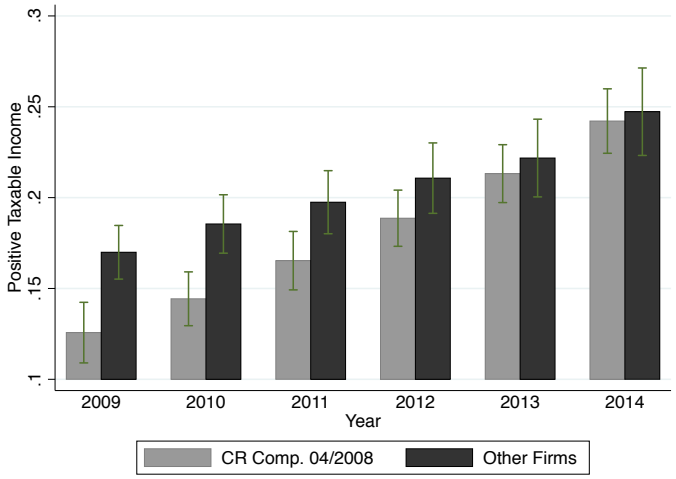
Commercial Registry Comparison 2008



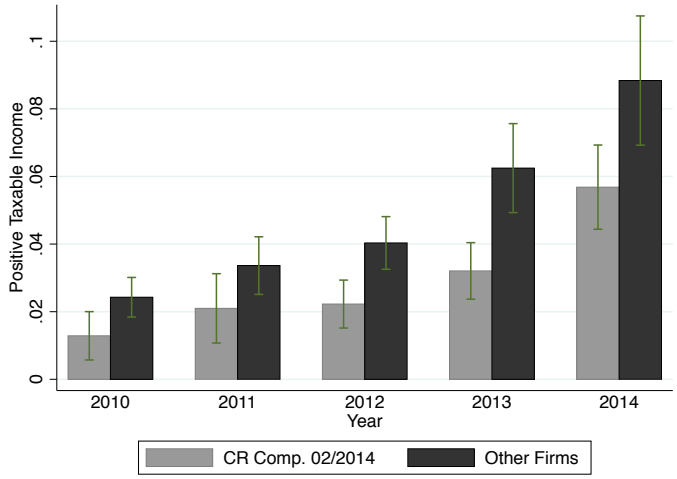
Commerical Registry Comparison 2014

Figure 4: Propensity to Submit a Tax Return

Notes: The figure depicts the average propensity of firms identified as non-compliant with their SARS registration requirement in the 2008 commercial register comparison (left) and the 2014 commercial register comparison (right) to submit a tax return with SARS in years in which they were active, relative to control groups of firms which complied with the obligation to register with SARS for business tax purposes. The sample is restricted to firms that registered with CIPC between 2001 and 2008 (left panel) and between 2008 and 2014 (right panel) respectively. Observations are weighted using CEM-weights that account for firms' host area and registration year with CIPC. Moreover, 95% confidence bounds are reported, where the calculation of standard errors allows for clustering of errors at the level of tax authority districts.



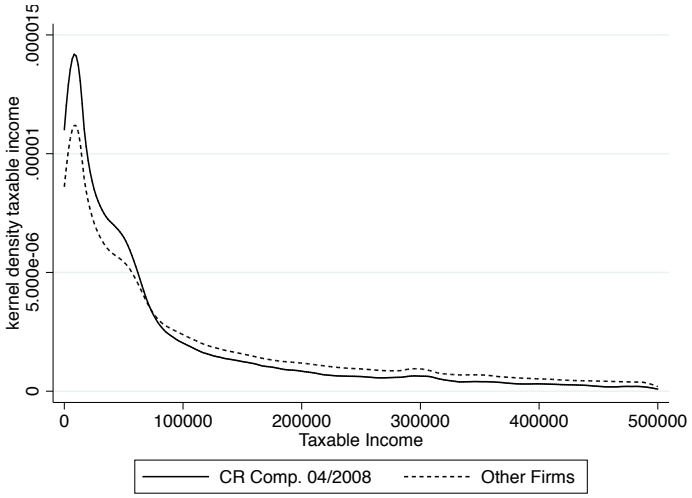
Commercial Registry Comparison 2008



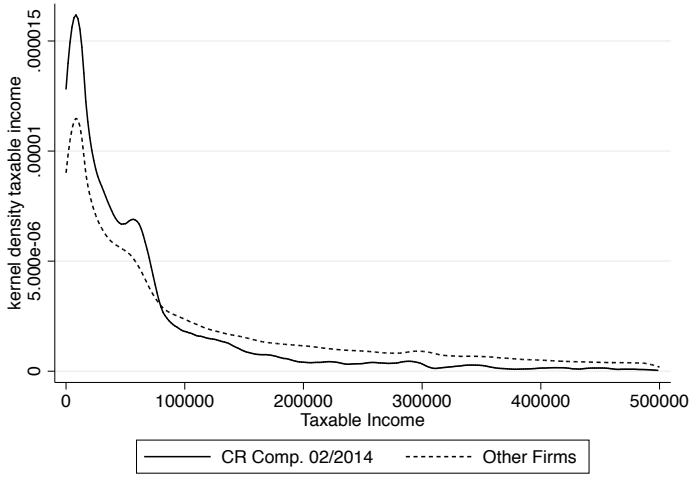
Commercial Registry Comparison 2014

Figure 6: Propensity to Submit Positive Taxable Income, Conditional on Submitting a Tax Return

Notes: The figure depicts the average propensity of firms identified as non-compliant with their SARS registration requirement in the 2008 commercial register comparison (left) and the 2014 commercial register comparison (right) to report positive taxable income, conditional on submitting a tax return, relative to control groups of firms which complied with the obligation to register with SARS for business tax purposes. Observations are weighted using CEM-weights that account for firms' host area, registration year with CIPC and 2 digit-industry affiliation. Moreover, 95% confidence bounds are reported, where the calculation of standard errors allows for clustering of errors at the level of tax authority districts.



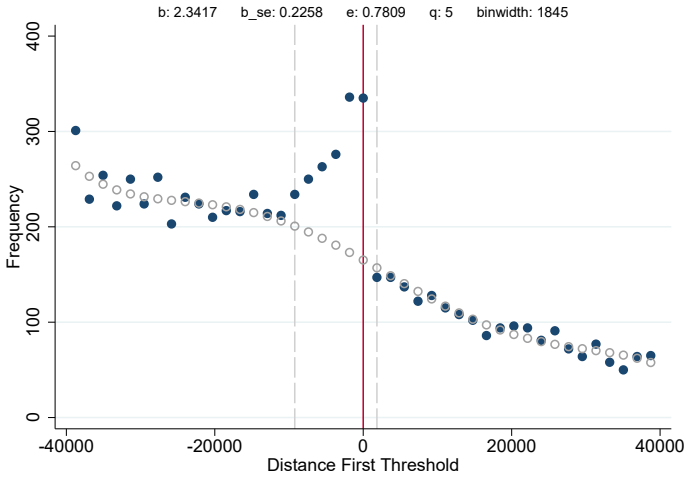
Commercial Registry Comparison 2008



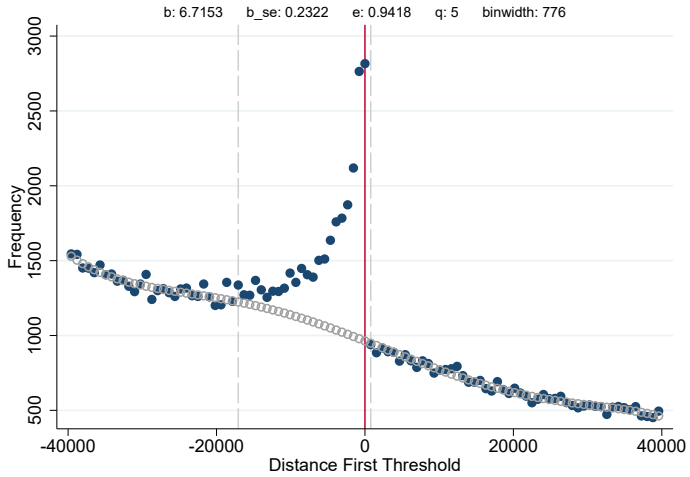
Commercial Registry Comparison 2014

Figure 7: Income Distribution, Conditional on Submitting a Tax Return with Positive Taxable Income

Notes: The figure depicts kernel density estimates for the taxable income of firms that were identified as non-compliant with their SARS registration requirement in the 2008 commercial register comparison (left) and the 2014 commercial register comparison (right), see the kernel density estimates indicated by the solid lines. These are compared to kernel density estimates for control firms indicated by the dashed lines. Observations are moreover reweighted using CEM-weights that account for firms' host area, registration year with CIPC and 2 digit-industry affiliation.



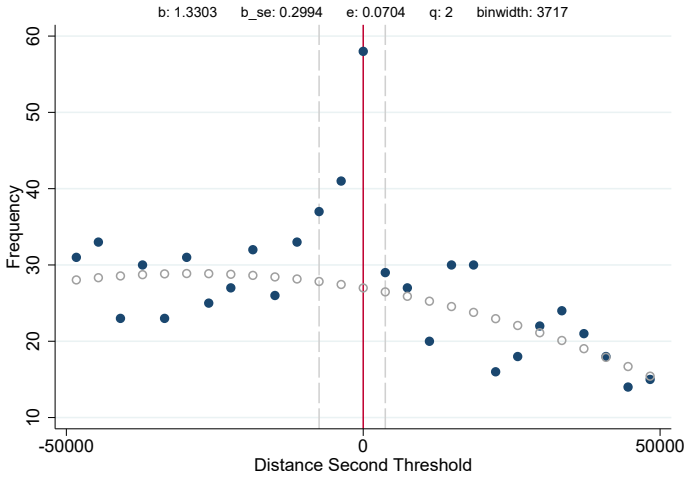
Extensive-margin Evaders



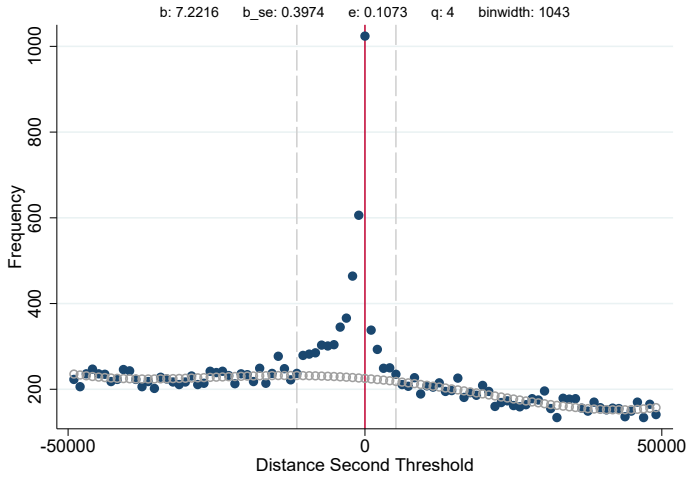
Fully Formal Control Firms

Figure 8: Bunching at Kinks in the SBC Tax Schedule - First Kink

Notes: The figure depicts the distribution of taxpayers around the first kink in the progressive South African tax schedule for Small Business Corporations, separately for firms that were identified as non-compliant with their SARS registration requirement in either the 2008 or the 2014 commercial registry comparisons (left panel, Extensive-margin Evaders) and for other firms that voluntarily registered with SARS for business tax purposes (right panel, Fully Formal Control Firms). The figures show the observed distribution (blue filled dotted line) and the estimated counterfactual (grey non-filled dotted line) of taxable income at the first threshold for the tax years 2010 to 2016. b indicates the excess mass of taxpayers, be_se the corresponding standard error and e the elasticity estimate. 'binwidth' is the width of the income bins, q is the number of polynomials in the local polynomial regression chosen based on the BIC criterion. See Lediga et al. (2018) for details.



Extensive-margin Evaders



Fully Formal Control Firms

Figure 9: Bunching at Kinks in the SBC Tax Schedule - Second Kink

Notes: The figure depicts the distribution of taxpayers around the second kink in the progressive South African tax schedule for Small Business Corporations, separately for firms that were identified as non-compliant with their SARS registration requirement in either the 2008 or the 2014 commercial registry comparison (left panel, Extensive-margin Evaders) and for other firms that voluntarily registered with SARS for business tax purposes (right panel, Fully Formal Control Firms). The figures show the observed distribution (blue filled dotted line) and the estimated counterfactual (grey non-filled dotted line) of taxable income at the second threshold for the tax years 2010 to 2016. b indicates the excess mass of taxpayers, be_se the corresponding standard error and e the elasticity estimate. 'binwidth' is the width of the income bins constructed, q is the number of polynomials in the local polynomial regression chosen using the BIC criterion. See Lediga et al. (2018) for details.

Table 1: Area Distribution of Forcedly Registered Firms

	(1)	(2)	(3)	(4)	(5)	(6)
Registrations 04/2008	0.0161***		0.0133***		0.0107***	
	(0.0010)		(0.0010)		(0.0007)	
Registrations 02/2014		0.0212***		0.0157***		0.0213***
		(0.0010)		(0.0010)		(0.0007)
Dep. Var	Gauteng	Gauteng	High Dens1	High Dens 1	High Dens 2	High Dens 2
Observations	1,605,648	1,389,487	1,605,648	1,389,487	1,605,648	1,389,487
Year Reg.	2001-2008	2009-2014	2001-2008	2009-2014	2001-2008	2009-2014

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. The dependent variable are indicators for firm location in Gauteng province (Specifications (1) and (2)), Gauteng or KwaZulu Natal province ('High Dens1', Specifications (3) and (4)) and Gauteng, KwaZulu Natal, Mpumalanga, Limpopo or Western Cape province ('High Dens 2', Specifications (5) and (6)) respectively. 'Registrations 04/2008' is a dummy variable indicating firms that were identified as 'semi-formal' in SARS's 2008 commercial register comparison, 'Registrations 02/2014' is a dummy variable indicating firms that were identified as 'semi-formal' in SARS's 2014 commercial register comparison. In Specifications (1), (3), (5), the sample is restricted to firms that registered with CIPC between 2001 and 2008 (as most of the extensive-margin evaders identified in the 2008 comparison registered with CIPC in that time period); in Specifications (2), (4), (6), the sample is restricted to firms that registered with CIPC between 2009 and 2014 (as most extensive-margin evaders identified in the 2014 comparison registered with CIPC in that time period).

Table 2: Propensity to Submit Tax Returns - Registrations 04/2008

	(1)	(2)	(3)	(4)	(5)	(6)
Registrations 04/2008	-0.1783***	-0.1816***	-0.1680***	-0.1669***	-0.1562***	-0.1528***
	(0.0009)	(0.0009)	(0.0009)	(0.0009)	(0.0009)	(0.0009)
Constant	0.3206***	0.3239***	0.2931***	0.2920***	0.2657***	0.2623***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Observations	1,545,257	1,542,887	1,526,128	1,523,814	1,512,063	1,509,899
Match	No	Yes	No	Yes	No	Yes
Tax Year	2009	2009	2010	2010	2011	2011
	(7)	(8)	(9)	(10)	(11)	(12)
Registrations 04/2008	-0.1434***	-0.1388***	-0.1276***	-0.1219***	-0.1072***	-0.1012***
	(0.0004)	(0.0004)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Constant	0.2373***	0.2327***	0.2049***	0.1993***	0.1673***	0.1613***
	(0.0004)	(0.0004)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Observations	1,497,665	1,495,562	1,484,780	1,482,814	1,467,179	1,465,237
Matched	No	Yes	No	Yes	No	Yes
Tax Year	2012	2012	2013	2013	2014	2014

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Specifications are run separately for each tax year. The sample is restricted to firms which registered with CIPC between 2001 and 2008 and are deemed to be active in a given tax year. Dependent variable: Dummy indicating whether firm i submitted a tax return in year t. 'Registrations 04/2008' is a dummy variable indicating firms that were identified as 'semi-formal' in SARS's 2008 commercial registry comparison. 'No'-entries in the 'Matched'-row refer to models without control variables, 'Yes'-entries to CEM models which account for firms' registration year with SARS and their host area (as determined by the tax office responsible for the taxpayer). 'Tax Year' is the year of the tax return. Note that observation numbers slightly differ between the models without control variables and the CEM models as 'area-CIPC registration year cells' that do not include both, 'semi-formal' and 'fully-formal' firms are assigned a CEM-weight of zero and drop out of the analysis.

Table 3: Propensity to Submit Tax Returns - Registrations 02/2014

	(1)	(2)	(3)	(4)	(5)	(6)
Registrations 02/2014	-0.2062***	-0.2026***	-0.2533***	-0.2917***	-0.2500***	-0.3438***
	(0.0066)	(0.0065)	(0.0022)	(0.0023)	(0.0014)	(0.0014)
Constant	0.2736***	0.2950***	0.3096***	0.3480***	0.3143***	0.4082***
	(0.0009)	(0.0009)	(0.0007)	(0.0007)	(0.0006)	(0.0006)
Observations	267,392	266,396	476,512	474,582	671,401	669,267
Match	No	Yes	No	Yes	No	Yes
Tax Year	2010	2010	2011	2011	2012	2012
	(7)	(8)	(9)	(10)	(11)	(12)
Registrations 02/2014	-0.2142***	-0.2946***	-0.1452***	-0.1996***	-0.2466***	-0.2719***
	(0.0010)	(0.0011)	(0.0008)	(0.0009)	(0.0009)	(0.0009)
Constant	0.2964***	0.3768***	0.2468***	0.3012***	0.3577***	0.3830***
	(0.0005)	(0.0005)	(0.0004)	(0.0005)	(0.0005)	(0.0005)
Observations	912,283	909,960	1,175,525	1,173,235	1,178,986	1,176,683
Matched	No	Yes	No	Yes	No	Yes
Tax Year	2013	2013	2014	2014	Reg. Year	Reg. Year

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Specifications are run separately for each tax year. The sample is restricted to firms which registered with CIPC between 2009 and 2014 and are deemed to be active in the indicated tax year. Dependent variable: Dummy indicating whether firm i submitted a tax return in year t. 'Registrations 02/2014' is a dummy variable indicating firms that were identified as 'semi-formal' in SARS's 2014 trade register comparison. 'No'-entries in the 'Match'-row refer to models without control variables, 'Yes'-entries to CEM models which account for firms' registration year with SARS and their host area (as determined by the tax office responsible for the taxpayer). 'Tax Year' is the year of the tax return. Note that observation numbers slightly differ between the models without control variables and the CEM models as 'area-CIPC registration year cells' that do not include both, 'semi-formal' and 'fully-formal' firms are assigned a CEM-weight of zero and drop out of the analysis. Moreover, we do not report results for the tax year 2009, as only 39 extensive-margin evaders identified in the 2014 commercial registry comparison were required to submit a return in that year.

Table 4: Propensity to Report Positive Taxable Income - Registrations 04/2008

	(1)	(2)	(3)	(4)	(5)	(6)
Registration 04/2008	-0.0772***	-0.0442***	-0.0695***	-0.0412***	-0.0615***	-0.0322***
	(0.0021)	(0.0019)	(0.0023)	(0.0022)	(0.0025)	(0.0024)
Constant	0.2029***	0.1699***	0.2138***	0.1855***	0.2268***	0.1975***
	(0.0006)	(0.0006)	(0.0007)	(0.0006)	(0.0007)	(0.0007)
Observations	445,651	437,103	400,882	395,271	359,112	354,463
Match	No	Yes	No	Yes	No	Yes
Tax Year	2009	2009	2010	2010	2011	2011
	(7)	(8)	(9)	(10)	(11)	(12)
Registration 04/2008	-0.0545***	-0.0221***	-0.0496***	-0.0086***	-0.0451***	-0.0051
	(0.0028)	(0.0026)	(0.0031)	(0.0030)	(0.0037)	(0.0035)
Constant	0.2432***	0.2107***	0.2628***	0.2218***	0.2872***	0.2473***
	(0.0008)	(0.0008)	(0.0009)	(0.0008)	(0.0010)	(0.0010)
Observations	316,726	311,415	270,211	263,853	217,154	211,984
Match	No	Yes	No	Yes	No	Yes
	2012	2012	2013	2013	2014	2014

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. The specifications are run separately for each tax year. The sample is restricted to firms which registered with CIPC between 2001 and 2008 and firms only enter in tax years in which they submitted a tax return. Dependent variable: Dummy indicating whether firm i submitted positive taxable income in year t, conditional on submitting a tax return. 'Registrations 04/2008' is a dummy variable indicating firms that were identified as 'semi-formal' in SARS's 2008 commercial registry comparison. 'No'-entries in the 'Match'-row refer to models without control variables, 'Yes'-entries to CEM models which account for firms' registration year with SARS, their host area (as determined by the tax office responsible for the taxpayer) and 2-digit industry affiliation. 'Tax Year' is the year of the tax return. Note that observation numbers slightly differ between the models without control variables and the CEM models as 'area-CIPC registration year-industry cells' that do not include both, 'semi-formal' and 'fully-formal' firms are assigned a CEM-weight of zero and drop out of the analysis.

Table 5: Propensity to Report Positive Taxable Income - Registrations 02/2014

	(1)	(2)	(3)	(4)	(5)
Registration 02/2014	-0.0632***	-0.0114**	-0.0836***	-0.0127***	-0.0939***
	(0.0095)	(0.0055)	(0.0053)	(0.0032)	(0.0032)
Constant	0.0760***	0.0243***	0.1046***	0.0336***	0.1162***
	(0.0009)	(0.0005)	(0.0008)	(0.0005)	(0.0007)
Observations	94,818	90,603	153,604	149,968	208,997
Match	No	Yes	No	Yes	No
Tax Year	2010	2010	2011	2011	2012
	(6)	(7)	(8)	(9)	(10)
Registration 02/2014	-0.0181***	-0.0983***	-0.0304***	-0.0943***	-0.0315***
	(0.0020)	(0.0023)	(0.0017)	(0.0020)	(0.0016)
Constant	0.0403***	0.1303***	0.0625***	0.1512***	0.0884***
	(0.0004)	(0.0007)	(0.0005)	(0.0007)	(0.0006)
Observation	205,255	251,304	248,662	254,848	252,668
Match	Yes	No	Yes	No	Yes
Tax Year	2012	2013	2013	2014	2014

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. The specifications are run separately for each tax year. The sample is restricted to firms which registered with CIPC between 2009 and 2014 and firms only enter in tax years in which they submitted a tax return. Dependent variable: Dummy indicating whether firm i submitted positive taxable income in year t, conditional on submitting a tax return. 'Registrations 02/2014' is a dummy variable indicating firms that were identified as 'semi-formal' in SARS's 2014 commercial registry comparison. 'No'-entries in the 'Match'-row refer to models without control variables, 'Yes'-entries to CEM models which account for firms' registration year with SARS, their host area (as determined by the tax office responsible for the taxpayer) and 2-digit industry affiliation. 'Tax Year' is the year of the tax return. Note that we do not report results for the tax year 2009 as only 22 extensive-margin evaders identified in the 2014 comparison handed in tax returns for that year. Note that observation numbers slightly differ between the models without control variables and the CEM models as 'area-CIPC registration year-industry cells' that do not include both, 'semi-formal' and 'fully-formal' firms are assigned a CEM-weight of zero and drop out of the analysis.

Table 6: Taxable Income - Registrations 04/2008

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Registrations 04/2008	-1.3848***	-0.7946***	-0.5493***	-0.3274***	-0.8783***	-0.5493***	-0.5410***	-0.3203***	-0.8150***	-0.5249***	-0.5354***	-0.2957***
	(0.0393)	(0.0431)	(0.0336)	(0.0294)	(0.0341)	(0.0339)	(0.0325)	(0.0285)	(0.0330)	(0.0330)	(0.0326)	(0.0284)
Log Total Assets				0.3127***				0.2980***				0.3015***
				(0.0020)				(0.0020)				(0.0020)
Observations	87,302	80,417	75,759	75,756	83,269	76,978	72,565	72,557	79,575	75,100	68,770	68,680
Match	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Tax Year	2009	2009	2009	2009	2010	2010	2010	2010	2011	2011	2011	2011
Sample	All	All	Tas	Tas	All	All	Tas	Tas	All	All	Tas	Tas
	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
Registrations 04/2008	-0.8203***	-0.5463***	-0.5901***	-0.3082***	-0.7640***	-0.5317***	-0.4472***	-0.1166***	-0.7547***	-0.5455***	-0.4555***	-0.1077***
	(0.0326)	(0.0327)	(0.0351)	(0.0304)	(0.0333)	(0.0338)	(0.0425)	(0.0349)	(0.0356)	(0.0359)	(0.0453)	(0.0376)
Log Total Assets				0.3438***				0.5814***				0.5906***
				(0.0025)				(0.0044)				(0.0052)
Observations	75,633	70,299	56,783	56,780	69,965	65,085	37,448	37,362	61,643	55,733	29,423	29,347
Match	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Tax Year	2012	2012	2012	2012	2013	2013	2013	2013	2014	2014	2014	2014
Sample	All	All	Tas	Tas	All	All	Tas	Tas	All	All	Tas	Tas

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Specifications are run separately for each tax year. The sample is restricted to firms which registered with CIPC between 2001 and 2008 and firms only enter in tax years in which they submitted a tax return with positive taxable income. Dependent variable: Natural log of taxable income of firm i in year t, conditional on submitting a tax return with positive taxable income. 'Registrations 04/2008' is a dummy variable indicating firms that were identified as semi-formal in SARS's 2008 commercial registry comparison. 'Log Total Assets' is the natural log of firms' total assets. 'No'-entries in the 'Match'-row refer to models without control variables, 'Yes'-entries to CEM models which account for firms' registration year with SARS, their host area (as determined by the tax office responsible for the taxpayer) and 2-digit industry affiliation. 'Tax Year' is the year of the tax return. 'Sample' 'All' indicates that all observations are used in the empirical analysis, while 'Tas' indicates restriction of the sample to firm-years with non-missing information on firms' total assets. Note that the sample size between Specification (1) and (2) differs because observations in 'area-CIPC registration year-industry cells' that do not include both, 'semi-formal' and 'fully-formal' firms are assigned a CEM-weight of zero and drop out of the analysis. An analogous explanations (now also accounting for categories of the coarsened total asset variable) rationalizes the slight difference in observation numbers between Specification (3) and (4). The same reasoning applies for the models run for the other tax years.

Table 7: Taxable Income - Registrations 02/2014

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Registrations 02/2014	-1.0469***	-0.5382*	0.6166	0.0444	-0.9453***	-0.6531***	-0.0761	-0.3027
	(0.2624)	(0.2885)	(0.6728)	(0.6003)	(0.1427)	(0.1489)	(0.4580)	(0.3859)
Log Total Assets				0.2420***				0.3286***
				(0.0088)				(0.0064)
Observations	15,790	9,307	2,950	2,950	23,328	19,663	6,640	6,616
Match	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Tax Year	2011	2011	2011	2011	2012	2012	2012	2012
Sample	All	All	Tas	Tas	All	All	Tas	Tas
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Registrations 02/2014	-0.8765***	-0.5271***	-0.7487***	-0.5590***	-0.7328***	-0.4142***	-0.4842***	-0.3584***
	(0.0831)	(0.0871)	(0.2093)	(0.1869)	(0.0512)	(0.0520)	(0.1098)	(0.1033)
Log Total Assets				0.4023***				0.3390***
				(0.0092)				(0.0092)
Observation	30,580	27,519	8,283	8,242	35,379	33,081	10,418	10,387
Match	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Tax Year	2013	2013	2013	2013	2014	2014	2014	2014
Sample	All	All	Tas	Tas	All	All	Tas	Tas

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Specifications are run separately for each tax year. The sample is restricted to firms which registered with CIPC between 2009 and 2014 and firms only enter in tax years in which they submit tax returns with positive taxable income. Dependent variable: Natural log of taxable income of firm i in year t, conditional on submitting a tax return with positive taxable income. 'Registrations 02/2014' is a dummy variable indicating firms that were identified as semi-formal in SARS's 2014 commercial registry comparison. 'Log Total Assets' is the natural log of firms' total assets. 'No'-entries in the 'Match'-row refer to models without control variables, 'Yes'-entries to CEM models which account for firms' registration year with SARS, their host area (as determined by the tax office responsible for the taxpayer) and 2-digit industry affiliation. 'Tax Year' is the year of the tax return. 'Sample' 'All' indicates that all observations are used in the empirical analysis, while 'Tas' indicates restriction of the sample to firm-years with non-missing information on firms' total assets. Note that the sample size between Specification (1) and (2) differs because observations in 'area-CIPC registration year-industry cells' that do not include both, 'semi-formal' and 'fully-formal' firms are assigned a CEM-weight of zero and drop out of the analysis. An analogous explanations (now also accounting for categories of the coarsened total asset variable) rationalizes the slight difference in observation numbers between Specification (3) and (4). The same reasoning applies for the models run for the other tax years. Note that we only report results for the tax years from 2011 onwards as a tiny fraction of extensive-margin evaders identified in the 2014 comparison submitted tax returns with a positive taxable income in 2009 and 2010.

Table 8: Audit Analysis		
	(1)	(2)
Registration 04/2008	-0.0209	0.0221
	(0.0585)	(0.0176)
Log Total Assets	-0.0045*	-0.0086***
	(0.0024)	(0.0007)
Profitability Dev. Industry Avg.	0.0424	-0.1256*
	(0.2156)	(0.0666)
Profitability	-0.0789	0.0739
	(0.2164)	(0.0669)
Dep. Var.	binary	level
Observations	11,712	11,148

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. The sample comprises all business tax returns with positive taxable income which were subject to major taxpayer audits between 2011 and 2015 (covering tax returns for the tax years 2009 to 2014). In Specification (1), the dependent variable is an indicator whether the firm experienced an upward adjustment in its taxable income in the course of the audit ('binary'). In Specification (2), the dependent variable captures the audit-related change in firms' tax payments due, relative to the taxable income initially reported by the firm ('level'). Both specifications control for firms' year of registration with CIPC, the tax year of the audited return, geographic location (as measured by tax office affiliation) and 2-digit industry. As reported, the models moreover control for firm size as measured by the log of total assets, firm profitability (as measured by firms' initially reported taxable income relative to total assets (winsorized at the 5% level)) and firm profitability relative to a 2-digit industry average.

Table 9: Sales and Asset Growth

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Registration 04/2008	0.0304***	0.0180***	0.0082***	0.0182***	0.0063*	0.0081*	0.0026
	(0.0025)	(0.0028)	(0.0029)	(0.0035)	(0.0036)	(0.0047)	(0.0051)
Log Sales, Lag		0.0079***		0.0072***		0.0000	
		(0.0005)		(0.0007)		(0.0014)	
Log Assets, Lag			0.0221***		0.0189***		0.0213***
			(0.0005)		(0.0006)		(0.0012)
Dep. Var.	Sales Growth	Sales Growth	Asset Growth	Sales Growth	Asset Growth	Sales Growth	Asset Growth
Observations	395,071	231,798	231,798	162,073	162,073	38,040	38,040
Match	No	Yes	Yes	Yes	Yes	Yes	Yes
Sample	All	All	All	VAT	VAT	No VAT	No VAT

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. The sample comprises all firms that registered with CIPC between 2001 and 2008 with their tax returns between 2009 and 2014. The dependent variable is the growth rate of the sales of firm i in year t in Specifications (1), (2), (4) and (6) and the growth rate of firms' total assets in Specifications (3), (5) and (7). To mitigate the impact of outliers, we disregard observations with sales or asset growth rates beyond 100%. 'Log Sales, Lag' is the lag of the natural logarithm of firm sales, 'Log Assets, Lag' is the lag of the natural logarithm of firms' total assets. 'No'-entries in the 'Match'-row refer to models without control variables, 'Yes'-entries to CEM models which account for firms' registration year with SARS, their host area (as determined by the tax office responsible for the taxpayer), 2-digit industry affiliation and lagged firm sales and total assets respectively. Sample splits in 'VAT' and 'No VAT' indicate firms that are and are not registered for VAT.

7 Online Appendix

The online appendix presents estimation models that compare reported tax payable (i.e. tax payments due) of extensive-margin evaders identified in the 2008 and 2014 commercial registry comparisons and control entities, conditional on tax return submission. Table A1 focuses on the 2008 commercial registry comparison and tests whether the propensity to report positive tax payable significantly differs between firms that were identified as 'semi-formal' in the commercial registry comparison and other entities that voluntarily registered with SARS. As with the analogous taxable income regressions presented in the main text, the sample is restricted to firms that entered the CIPC registry between 2001 and 2008 and table columns with uneven numbers refer to models that compare the unconditional mean across the two groups of firms, while columns with even numbers report CEM-estimates that account for firm age, host area and 2-digit industry affiliation. The results suggest that the propensity for positive tax payable is significantly reduced for the subgroup of extensive-margin evaders. This holds true for all sample years. For the tax year 2009, Specification (2) e.g. suggests that the propensity for positive tax payable is by 4.47 percentage points lower for the group of extensive-margin evaders relative to 'fully-formal' control firms (corresponding to 35.8%, evaluated at the average propensity for positive tax payments among 'fully-formal' entities (12.47%)). The significance of the findings moreover remains largely unaltered when the calculation of standard errors accounts for clustering of errors at the level of tax districts. Columns (3) to (12) in Table A1 report similar results for the other tax years and Table A2 shows that a similar result pattern emerges for the 2014 trade register comparison.

Analogously to the taxable income analysis in the main text, Table A3 moreover studies whether firms identified as non-compliant with their registration requirement in the 2008 comparison differ from fully formal firms in terms of the level of their taxes payable, conditional on tax payments being positive. Again the sample is restricted to firms that registered with CIPC between 2001 and 2008 and the organization of the table follows Table 6 in the main text. The models regress the log of tax payable on a dummy variable indicating extensive-margin evaders identified in the 2008 commercial registry comparison. The results show that 'semi-formal' firms report significantly lower levels of taxable income than their fully-formal counterparts. Column (3) of Table A3 suggests that the gap amounts to about 75%. Controlling for firm size (cf. Specification (4)) significantly reduces this difference by around 59%. Similar results are reported for the other tax years and for the commercial registry comparison in 2014, see Table A4.

Table A1: Propensity to Report Positive Tax Payable - Registrations 04/2008

	(1)	(2)	(3)	(4)	(5)	(6)
Registration 04/2008	-0.0820***	-0.0447***	-0.0783***	-0.0451***	-0.0743***	-0.0396***
	(0.0019)	(0.0017)	(0.0021)	(0.0019)	(0.0023)	(0.0021)
Observations	445,651	437,103	400,882	395,271	359,112	354,463
Match	No	Yes	No	Yes	No	Yes
Tax Year	2009	2009	2010	2010	2011	2011
	(7)	(8)	(9)	(10)	(11)	(12)
Registration 04/2008	-0.0721***	-0.0346***	-0.0681***	-0.0228***	-0.0667***	-0.0211***
	(0.0026)	(0.0024)	(0.0029)	(0.0027)	(0.0035)	(0.0032)
Observations	316,726	311,415	270,211	263,853	217,154	211,984
Match	No	Yes	No	Yes	No	Yes
Tax Year	2012	2012	2013	2013	2014	2014

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. The sample is restricted to firms which registered with CIPC between 2001 and 2008. Dependent variable: Dummy indicating whether firm i's tax return in year t results in positive tax payments due, conditional on submitting a tax return. 'Registrations 04/2008' is a dummy variable indicating firms that were identified as 'semi-formal' in SARS's 2008 commercial registry comparison. 'No'-entries in the 'Match'-row refer to models without control variables, 'Yes'-entries to CEM models which account for firms' registration year with SARS, their host area (as determined by the tax office responsible for the taxpayer) and 2-digit industry affiliation. 'Tax Year' is the year of the tax return. Note that observation numbers slightly differ between the models without control variables and the CEM models as 'area-CIPC registration year-industry cells' that do not include both, 'semi-formal' and 'fully-formal' firms are assigned a CEM-weight of zero and drop out of the analysis.

Table A2: Propensity to Report Positive Tax Payable - Registrations 02/2014

	(1)	(2)	(3)	(4)	(5)
Registration 02/2014	-0.0386***	-0.0096**	-0.0581***	-0.0136***	-0.0680***
	(0.0073)	(0.0043)	(0.0043)	(0.0025)	(0.0026))
Observations	94,818	90,603	153,604	149,968	208,997
Match	No	Yes	No	Yes	No
Tax Year	2010	2010	2011	2011	2012
	(6)	(7)	(8)	(9)	(10)
Registration 02/2014	-0.0169***	-0.0741***	-0.0250***	-0.0746***	-0.0266***
	(0.0016)	(0.0019)	(0.0014)	(0.0017)	(0.0013)
Observations	205,255	251,304	248,662	254,848	252,668
Match	Yes	No	Yes	No	Yes
Tax Year	2012	2013	2013	2014	2014

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. The sample is restricted to firms which registered with CIPC between 2009 and 2014. Dependent variable: Dummy indicating whether firm i's tax return in year t results in positive tax payments due, conditional on submitting a tax return. 'Registrations 02/2014' is a dummy variable indicating firms that were identified as 'semi-formal' in SARS's 2014 commercial registry comparison. 'No'-entries in the 'Match'-row refer to models without control variables, 'Yes'-entries to CEM models which account for firms' registration year with SARS, their host area (as determined by the tax office responsible for the taxpayer) and 2-digit industry affiliation. 'Tax Year' is the year of the tax return. Note that we do not report results for the tax year 2009 as only 22 extensive-margin evaders identified in the 2014 comparison handed in tax returns for that year. Note that observation numbers slightly differ between the models without control variables and the CEM models as 'area-CIPC registration year-industry cells' that do not include both, 'semi-formal' and 'fully-formal' firms are assigned a CEM-weight of zero and drop out of the analysis.

Table A3: Tax Payable - Registrations 04/2008

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Registration 04/2008	-1.8391***	-1.1240***	-0.7528***	-0.4452***	-1.1119***	-0.7650***	-0.7608***	-0.4509***	-1.0244***	-0.7478***	-0.7512***	-0.4363***
	(0.0532)	(0.0613)	(0.0452)	(0.0397)	(0.0457)	(0.0468)	(0.0437)	(0.0385)	(0.0437)	(0.0443)	(0.0425)	(0.0374)
Log Total Assets				0.3540***				0.3461***				0.3490***
				(0.0027)				(0.0027)				(0.0028)
Observations	68,673	62,239	57,955	57,951	64,739	58,929	55,068	55,061	62,857	58,161	53,092	53,011
Match	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Tax Year	2009	2009	2009	2009	2010	2010	2010	2010	2011	2011	2011	2011
Sample	All	All	Tas	Tas	All	All	Tas	Tas	All	All	Tas	Tas
	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
Registration 04/2008	-1.0187***	-0.7204***	-0.7148***	-0.3960***	-0.9785***	-0.7236***	-0.6326***	-0.2104***	-0.9589***	-0.7379***	-0.5536***	-0.1330***
	(0.0423)	(0.0434)	(0.0452)	(0.0391)	(0.0426)	(0.0442)	(0.0541)	(0.0440)	(0.0447)	(0.0460)	(0.0575)	(0.0473)
Log Total Assets				0.4221***				0.7227***				0.7154***
				(0.0034)				(0.0056)				(0.0063)
Observations	60,945	55,391	46,408	46,404	57,641	52,004	33,152	33,066	51,493	46,142	26,630	26,560
Match	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Tax Year	2012	2012	2012	2012	2013	2013	2013	2013	2014	2014	2014	2014
Sample	All	All	Tas	Tas	All	All	Tas	Tas	All	All	Tas	Tas

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. The sample is restricted to firms which registered with CIPC between 2001 and 2008. Dependent variable: Natural log of tax payable of firm i in year t, conditional on submitting a tax return which results in positive tax payable. 'Registrations 04/2008' is a dummy variable indicating firms that were identified as 'semi-formal' in SARS's 2008 commercial registry comparison. 'Log Total Assets' is the natural log of firms' total assets. 'No'-entries in the 'Match'-row refer to models without control variables, 'Yes'-entries to CEM models which account for firms' registration year with SARS, their host area (as determined by the tax office responsible for the taxpayer) and 2-digit industry affiliation. 'Tax Year' is the year of the tax return. 'Sample' 'All' indicates that all observations are used in the empirical analysis, while 'Tas' indicates restriction of the sample to firm-years with non-missing information on firms' total assets. Note that the sample size between Specification (1) and (2) differs because observations in 'area-CIPC registration year-industry cells' that do not include both, 'semi-formal' and 'fully-formal' firms are assigned a CEM-weight of zero and drop out of the analysis. An analogous explanations (now also accounting to categories of the coarsened total asset variable) rationalizes the slight difference in observation numbers between Specification (3) and (4). The same reasoning applies for the models run for the other tax years.

Table A4: Tax Payable - Registrations 02/2014

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Registration 02/2014	-1.3960***	-0.7879	0.1187	-0.4141	-1.1080***	-0.8694***	0.0503	-0.3366
	(0.5018)	(0.5489)	(0.8112)	(0.7180)	(0.2598)	(0.2683)	(0.6540)	(0.5455)
Log Total Assets				0.2871***				0.4354***
				(0.0137)				(0.0107)
Observations	9,723	4,094	1,568	1,568	15,171	9,358	4,052	4,029
Match	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Tax Year	2011	2011	2011	2011	2012	2012	2012	2012
Sample	All	All	tas	tas	All	All	tas	tas
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Registration 02/2014	-1.1687***	-0.8425***	-0.7933***	-0.6640***	-0.9118***	-0.5868***	-0.5781***	-0.5023***
	(0.1311)	(0.1382)	(0.2909)	(0.2576)	(0.0752)	(0.0777)	(0.1447)	(0.1345)
Log Total Assets				0.4760***				0.4244***
				(0.0122)				(0.0112)
Observations	20,889	17,078	6,243	6,202	24,478	21,387	8,814	8,785
Match	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Tax Year	2013	2013	2013	2013	2014	2014	2014	2014
Sample	All	All	tas	tas	All	All	tas	tas

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. The sample is restricted to firms which registered with CIPC between 2009 and 2014. Dependent variable: Natural log of tax payable of firm i in year t, conditional on submitting a tax return which results in positive tax payable. 'Registrations 02/2014' is a dummy variable indicating firms that were identified as 'semi-formal' in SARS's 2014 commercial registry comparison. 'Log Total Assets' is the natural log of firms' total assets. 'No'-entries in the 'Match'-row refer to models without control variables, 'Yes'-entries to CEM models which account for firms' registration year with SARS, their host area (as determined by the tax office responsible for the taxpayer) and 2-digit industry affiliation. 'Tax Year' is the year of the tax return. 'Sample' 'All' indicates that all observations are used in the empirical analysis, while 'tas' indicates restriction of the sample to firm-years with non-missing information on firms' total assets. Note that the sample size between Specification (1) and (2) differs because observations in 'area-CIPC registration year-industry cells' that do not include both, 'semi-formal' and 'fully-formal' firms are assigned a CEM-weight of zero and drop out of the analysis. An analogous explanations (now also accounting to categories of the coarsened total asset variable) rationalizes the slight difference in observation numbers between Specification (3) and (4). The same reasoning applies for the models run for the other tax years. Note that we only report results for the tax years from 2011 onwards as a tiny fraction of extensive-margin evaders identified in the 2014 comparison submitted tax returns with a positive taxable income in 2009 and 2010.

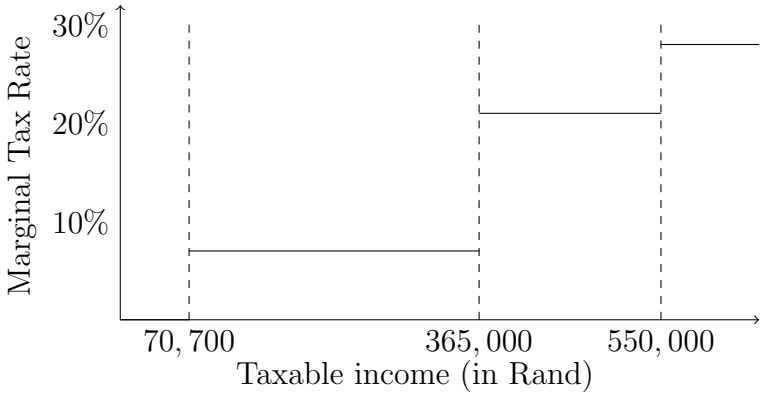


Figure A1: SBC-Tax Schedule in 2015