

Quasi-Experimental Evidence on Tobacco Tax Regressivity

Steven F. Koch

October 2017

Abstract

Tobacco taxes are known to reduce tobacco consumption and to be regressive, such that tobacco control policy may have the perverse effect of further harming the poor. However, if tobacco consumption falls faster amongst the poor than the rich, tobacco control policy can actually be progressive. We take advantage of persistent and committed tobacco control activities in South Africa to examine the household tobacco expenditure burden. For the analysis, we make use of two South African Income and Expenditure Surveys (2005/06 and 2010/11) that span a series of such tax increases and have been matched across the years, yielding 7806 matched pairs of tobacco consuming households and 4909 matched pairs of cigarette consuming households. By matching households across the surveys, we are able to examine both the regressivity of the household tobacco burden, and any change in that regressivity, and since tobacco taxes have been a consistent component of tobacco prices, our results also relate to the regressivity of tobacco taxes. Like previous research into cigarette and tobacco expenditures, we find that the tobacco burden is regressive; thus, so are tobacco taxes. However, we find that over the five-year period considered, the tobacco burden has decreased, and, most importantly, falls less heavily on the poor. Thus, the tobacco burden and the tobacco tax is less regressive in 2010/11 than in 2005/06. Thus, increased tobacco taxes can, in at least some circumstances, reduce the financial burden that tobacco places on households.

Keywords: Tobacco Tax Burden; Tobacco Control Policy; Tax Regressivity; South Africa; Genetic Matching; Inequality; Concentration Index

1 Introduction

According to the Cancer Society of South Africa (2013), around 44 000 people die each year of tobacco related diseases in South Africa. Thus, effecting a reduction in the number of tobacco linked deaths will hinge upon the ability to reduce tobacco usage. Increased cigarette taxation, along with a number of other control strategies, have been implemented around the world in order to increase either the direct or indirect cost of purchasing or consuming tobacco products, and, thus, reduce consumption. Raising the price of cigarettes is one of the most effective tobacco control strategies (Centres for Disease Control and Prevention, 2007) primarily because the empirical tobacco demand literature repeatedly provides support for the law of demand (Chaloupka et al., 2012).

However, tobacco taxes are regressive, and, therefore, increases in tobacco taxes could burden poor households, unless they reduce their smoking expenses at least enough to offset the increase in tax. Relatedly, taxation may yield additional burdens in smoking households, due to the budget reallocations that would be required to cover any additional tobacco consumption costs (see John, 2008; Tshiswaka-Kashalala and Koch, 2008; John et al., 2012; Chelwa and van Walbeek, 2014). Reversing these negative budget consequences will also hinge on the ability to reduce tobacco usage, or at least tobacco expenditure. Such a reduction could come in the form of reduced purchases, even cessation, as well as substitution towards less expensive tobacco products. Although there is evidence that higher prices lead to cessation (Gallet and List, 2003), more recent studies find lower quit elasticities than earlier studies (Franks et al., 2007; Farrelly et al., 2008). There is also evidence of substitution across tobacco products, in response to increased prices (Connelly et al., 2009; DeCicca et al., 013a; DeCicca et al., 013b).

Despite the preceding evidence on cessation and substitution, there is little evidence to suggest that these reductions are generally large enough to offset the regressive nature of tobacco taxes. However, Warner (2000), Chaloupka et al. (2011) and Chaloupka et al. (2012) argue that such an effect is very plausible. They argue that tobacco tax increases can be progressive, because the poor are more responsive to price increases than the rich. There is ample empirical evidence supporting the hypothesis that price elasticities are higher for poorer individuals

(Siahpush et al., 2009; DeCicca and McLeod, 2008; Matire et al., 2010), although not all studies support this finding (Franks et al., 2007). Despite Franks et al.'s (2007) finding, the evidence is overwhelming. Thomas et al.'s (2008) review of 84 papers argues that prices are more effective in reducing consumption amongst the poor. Hill et al.'s (2013) recent review of 77 papers offers further evidence. Thus, researchers and anti-smoking advocates are warming to the idea that, not only are tobacco taxes an important instrument in the fight for improved public health, they may also be a useful instrument in the fight against poverty.

Even though the empirical evidence overwhelmingly suggests that the poor are more responsive to price increases than those in better financial circumstances, there remains little empirical support that the medium term impact of tobacco taxation can be positive for the finances of, especially, poorer households. In other words, there is little evidence that tax increases might, over time, become less regressive. The evidence that does exist is negative. Brown et al.'s (2015) recent review of 29 studies finds only one suggesting that tobacco control programs have had a positive equity impact. Both Coleman and Remler (2008) and, to a lesser extent, Farrelly et al. (2012) find that the price responsiveness is not large enough to lead to reductions in tax regressivity. Similarly, Hoffer et al. (2015) conclude that cigarette purchases are inelastic, while tobacco taxes burden the poorest. Siahpush et al. (2013) finds similar evidence in Mexico. Following successive tax increases, expenditure on cigarettes in Mexico has risen, which suggests an increase in the tobacco burden; however, they did not directly consider regressivity. In South Africa, Vellios and van Walbeek (2016) find that higher prices reduce smoking 'onset' for men, but not for women. Given that women are generally poorer than men, their analysis suggests that tobacco taxes may have worsened regressivity. Although the preceding literature broadly suggests that increased taxes increase the household financial burden, Hu et al.'s (2016) recent findings are more positive. They find some evidence that (i) policies are associated with a reduction in prevalence, and (ii) that the prevalence reduction appears to be concentrated at the lower end of the socioeconomic spectrum. Our study seeks to contribute further recent evidence to this question, focusing attention on a country that has followed a consistent tobacco taxation policy, South Africa, as outlined by van Walbeek (2014).

In what follows, we examine the hypothesis that increases in cigarette taxes might lead to an increase in the progressiveness of tobacco taxes, and we do so in a middle income country.

The analysis is undertaken at the beginning and end of a five-year period (2005/06-2010/11), during which tobacco taxes increased approximately 30% (see Figure 1). Given the steady increases in tobacco tax during that period, and the relatively consistent ratio of tax to price (also see Figure 1), the tobacco tax underscores a near constant share of the overall financial burden of tobacco expenditure. Therefore, we compare the relative regressivity of the tobacco burden (tobacco tax) at the beginning and end of the period. Our analysis hinges upon genetic matching, an extension to both Mahalanobis and propensity score matching, which we undertake across the two most recent South African Income and Expenditure Surveys, in an effort to match tobacco and cigarette consuming households over time. Once households are matched, we consider the relative concentration of tobacco (cigarette) expenditures in those households, relative to their income. We find, as expected, that tobacco purchases are a greater financial burden to poorer households; thus, tobacco taxes are regressive. However, we also find that such purchases are less of a burden on poorer households at the end of the time period than at the beginning. In other words, tobacco taxes can yield an equity benefit; households can reduce their consumption enough to offset the increased costs, and tobacco taxes can be regressive.

Figure 1 about here.

2 The Data

Unfortunately, information not widely available in South Africa is individual level data on cigarette or tobacco consumption (survey questions regarding the number of cigarettes smoked per day, for example) and the price paid per pack. Therefore, in this analysis, we make use of consumption expenditure data, sourced from the 2005/06 and 2010/11 Income and Expenditure Surveys (IES) (Statistics South Africa, 2008, 2013). These are secondary and publicly available data, and ethical clearance for the project was approved by the University of Pretoria Faculty of Economics Research Ethics Committee. Starting in 2005/06 the quinquennial IESs were completed over a lengthy period of time, rather than a few weeks; in these surveys, the time frame was September through August. Since data was collected over time, all expenditure data was adjusted to March 2006 (2011) using the consumer price index. Furthermore, all

expenditure data was adjusted to 2008 real values using the appropriate CPI figures; income and expenditure were adjusted using the income CPI, while tobacco consumption was adjusted using the tobacco CPI.

Because the surveys contain information about consumption, they underpin South Africa's Consumer Price Index. In addition, the data allows for the examination of expenditure patterns in South African households, especially when it is combined with other information regarding individuals in the household. Previous South African tobacco studies have made extensive use of the 2000 Income and Expenditure Survey data, (van Walbeek, 2002a; Ground and Koch, 2008; Tshiswaka-Kashalala and Koch, 2008), the All Products Market Survey (AMPS) data (van Walbeek, 2002b), the 1993 Living Standards Measurement Survey (Berg and Kaempfer, 2001), wholesale cigarette shipments from one of the large manufacturers in South Africa (Boshoff, 2008) and the National Income Dynamics Study (Vellios and van Walbeek, 2016).

In comparing two different samples that do not arise from a panel, there are two primary options. The first is to use the given population weights, in an effort to make the analysis nationally representative, and compare those results. However, population weights in the IESs are defined to be representative by gender, race and five-year age band at the national level, and by gender, race and broader age bands at the provincial level. Our interest is in tobacco purchasing and cigarette purchasing households, which are not expected to be representative, if weighted by the aforementioned race, gender and age groups. Thus, we turn to a second option, that of matching households across the survey years, in order to create a pooled database of broadly similar tobacco consuming households, even though the results cannot be extended beyond the matches. Despite that limitation, we believe that comparing like with like is preferable.

A primary reason to pursue matching, rather than a simple comparative analysis of the two surveys, is that in a simple comparison some of the differences that we might observe across surveys may also be correlated with tobacco consumption decisions; thus, our estimates of tobacco burden regressivity and its changes could be confounded (Cochran and Rubin, 1973). A common approach follows Rosenbaum and Rubin (1983), wherein households would be matched by their propensity score; here that score measures their propensity to be in the 2010/11 survey,

based on their household characteristics (see below). However, matching only on the propensity score may yield poor balance.

Genetic matching (Diamond and Sekhon, 2013) is an alternative to the standard approach. It is an iterative matching algorithm that generalizes propensity score and Mahalanobis distance matching, and, therefore, it allows for additional dimensions to the matching algorithm, and, potentially results in better match balance. It makes use of an evolutionary search algorithm (Sekhon and Mebane, Jr., 1998) that optimizes balance between the observed covariates. In their Monte Carlo experiments, Diamond and Sekhon (2013) show that it performs better than common alternatives for matching, such as binary probability propensity scores and machine learning algorithms that have been shown to operate well for matching (Lee et al., 2010); such machine learning algorithms include boosted Classification and Regression Trees (Brieman et al., 1984; Bühlman and Yu, 2003) and random forests (Brieman, 2001).

For our implementation, in addition to matching on a predicted propensity score, we also included real household income and expenditure, household head schooling and two variables used in the cigarette budget share ‘burden’ literature. These variables are the proportion of adults in the household and the proportion of adult males amongst household adults (see John, 2008; Tshiswaka-Kashalala and Koch, 2008; John et al., 2012; Chelwa and van Walbeek, 2014). For the propensity score, all of the household structure and location variables listed in Table 1 were included in a logit regression, from which a predicted probability was determined. Finally, matching was undertaken, separately, for race and gender of household head, as well as urban and rural location of the household, and the resulting matched samples were re-pooled. Results for each of the matches are available from the authors, upon request. We undertook genetic matching (Diamond and Sekhon, 2013), using the “Matching” package (Sekhon, 2011) in R (R Core Team, 2016) to create matched and balanced 2005/06 and 2010/11 samples, and use those for the analysis.

2.1 Descriptive Statistics of Matched Data

We present post-match sample means from the IES 2005/06 and 2010/11 in Tables 1 and 2 along with the p -values arising from t -tests of mean differences across the samples and Kolmogorov-Smirnov (ks) tests of distributional differences; distributional differences are estimated only for continuous and ordinal variables. Table 1 contains the information for tobacco consuming households, while Table 2 contains the information for cigarette purchasing households. As can be seen in the tables, even after matching, the data is not perfectly aligned. However, the post-match balance is an improvement over the pre-match balance; additional analysis of both the pre-match data, and post-match balance **can be found in the online supplementary material**.

Table 1 about here.

In Table 1 we see that tobacco purchasing households have relatively older household heads (age group 10 corresponds to the age range 50-54), and, therefore, is relatively adult dominant. The head of household has some education, but has not generally completed primary school. That household also has more males than females. We also see that average real household income and expenditure are low, at approximately ZAR4900 and ZAR3900, respectively (8.25ZAR=1USD in 2008). Further information about the distribution of real household income for the matched samples is available in Figure A.1.

Table 2 about here.

In contrast to what we find with tobacco consuming households, seen in Table 1, cigarette purchasing households are rather different; see Table 2. Heads of households purchasing cigarettes tend to be younger and more educated than those from the more broadly defined tobacco consuming households; note that we define tobacco consumption to include cigarette consumption. Cigarette purchasing households are also composed of relatively more males and more adults than the tobacco consuming average. Finally, cigarette purchasing households are richer; their

real income and expenditure is nearly 30% higher at approximately ZAR6400 and ZAR5025, respectively. Further information about the distribution of real household income for the matched samples is available in Figure A.2.

2.2 Descriptive Statistics of Tobacco Purchases

Although it is important for there to be balance across households, once they have been matched, the primary purpose of the analysis is an examination of tobacco and cigarette purchasing behaviour with respect to income. Before turning to the methods and the results of that analysis, we turn to describing household tobacco purchasing behaviour across the two surveys. To do so, we estimate the means in each survey, using matched data, as well as the mean difference in each of the underlying household tobacco consumption components; see Table 3. We also include the p -value associated with a Wilcoxon signed-rank test on the matched samples. These p -values provide information on whether or not to reject the null hypothesis that the various paired tobacco consumption items are likely to have been drawn from a symmetric distribution centered at zero. The top panel (Table 3a) focuses on matched tobacco consuming households and the bottom panel (Table 3b) includes only matched cigarette consuming households. The variables included in the tables are total cigarette and tobacco expenditure (annualised real ZAR) and the share of the budget (relative to both household income and household expenditure) devoted to either tobacco or cigarettes.

Table 3 about here.

Across each of the subsamples, there is a consistent result. Following matching, households in 2010/11 spent less (in real terms) on cigarettes and tobacco products, and had lower cigarette and tobacco shares relative to both income and expenditure. In other words, we observe reductions in the cigarette and tobacco ‘burden’, despite increased taxes on cigarettes and tobacco between the surveys. Although we cannot say whether this reduction is underscored by a reduction in smoking or a move towards less expensive tobacco products, the results are encouraging, even if not large. Given the prices in Figure 1, the estimated average reduction is equivalent to

about one pack of cigarettes for tobacco consuming households and 1.5 packs of cigarettes per cigarette consuming household.

3 Methodology

Our analysis of the regressivity of tobacco expenditures is underpinned by a common definition of regressivity, which is extended to incorporate the fact that the available data spans a five-year period replete with many tax and price hikes. Therefore, it is possible to tie the underlying estimate of regressivity to the underlying consistency of tax hikes.

3.1 Regressivity

Consider the ratio of the total tobacco expenditure (C) to the share of total income (y). Regressivity is concerned with the relationship of this share to total income. We denote regressivity by r_ω , and define it by the relationship in (1)

$$r_\omega = \frac{\partial(C/y)}{\partial y} \tag{1}$$

If $r_\omega < 0$ – that is, the tobacco burden is higher for poorer households – the tobacco expenditure burden is defined to be regressive. It is also possible to examine changes in the regressivity of the burden. For example, consistently increased taxes can lessen regressivity if $r_\omega^{2010} > r_\omega^{2005}$, where superscripts refer to the survey year.

Tobacco taxes, as outlined in policy documents, are set to represent 52% of the retail price of the most popular brand. For this to be true, annually, tobacco taxes rise, annually (see Figure 1). Unfortunately, our data does not contain brand information, or even total pack purchases; thus, for purposes of this analysis, we need reasonable assumptions for identifying the regressivity in both surveys, along with the change in that regressivity.

Our primary assumption along these lines is that tobacco taxes account for a near constant share (ρ) of total tobacco expenditure. As can be seen in Figure 1, this is not an outrageous assumption for aggregate cigarette purchases. Under our assumption, with T defined as the tobacco tax burden, $T = \rho C$. From this, we can define tobacco tax regressivity, r_τ as a simple linear function of the tobacco burden regressivity measure.

$$r_\tau = \frac{\partial(T/y)}{\partial y} = \frac{\partial(\rho C/y)}{\partial y} = \rho \cdot r_\omega \quad (2)$$

Assuming that the the tobacco tax represents a constant share of total tobacco expenditures could be too optimistic. Tobacco users may choose to purchase less expensive tobacco, for example. Given that the policy rate is 52% of the most popular brand, the effect will depend on the relationship between the price of the actual brand smoked and that of the most popular brand. If smokers were already indulging in less expensive (relative to the most popular), the tobacco tax regressivity measure (for 2010) could be understated by our definition, although the overall tobacco burden regressivity would not be affected. Furthermore, if the measure of tax regressivity in 2010 is understated, observing $r_\tau^{2010} > r_\tau^{2005}$ would strengthen the case that tobacco tax regressivity has dropped; however, not observing would not necessarily mean tax regressivity has not dropped.

3.2 Estimating Regressivity

The preceding discussion suggests that the effect of tobacco tax changes can be examined through the analysis of tobacco expenditure shares at the beginning and the end of the period. The simplest comparison to make is based on tobacco share concentration curves and the concentration index (see Kakwani, 1977; O'Donnell et al., 2008, for more details). The curve itself provides an illustration of the concentration of tobacco budget shares amongst the poorest households relative to the richest. It is a plot of the cumulative percentage of tobacco burdens against the cumulative percentage of the sample, ranked from poorest to richest; therefore, the illustration is easily interpretable. If poor households face relatively higher tobacco burdens, then the curve will lie above the line of equality, the 45° line; further, tobacco consumption

(and thus the tax) will then be deemed regressive. The index measures the area between the concentration curve and the line of equality; if the curve lies above, the index will be negative; if below, the index will be positive. Thus, the index is one estimate for r_ω , it can be estimated for each year, and the difference between them can be calculated. The concentration curve analysis is undertaken using the “IC2” package (Plat, 2012) in R (R Core Team, 2016).

A different option is to estimate the parameter r_ω and its change through a linear conditional expectation of the tobacco or cigarette expenditure share ($\omega = C/y$), where conditioning is over income (we use real log income) a 2010 survey indicator (D_{2010}), and the interaction between the two.

$$\omega_i = \beta_0 + \beta_1 \ln y_i + \beta_2 D_{2010,i} + \beta_3 D_{2010,i} \ln y_i + \nu_i \quad (3)$$

However, such a regression only tells us the relationship at the mean of the data, and, since we are interested in the relationship for the poor, we use the linear structure to underpin a quantile regression (Koenker and Basset, Jr., 1978), which we estimate at the 25th, 50th and 75th percentiles of the distribution of tobacco (or cigarette) expenditure shares. The primary parameter of interest in (3) is β_3 , which is an estimate of $r_\omega^{2010} > r_\omega^{2005}$, the change in regressivity over time. We estimate the quantile regressions using Koenker’s (2016) “quantreg” package for R (R Core Team, 2016).

4 Results

4.1 Concentration Curves and Concentration Index

Initial descriptive statistics, see Table 3, suggest that the tobacco burden has lessened in South Africa between 2005/06 and 2010/11. However, that average reduction does not provide us with any information on distributional aspects. Therefore, we turn to concentration curves (see Figures 2 and 3) and the estimation of Kakwani’s (1977) concentration index for household tobacco and cigarette budget burdens (see Table 4). As can be seen in the figures, the curve lies above the 45° line. Thus, cigarette and tobacco burdens are regressive. In other words, when

poor households purchase tobacco and cigarettes, poor households will devote a larger share of their earnings to those purchases than rich households.

Figures 2 and 3, as well as Table 4 about here.

The concentration of the tobacco and cigarette burdens, as can be seen in the figures and table, has fallen over time. Despite the regressive nature of these purchases, they are now less concentrated amongst the poor. The improvement is larger for matched tobacco consuming households (0.0225), and smaller for matched cigarette consuming households (0.0113). Thus, not only has household average cigarette and tobacco expenditure, as well as their burdens, fallen over time, that reduction has been relatively more pronounced amongst the poor. However, even though the average reduction in cigarette and tobacco consumption has been larger amongst matched cigarette consuming households (see Table 3), and, overall, we see that the reduction has been relatively larger amongst poorer than richer households (see Table 4), the regressivity reduction has been less pronounced in the matched cigarette consuming subgroup. Plausibly, poor cigarette consuming households have fewer substitution options – they may already be smoking the least expensive brands – and, therefore, the burden reduction has not been as noticeable. Another possibility is that poor households have made a switch to other forms of less expensive tobacco, rather than cigarettes, which is why the burden reduction is higher when comparing matched tobacco consuming households.

4.2 Quantile Regression

The previous discussion has provided support for the hypothesis that poorer households reduced their tobacco and cigarette burdens more than richer households. That result arises despite the consistent increase in taxes during the period under study. These same results suggest that, as long as cigarette and tobacco taxes make up a constant share of the price of tobacco and cigarettes, tobacco and cigarette taxes have become less regressive over time. Assuming that the tax share is near 50%, the tobacco tax concentration index would be about 1/2 of the index

calculated, yielding an index reduction of 0.0113. A similar assumption would yield a cigarette tax concentration index reduction of 0.0057.

Although the concentration curves and their associated indexes describe a broad change in regressivity that has benefitted poorer households, it is a broad description. With quantile regression, it is possible consider other dimensions of the burden. Table 5 presents estimates from quantile regressions (Koenker and Basset, Jr., 1978). Specifically, we assume that one-half of the tobacco and cigarette burdens are made up of tax. Thus, we take the share in (3) and half it. Thus, we report tax regressivity, β_1 and the change in that regressivity, β_3 . Estimates are presented for low burden (25th percentile), median burden and high burden (75th percentile) households. For tobacco consuming households the tobacco tax burdens are 0.004, 0.011, 0.023 at the 25th, 50th and 75th quantile of the tobacco share distribution, respectively. For cigarette consuming households the cigarette tax burdens are 0.004, 0.01, 0.023 at the 25th, 50th and 75th quantile of the cigarette share distribution.

Table 5 about here.

The quantile regression results confirm the concentration curve and index results, at least qualitatively, as the index and regression results are not directly comparable. Tobacco and cigarette taxes are regressive, and regressivity is better in 2010/11 than it was in 2005/06. Regressivity is relatively better for households facing the largest tobacco and cigarette tax burdens. For tobacco and cigarette consuming households facing the smallest burden, i.e., those in 25th quantile, the burden reduction has been the smallest; recall that positive values represent reductions, since regressivity is a negative value. Furthermore, the reduction at the 75th quantile exceeds that of the 50th, which exceeds that at the 25th quantile. For cigarette consuming households, the reduction at the 50th and 75th percentiles are statistically similar, but are larger than for the reduction at the 25th percentile. Although the improvements in regressivity might be larger for higher tax burden households, regressivity tends also to be larger for those households.

5 Conclusion

This research made use of the 2005/06 and 2010/11 South African IESs to examine the regressivity of cigarette and tobacco expenditure and taxes. The analysis was based on genetic matching across the survey years, so that similar households could be compared over time. Once the data was matched, regressivity was assessed through the concentration of tobacco expenditure (and taxes) amongst the poor, which was underscored by concentration curves and the concentration index. Throughout the analysis, we assumed that tobacco taxes represented a fixed share of household tobacco expenditure, and that share was 50%. In addition to examining regressivity through concentration curves and the concentration index, we estimated quantile regressions of tax regressivity, and the change in tax regressivity across households subject to different levels of tax burden.

In the analysis, we found that the tobacco (cigarette) burden and tobacco (cigarette) tax burden are regressive. We also found that households facing the highest tobacco and tobacco tax burden faced the most regressivity. Most importantly, we found that tobacco expenditure and taxation regressivity has improved between 2005/06 and 2010/11, despite the fact that during that period tobacco and cigarette taxes have risen, annually; the increase was near 30% over the period. In particular, we find statistically significant evidence that tobacco tax increases can reduce tobacco tax regressivity, as proposed by Warner (2000), Chaloupka et al. (2011) and Chaloupka et al. (2012). Counter to Coleman and Remler (2008) and, to a lesser extent, Farrelly et al. (2012), who find, in the US, that tobacco price responsiveness amongst the poor is not large enough to lead to reductions in tax regressivity, we find that tobacco price responsiveness can be large enough.

Since we limited our analysis to matched tobacco and cigarette purchasing households over the time period, such that our data contains only tobacco using households, our findings are not underpinned by cessation. Instead, the effect must be driven by overall reductions in purchases, which could reflect substitution towards cheaper tobacco products, or, even, contraband tobacco products. If it is the former, it is likely that our tobacco tax regressivity reduction is understated; cheaper tobacco products sell for less than the most popular brand, such that the actual tax

should be higher on cheaper products. If it is the latter, our tobacco tax regressivity reduction is understated, since the purchasers of contraband cigarettes would not pay the actual cigarette tax. Furthermore, such behaviour would imply that the public health benefits associated with the reduction in purchases are also not being realized. Although van Walbeek (2014) suggests that contraband cigarettes are not a particular problem in South Africa, increased taxation of tobacco products does raise the legal price, relative to the illegal price, creating opportunities for smuggling.

References

- Berg, G. D. and Kaempfer, W. H. (2001). Cigarette demand and tax policy for race groups in South Africa. *Applied Economics*, 33(9):1167–1173.
- Boshoff, W. H. (2008). Cigarette demand in South Africa over 1996-2006: the role of price, income and health awareness. *South African Journal of Economics*, 76(1):118–131.
- Brieman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.
- Brieman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). *Classification and Regression Trees*. Chapman & Hall, New York.
- Brown, T., Platt, S., and Amos, A. (2015). Equity impact of european individual-level smoking cessation interventions to reduce smoking in adults: a systematic review. *European Journal of Public Health*, 24(4):551–556.
- Bühlman, P. and Yu, B. (2003). Boosting with the l_2 -loss: Regression and classification. *Journal of the American Statistical Association*, 98(462):324–339.
- Cancer Society of South Africa (2013). Fact sheet on tobacco products.

- Centres for Disease Control and Prevention (2007). *Best Practices for Comprehensive Tobacco Control Programs - 2007*. Centres for Disease Control and Prevention, Atlanta, GA.
- Chaloupka, F. J., Straif, K., and Leon, M. E. (2011). Effectiveness of tax and price policies on tobacco control. *Tobacco Control*, 20:235–238.
- Chaloupka, F. J., Yurekli, A., and Fong, G. T. (2012). Tobacco taxes as a tobacco control strategy. *Tobacco Control*, 21:172–180.
- Chelwa, G. and van Walbeek, C. (2014). Assessing the causal impact of tobacco expenditure on household spending patterns in Zambia. Economic Research Southern Africa Working Paper No. 453.
- Cochran, W. G. and Rubin, D. B. (1973). Controlling bias in observational studies: A review. *Sankhya, Series A*, 35(4):417–446.
- Coleman, G. J. and Remler, D. K. (2008). Vertical equity consequences of very high cigarette tax increases: if the poor are the ones smoking, how could cigarette tax increases be progressive? *Journal of Policy Analysis and Management*, 27(2):376–400.
- Connelly, R., Goel, G., and Ram, R. (2009). Demand for cigarettes in the United States: effects of prices in bordering states and contiguity with Mexico and Canada. *Applied Economics*, 41:2255–2260.
- DeCicca, P., Kenkel, D., and Liu, F. (2013a). Who pays cigarette taxes? the impact of consumer price search. *The Review of Economics and Statistics*, 95(2):516–529.
- DeCicca, P., Kenkel, D., and Liu, F. (2013b). Excise tax avoidance: the case of state cigarette taxes. *Journal of Health Economics*, 32(6):1130–1141.

- DeCicca, P. and McLeod, L. (2008). Cigarette taxes and older adult smoking: evidence from recent large tax increases. *Journal of Health Economics*, 27:918–929.
- Diamond, A. and Sekhon, J. S. (2013). Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. *The Review of Economics and Statistics*, 95(3):932–945.
- Farrelly, M. C., Nonnemaker, J. M., and Watson, K. A. (2012). The consequences of high cigarette excise taxes for low-income smokers. *PLoS ONE*, 7(9):e43838.
- Farrelly, M. C., Pechachek, T., Thomas, K., and Nelson, D. (2008). The impact of tobacco control programs on adult smoking. *American Journal of Preventive Medicine*, 98:304–309.
- Franks, P., Jerant, A., Leigh, L., Lee, D., and Lewis, II, S. (2007). Cigarette prices, smoking and the poor: implications of recent trends. *American Journal of Preventive Medicine*, 97(1873-1877).
- Gallet, C. A. and List, J. A. (2003). Cigarette demand: a meta-analysis of elasticities. *Health Economics*, 12:821–835.
- Ground, M. and Koch, S. F. (2008). Hurdle models of alcohol and tobacco expenditure in South African households. *South African Journal of Economics*, 76(1):132–143.
- Hill, S., Clifford, D., Platt, S., and Amos, A. (2013). Impact of tobacco control interventions on socioeconomic inequalities in smoking: A review of the evidence. *PloS One*, 10.
- Hoffer, A., Gvillo, R., Shughart, II, W. F., and Thomas, M. D. (2015). Regressive effects: Causes and consequences of selective consumption taxation. Mercatus Center Working Paper, George Mason University.

- Hu, Y., van Lenthe, F. J., Platt, S., Bosdriesz, J. R., Lahelma, E., Menvielle, G., Regidor, E., Santana, P., de Gelder, R., and Mackenbach, J. P. (2016). The impact of tobacco control policies on smoking among socioeconomic groups in nine European countries, 1990-2007. *Nicotene & Tobacco Research*.
- John, R. M. (2008). Crowding out effect of tobacco expenditure and its implications on household resource allocation in India. *Social Science & Medicine*, 66(6):1356–1367.
- John, R. M., Ross, H., and Blecher, E. (2012). Tobacco expenditure and its implications for household resource allocation in Cambodia. *Tobacco Control*, 21(3):341–346.
- Kakwani, N. C. (1977). Measurement of tax progressivity: an international comparison. *The Economic Journal*, 87(345):71–80.
- Koenker, R. (2016). *quantreg: Quantile Regression*. R package version 5.29.
- Koenker, R. and Basset, Jr., G. (1978). *Quantile Regressions*.
- Lee, B., Lessler, J., and Stuart, E. A. (2010). Improving propensity score weighting using machine learning. *Statistics in Medicine*, 29(3):337–346.
- Maire, K. A., Mattick, R. P., Doran, C. M., and Hall, W. D. (2010). Cigarette tax and public health: what are the implications of financially stressed smokers for the effects of price increase on smoking prevalence. *Addiction*, 106(3):622–630.
- O'Donnell, O., van Doorslaer, E., Wagstaff, A., and Lindelow, M. (2008). *Analyzing health equity using household survey data*. World Bank.
- Parzen, M. I., Wei, L., and Ying, Z. (1994). *Quantile Regressions*.

- Plat, D. (2012). *IC2: Inequality and Concentration Indices and Curves*. R package version 1.0-1.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.
- Sekhon, J. S. (2011). Multivariate and propensity score matching software with automated balance optimization: The matching package for r. *Journal of Statistical Software*, 42(7):1–52.
- Sekhon, J. S. and Mebane, Jr., W. R. (1998). Genetic optimization using derivatives: theory and application to nonlinear models. *Political Analysis*, 7:189–203.
- Siahpush, M., Thrasher, J. T., Yong, H. H., Cummings, K. M., Fong, G. T., Saenz de Miera, B., and Borland, R. (2013). Cigarette prices, cigarette expenditure and smoking-induced deprivation: findings from the international tobacco control mexico survey. *Tobacco Control*, 22(4):223–236.
- Siahpush, M., Wakefield, M. A., Spittal, M. J., Durkin, S. J., and Scollo, M. M. (2009). Taxation reduces social disparities in adult smoking prevalence. *American Journal of Preventive Medicine*, 36(4):285–291.
- Statistics South Africa (2008). Income and Expenditure Survey, 2005/06 [Computer File]. S0160. Pretoria: Statistics South Africa [producer], 2006. Pretoria: South African Data Archive, National Research Foundation [distributor], 2008.

- Statistics South Africa (2013). Income and Expenditure Survey, 2010/11 [Computer File]. SADA0193. Pretoria: Statistics South Africa [producer], 2013. Pretoria: South African Data Archive, National Research Foundation [distributor], 2013.
- Thomas, S., Fayter, D., Misso, K., Petticrew, M., Snowden, A., Whitehead, M., and Worthy, G. (2008). Population tobacco control interventions and their effects on social inequalities in smoking: systematic review. *Tobacco Control*, 17:230–237.
- Tshiswaka-Kashalala, G. and Koch, S. F. (2008). Assessing the crowding-out of effect of tobacco consumption. University of Pretoria, Department of Economics Working Paper.
- van Walbeek, C. (2002a). The distributional impact of tobacco excise increases. *South African Journal of Economics*, 70(3):560–578.
- van Walbeek, C. (2002b). Recent trends in smoking prevalence in South Africa - some evidence from AMPS data. *South African Medical Journal*, 92(6):468–472.
- van Walbeek, C. (2014). Measuring changes in the illicit cigarette market using government revenue data: the example of South Africa. *Tobacco Control*, 23:e69–e74.
- Vellios, N. and van Walbeek, C. (2016). Determinants of regular smoking onset in South Africa using duration analysis. *BMJ Open*, 6(7):e011076.
- Warner, K. E. (2000). The economics of tobacco: myths and realities. *Tobacco Control*, 9:78–89.

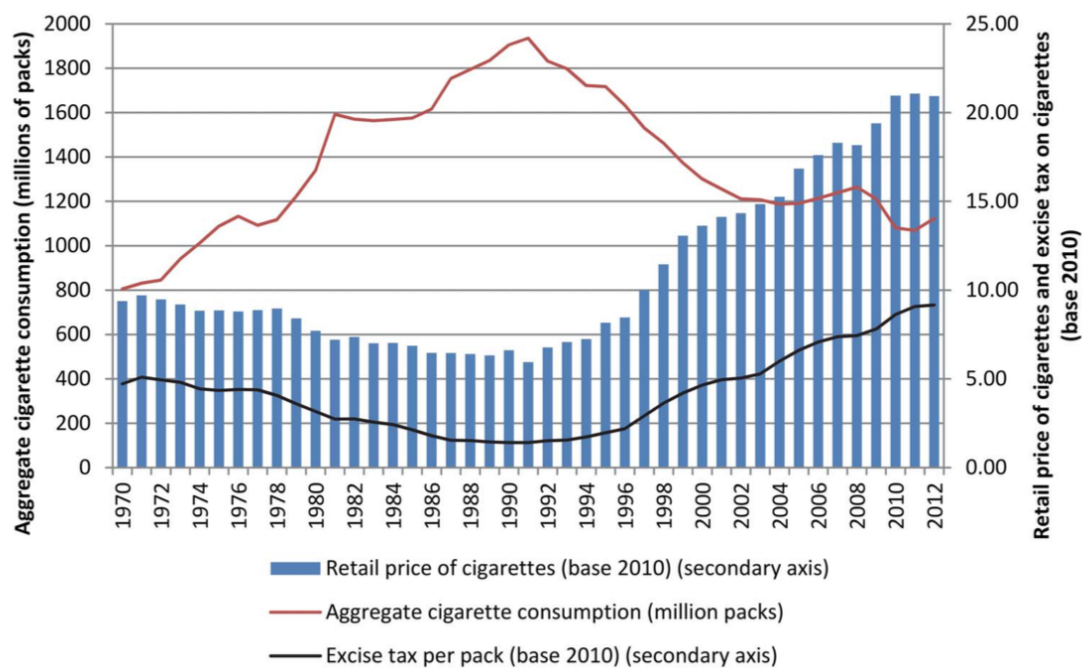


Figure 1: Cigarette Prices and Excise Taxes in Real 2010 South African Rand (ZAR) and Cigarette Consumption in Millions of Packs. Source: Vellios and van Walbeek (2016).

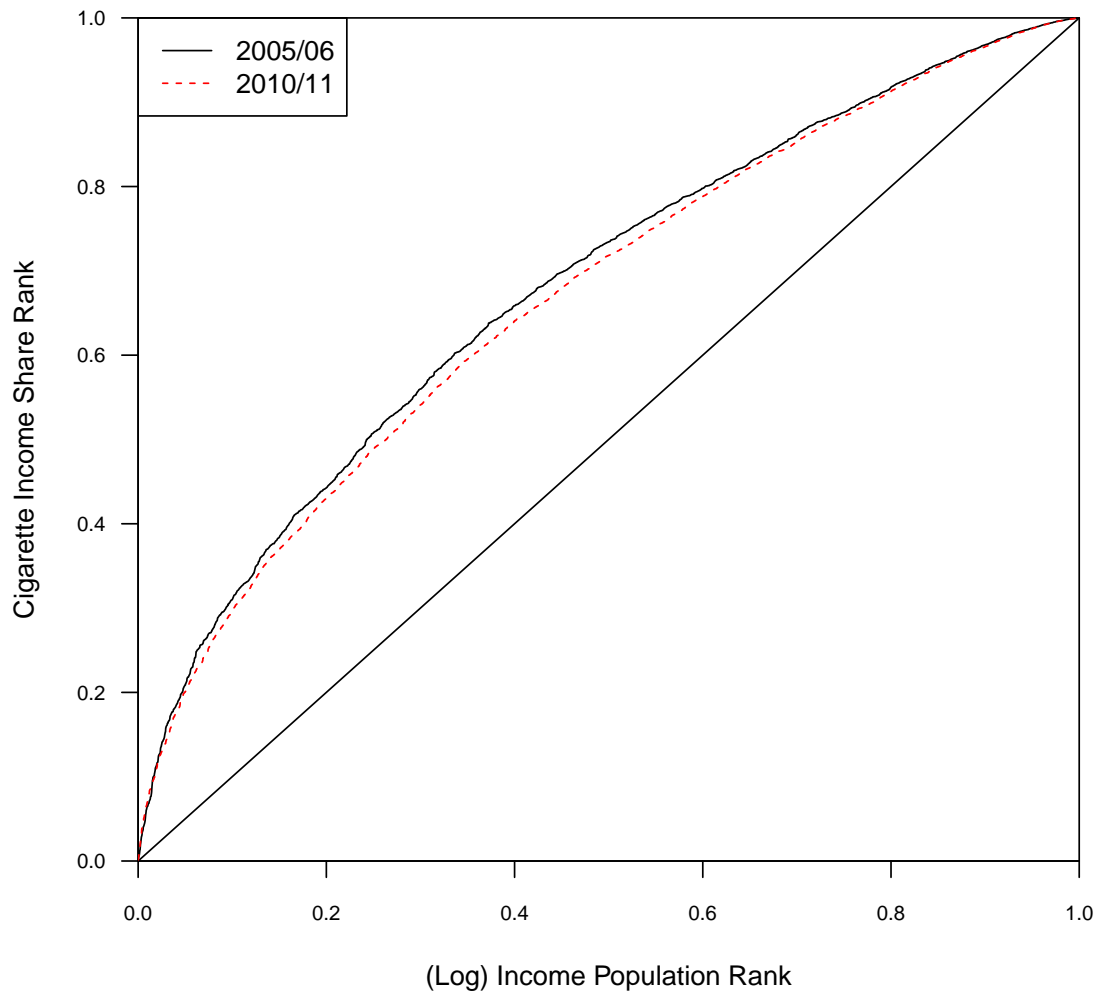


Figure 2: Tobacco consuming household concentration curves for tobacco expenditure as a share of income. Estimated indexes of concentration (see Kakwani, 1977) associated with the concentration curves are available in Table 4.

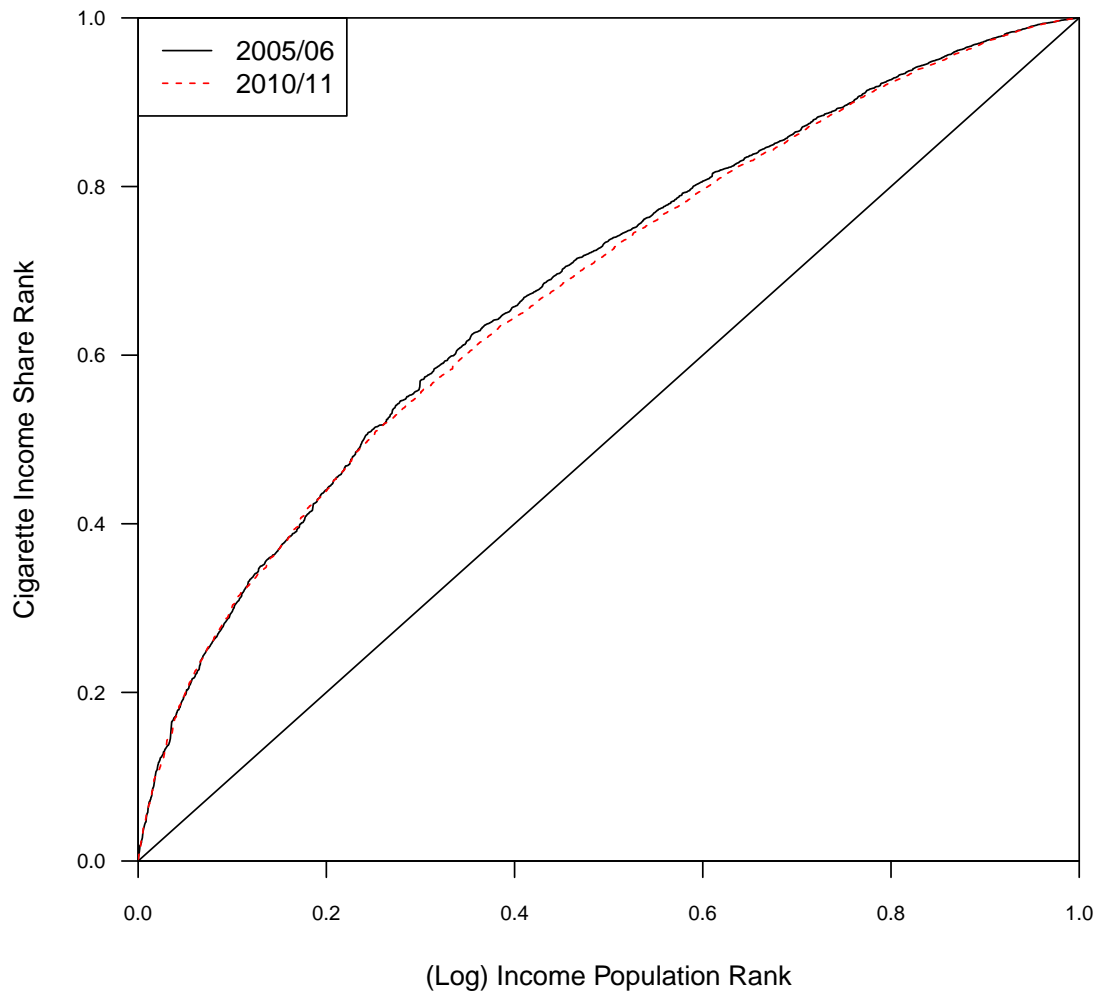


Figure 3: Cigarette consuming household concentration curves for cigarette expenditure as a share of income. Estimated indexes of concentration (see Kakwani, 1977) associated with the concentration curves are available in Table 4.

Table 1: Tobacco Consuming Household Descriptive Statistics After Matching

	Matched Sample Means		Statistical Tests	
	2005/06	2010/11	Pr(> t)	Pr(> ks)
Propensity Score	0.590	0.591	0.626	0.823
HH Head Age Group	10.472	10.435	0.462	0.000
HH Head Schooling	1.365	1.358	0.660	0.010
Real Income	4861.899	4932.619	0.603	0.309
Real Expenditure	3853.554	3899.656	0.651	0.006
HH Male Ratio	0.545	0.540	0.202	0.123
HH Adult Ratio	0.787	0.782	0.151	0.229
Girls (0-4) in HH	0.200	0.197	0.667	0.136
Boys (0-4) in HH	0.175	0.194	0.008	0.004
Girls (5-14) in HH	0.403	0.417	0.220	0.262
Boys (5-14) in HH	0.386	0.427	0.000	0.005
Men (15-64) in HH	1.259	1.295	0.019	0.075
Women (15-64) in HH	1.126	1.183	0.001	0.066
Men (65+) in HH	0.143	0.147	0.441	0.375
Women (65+) in HH	0.202	0.211	0.159	0.060
Eastern Cape	0.164	0.142	0.000	
Western Cape	0.140	0.141	0.765	
Northern Cape	0.088	0.118	0.000	
Free State	0.152	0.113	0.000	
Kwa-Zulu Natal	0.097	0.150	0.000	
Northwest Province	0.117	0.086	0.000	
Gauteng Province	0.106	0.115	0.055	
Mpumulanga Province	0.079	0.068	0.005	
Observations	7806	7806		

Means for 2005/06 and 2010/11 Income and Expenditure Surveys along with t -test p -values of mean differences and ks-test p -values for distributional differences. Note: for discrete variables, no KS-test has been performed, which is why those cells are blank.

Table 2: Cigarette Consuming Household Descriptive Statistics After Matching

	Matched Sample Means		Statistical Tests	
	2005/06	2010/11	Pr(> t)	Pr(> ks)
Propensity Score	0.556	0.557	0.570	0.592
HH Head Age Group	9.881	9.819	0.309	0.001
HH Head Schooling	1.648	1.659	0.582	0.492
Real Income	6396.054	6463.292	0.741	0.685
Real Expenditure	5014.314	5052.121	0.806	0.015
HH Male Ratio	0.587	0.583	0.494	0.698
HH Adult Ratio	0.802	0.801	0.808	0.030
Girls (0-4) in HH	0.174	0.175	0.963	0.627
Boys (0-4) in HH	0.172	0.178	0.507	0.314
Girls (5-14) in HH	0.348	0.375	0.044	0.259
Boys (5-14) in HH	0.361	0.386	0.072	0.349
Men (15-64) in HH	1.371	1.406	0.067	0.148
Women (15-64) in HH	1.093	1.150	0.006	0.078
Men (65+) in HH	0.116	0.120	0.502	0.479
Women (65+) in HH	0.154	0.159	0.497	0.617
Eastern Cape	0.209	0.179	0.000	
Western Cape	0.116	0.125	0.145	
Northern Cape	0.072	0.100	0.000	
Free State	0.123	0.096	0.000	
Kwa-Zulu Natal	0.108	0.141	0.000	
Northwest Province	0.089	0.075	0.011	
Gauteng Province	0.147	0.153	0.445	
Mpumulanga Province	0.083	0.072	0.059	
Observations	4909	4909		

Means for 2005/06 and 2010/11 Income and Expenditure Surveys along with t -test p -values of mean differences and ks-test p -values for distributional differences. Note: for discrete variables, no KS-test has been performed, which is why those cells are blank.

Table 3: Cigarette and Tobacco Usage Differences Across Surveys

(a) Tobacco Purchasing Households

	Matched Sample Means		Mean	Wilcoxon
	2005/06	2010/11	Difference	P Value
Annual Cigarette Expenditure	72.1942	50.6912	-21.5030	0.000
Cigarette Income Share	0.0286	0.0215	-0.0071	0.000
Cigarette Expenditure Share	0.0258	0.0206	-0.0052	0.000
Annual Tobacco Expenditure	88.6198	63.0568	-25.5630	0.000
Tobacco Income Share	0.0447	0.0340	-0.0107	0.000
Tobacco Expenditure Share	0.0404	0.0331	-0.0073	0.000

(b) Cigarette Purchasing Households

	Matched Sample Means		Mean	Wilcoxon
	2005/06	2010/11	Difference	P Value
Annual Cigarette Expenditure	116.2345	80.8052	-35.4293	0.000
Cigarette Income Share	0.0458	0.0343	-0.0114	0.000
Cigarette Expenditure Share	0.0407	0.0329	-0.0078	0.000
Annual Tobacco Expenditure	121.7212	86.0263	-35.6949	0.000
Tobacco Income Share	0.0497	0.0389	-0.0107	0.000
Tobacco Expenditure Share	0.0446	0.0373	-0.0073	0.000

Wilcoxon paired tests of mean differences (across the matched samples). Panel (3a) includes only tobacco purchasing households, ($n = 15612$). Panel (3b) includes only cigarette purchasing households, ($n = 9818$). Note: Shares listed as percentages rather than proportions.

Table 4: Concentration of Cigarette Purchase Burden

	Concentration Index		Index
	2005/06	2010/11	Difference
Matched Tobacco Consuming Households	-0.3617	-0.3392	0.0225
Matched Cigarette Consuming Households	-0.3633	-0.3520	0.0113

Estimates of the Kakwani's (1977) concentration index for matched households. Note: "Difference" is the 2010/11 index net of the 2005/06 index; thus, a positive difference represents a reduction in the regressivity of cigarette purchases.

Table 5: Quantile Regression Estimates of Tobacco and Cigarette Burden Regressivity Across Surveys

(a) Tobacco Burden in Tobacco Purchasing Households

	Estimate	Std. Error	t -value	Pr(> t)
25 th Percentile: Regressivity	-0.0027	0.0001	-21.3318	0.0000
25 th Change	0.0008	0.0002	5.3202	0.0000
50 th Percentile: Regressivity	-0.0056	0.0002	-26.3207	0.0000
50 th Change	0.0020	0.0002	8.5471	0.0000
75 th Percentile: Regressivity	-0.0100	0.0002	-43.4753	0.0000
75 th Change	0.0030	0.0003	9.7847	0.0000

(b) Cigarette Burden in Cigarette Purchasing Households

	Estimate	Std. Error	t -value	Pr(> t)
25 th Percentile: Regressivity	-0.0031	0.0002	-19.0930	0.0000
25 th Change	0.0014	0.0002	7.3954	0.0000
50 th Percentile: Regressivity	-0.0059	0.0002	-24.1954	0.0000
50 th Change	0.0023	0.0003	7.7359	0.0000
75 th Percentile: Regressivity	-0.0096	0.0003	-31.7388	0.0000
75 th Change	0.0021	0.0004	5.8695	0.0000

Quantile regression estimates of tobacco (cigarette) tax regressivity and the change in that regressivity between 2005/06 and 2010/11 using matched data; see (3) for details. Panel (5a) only tobacco purchasing households, ($n = 15612$). Panel (5b) includes only cigarette purchasing households, ($n = 9818$). Note: Negative regressivity estimates connote regressivity of the tax, while positive change estimates connote reduction in regressivity. Standard errors are bootstrapped following Parzen et al. (1994); 399 replications were used.

A Descriptive Illustrations of Matched Data

As further evidence of the validity of the matching mechanism, we illustrate post-match income densities for the same sets of households. For tobacco purchasing households, the post-match income densities (Panel A.1) point to similar distributions across the surveys. The same conclusion can be drawn from the income density for post-match cigarette purchasing households (Panel A.2), which gives us further confidence in the results we present in this research.

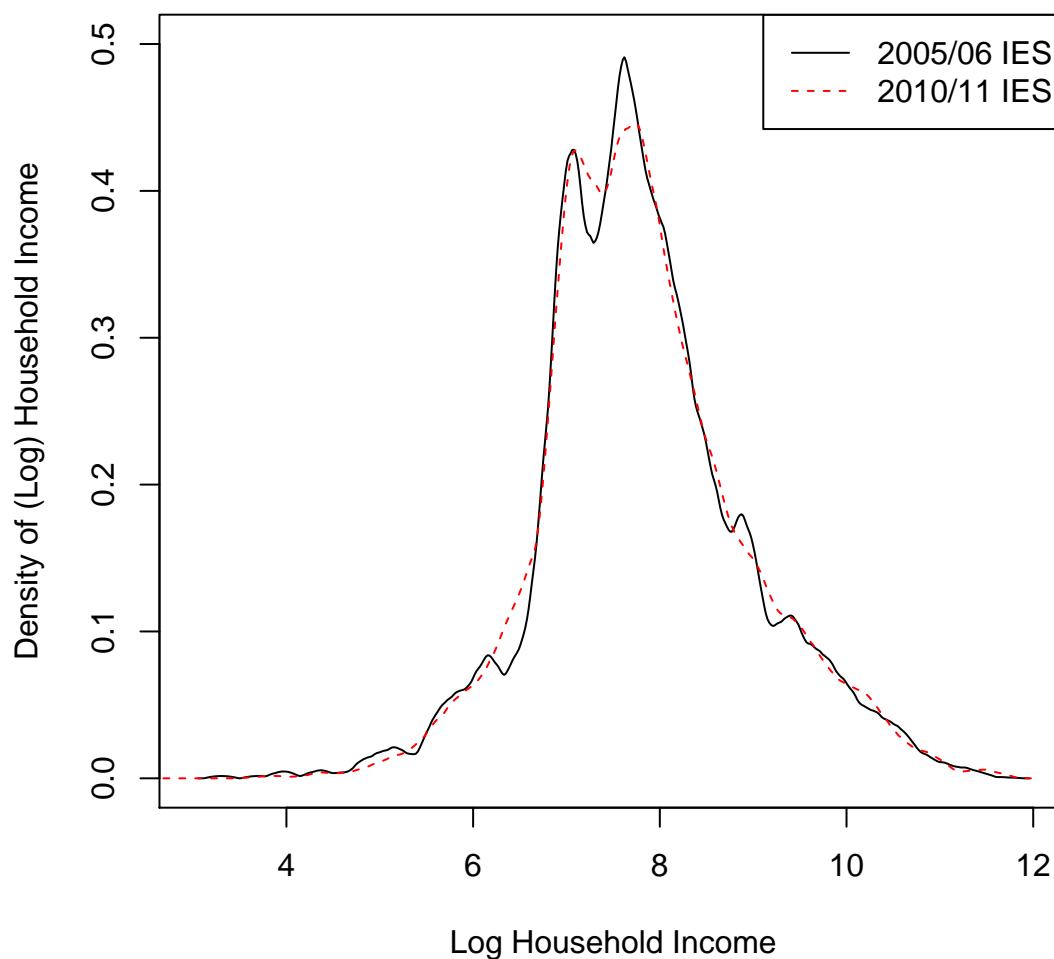


Figure A.1: Densities of the log of real household income for tobacco consuming households, following matching.

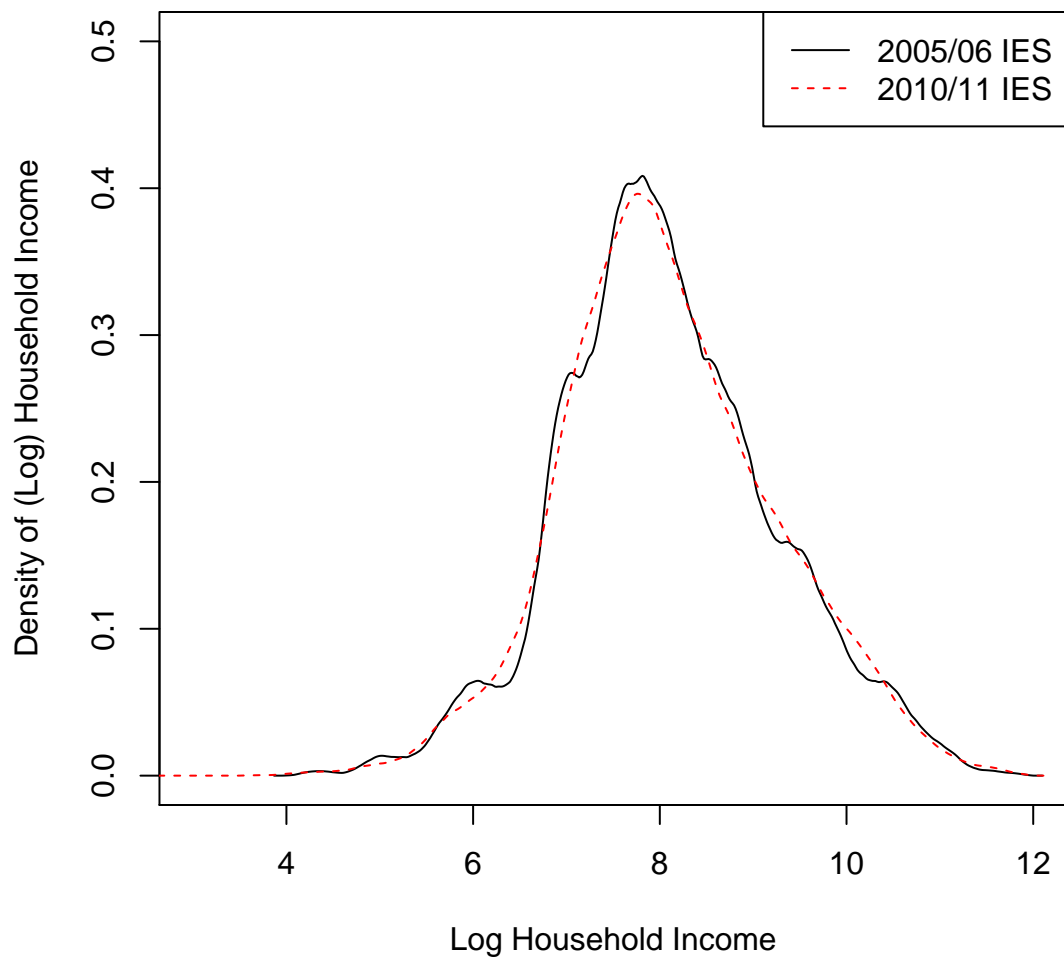


Figure A.2: Densities of the log of real household income for cigarette consuming households, following matching.