Appendix: An overlooked market: loose cigarettes, informal vendors, and their implications for tobacco taxation

Appendix A: Data and Method

This paper relies on two main sources of data: the African Cigarettes Prices Project from the Economics of Excisable Products Research Unit (REEP) and data from all low- and middle-income countries covered by the International Tobacco Control (ITC) Policy Evaluation Project which included information on loose cigarette prices. The only middle-income country which was not included from those for which ITC had data is Vietnam, as we could not verify the availability of data on loose cigarettes. To the best of our knowledge, these are the only two sources of data on tobacco consumption or pricing which also cover loose cigarettes and that include a time component. That is, some information on loose cigarette consumption and pricing is also available through the Global Adult Tobacco Surveys managed by the Centres for Disease Control and Prevention (CDC), but in the vast majority of cases this survey was only implemented once in each country. On the other hand, both of the data sources we use ensure coverage of each country in at least 2 years.

The Data on Aliments, Tobacco and Alcohol in Africa Project based at REEP at the University of Cape Town has been collecting prices of cigarettes sold at retail outlets and from street vendors in a number of African countries between 2016 and 2022, with data available through the DataFrist of the University of Cape Town at https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/927 . This survey represents the most extensive source of data on cigarette prices in the African continent, but, importantly, it is not nationally representative. The data is collected by students of REEP after rigorous training, but the sampling is purposive in nature, as it is up to the students to determine from which shops information is collected. While efforts are made to avoid duplicate in data collection where more students are working in the same area, there might be some geographic bias in coverage (e.g. less focus on poorer areas in which data collection might be more complex), and the way in which geographic coverage is determined is not well documented. Consequently, while certain shops are covered in more than one round of the survey, the data remains cross-sectional in nature rather than representing a true panel. While 10 rounds of the survey exist, we draw on the publicly available data here, which includes 2020 (for a full breakdown of the number of rounds used by country see Appendix B). Critically, its information about cigarette prices differentiates between cigarettes sold as loose cigarettes and cigarettes sold as packs. It also includes information about the brands as well as about features of the store, such as whether they are street vendors, kiosks, or retail outlets. The data is best described as a repeated cross-section rather than a panel, and does not present a nationally representative sample of retail outlets as selection was purposive rather than based on an underlying sampling frame. As its authors note, this may have caused some geographic bias, for example against poorer areas that are more difficult for enumerators to access. Nonetheless, the data is notable and useful for our purposes not merely for its specific treatment of loose cigarettes, but also because of its large number of price observations (for example, it includes 50,219 price observations over 5 years in South Africa).

Aside from this, we also draw on data from ITC that is based on a set of multi-country surveys which can allow researchers to compare the success of different tobacco control policies promoted by the WHO. Multiple waves of nationally representative surveys have been implemented in each of the 28 countries covered by the projects – most of which are high-income countries – targeting both people

Gallien M, *et al. Tob Control* 2023;0:1–7. doi: 10.1136/tc-2023-057965

who do and do not smoke and people who use other tobacco products such as chewing tobacco. The survey includes different types of information on both the respondents – such as gender, age, income and level of education – and, if any, their tobacco use – such as where did they last buy tobacco, in which form, of which brand and for which price. Furthermore, the data is collected in a panel format, although not all respondents could be tracked for each wave, so that it is possible to track changes in tobacco consumption over time – such as a switch between consuming loose and packed cigarettes. Due to attrition, most of our analysis uses this data as a repeated cross-section rather than a panel. As mentioned in the opening paragraph of the section, we initially targeted all ITC data from low and middle-income countries in which we could find evidence of loose cigarette consumption – these were Bangladesh, India, Kenya, Malaysia, Thailand and Zambia. While Vietnam could have also qualified, we did not pursue that data as the survey instrument is only available in Vietnamese on ITC website, so that we could not determine whether information about loose cigarettes consumption and pricing was available. Furthermore, after an initial examination of the data, we decided to exclude Malaysia from the analysis, as loose cigarettes accounted for less than 5% of consumption in five of the six survey waves. Summary statistics for both datasets are provided below.

There are two consequences of the limited availability of data on the sale and consumption of loose cigarettes. The first is that, although we believe that the phenomena we are trying to illustrate occur in a number of different low- and middle-income countries across the globe, the vast majority of those covered in the remainder of the paper are located in Sub-Saharan Africa, as those are covered in the REEP data.

A further consequence is that we mostly limit our analysis to exploratory statistics, such as testing for correlation or significant differences amongst variables, although we also perform a couple of multivariate regressions in cases in which enough data is available, which is mainly the case for the characteristics associated with smoking sticks (Table 2 in the paper). We decided to run these regressions separately for two main reasons. The first is that not enough information is generally available on this topic to assume that the characteristics associated with the decision to buy loose cigarettes will be the same across the 5 countries for which we have data, nor that they will have the same impact (i.e. in certain countries women who smoke might be more likely to buy singles, in others less). The second reason is that some of the data contained in the survey – education and income – might be categorised differently across each country survey, and hence require thorough harmonisation if we were to pool all countries together. Nonetheless, we have attempted some harmonisation of both of these variables to allows for some comparability, which we now describe. With regard to education, each country survey contained detailed information of what was the highest educational attainment of the respondents. While this included many categories, the diverse structure of schools' progression across different countries led us to decide to create three categories for which we could construct with relative certainty in each country case: "illiterate" (which we use as baseline for regression), "primary or lower secondary" and "higher secondary or tertiary". While this categorisation necessarily leads to some loss of information, this was necessary to ensure some comparability. With regard to income, each country survey contained an income variable with three categories which were however not labelled, nor corresponding to a clear variable in the survey instrument. In this case, we determined whether this variable was increasing from "low" to "high" income or decreasing from "high" to "low" income through the relative distribution of occupations. Using Thailand as an example, we determined that the category 1 was associated with "low income" and 3 with "high income" as professionals made up 0.06% and 3.59% of the two group respectively; those employed in the service sector 7.17% and 21.99% respectively; those employed in agriculture 50.81% and 20.69% respectively, etc. As the paper demonstrates, this level of analysis is sufficient to make our general case, but it also highlights the importance of further data collection in this area.

Appendix B: Summary Statistics

Table C.1 – Summary statistic of REEP data

Country	Observation	Rounds	Years	Main brand	Shop Type		
					Retail	Spaza/Kiosk	Street Vendor
Botswana	9,934	8	2016-2020	Peter Stuyvesant	83.93%	2.18%	13.88%
Chad	250	1	2019-2020	Fine	4.40%	9.20%	86.40%
Eswatini	723	3	2016-2020	Dunhill	96.82%	1.38%	1.80%
Ethiopia	5,989	3	2018-2020	Nyala	-	4.51%	95.49%
Ghana	485	1	2017-2018	Pall Mall	16.29%	61.65%	22.06%
Kenya	349	1	2018-2019	Sportsman	12.89%	64.18%	22.92%
Lesotho	25,653	9	2016-2020	Dunhill	9.69%	15.52%	74.78%
Madagascar	367	2	2019-2020	Good Look	24.52%	38.15%	37.33%
Malawi	2,334	4	2017-2020	Pall Mall	27.08%	18.34%	54.58%
Mauritius	232	1	2016	Dunhill	100.00%	-	-
Mozambique	850	2	2019-2020	Pall Mall	19.18%	4.47%	76.35%
Namibia	25,919	9	2016-2020	Camel	96.83%	1.32%	1.85%
Nigeria	740	2	2018-2020	Benson & Hedges	11.22%	73.51%	15.27%
South Africa	45,410	8	2016-2020	Peter Stuyvesant	54.34%	16.28%	29.38%
Tanzania	2,137	3	2017-2020	Embassy	28.69%	9.45%	61.86%
Uganda	128	2	2016-2019	Dunhill	29.69%	57.03%	13.28%
Zambia	556	5	2017-2019	Peter Stuyvesant	49.10%	43.17%	7.73%
Zimbabwe	22,046	8	2016-2020	Pacific	27.24%	15.40%	57.36%

Table C.2 – ITC Summary statistics

Country	Observation	Waves	Years	Consumption	
				Packed	Loose
Bangladesh	8,243	4	2009-2015	33.01%	66.99%
India	3,052	3	2007-2013	23.97%	76.03%
Kenya	1,776	2	2012-2018	12.70%	87.30%
Thailand	8,146	6	2005-2014	63.12%	36.88%
Zambia	813	2	2012-2014	17.90%	82.10%

Table C.3 – Tests of differences across individual variables and for multicollinearity, Bangladesh

	Bangladesh				
	Number	Gender	Income	Education	Age
People who smoke packs	2,646	0.02	2.35*	1.97**	38.88***
People who smoke loose cigarettes	5,777	0.02	2.33*	2.01**	36.37***

	VIF	SQRT VIF	Tolerance	R-Squared
sex	1.00	1.00	1.00	0.00
income	1.08	1.04	0.92	0.08
education	1.11	1.05	0.90	0.10
age_at_recruitment	1.02	1.01	0.98	0.02

Table C.4 – Tests of differences across individual variables and for multicollinearity, India

India					
	Number	Gender	Income	Education	Age
People who smoke packs	727	0.01	2.02***	2.60***	40.65***
People who smoke loose cigarettes	2,325	0.01	1.84***	2.38***	37.82***
	VIF	SQRT VIF	Tolerance	R-Squared	
sex	1.00	1.00	1.00	0.00	
income	1.04	1.02	0.97	0.04	
education	1.04	1.02	0.96	0.04	
age_at_recruitment	1.01	1.01	0.99	0.01	

Table C.5 – Tests of differences across individual variables and for multicollinearity, Kenya

Kenya					
	Number	Gender	Income	Education	Age
People who smoke packs	219	0.07	1.73***	2.45*	40.42
People who smoke loose cigarettes	1,557	0.08	1.42***	2.38*	40.82
	VIF	SQRT VIF	Tolerance	R-Squared	
sex	1.02	1.01	0.98	0.02	
income	1.03	1.01	0.97	0.03	
education	1.08	1.04	0.92	0.08	
age_at_recruitment	1.06	1.03	0.94	0.06	

Table C.6 – Tests of differences across individual variables and for multicollinearity, Thailand

Thailand					
	Number	Gender	Income	Education	Age
People who smoke packs	3,957	0.08***	2.36***	2.08***	41.07***
People who smoke loose cigarettes	1,534	0.13***	2.21***	2.01***	38.7***
	VIF	SQRT VIF	Tolerance	R-Squared	

sex	1.01	1.01	0.99	0.01
income	1.07	1.03	0.94	0.07
education	1.09	1.04	0.92	0.08
age_at_recruitment	1.09	1.04	0.92	0.08

Table C.7 – Tests of differences across individual variables and for multicollinearity, Zambia

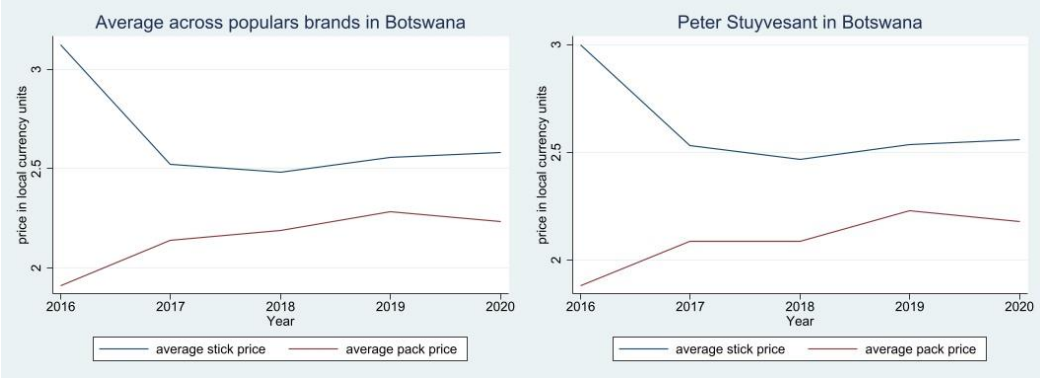
Zambia					
	Number	Gender	Income	Education	Age
People who smoke packs	144	0.04	2.07	2.48	36.49***
People who smoke loose cigarettes	669	0.04	2.17	2.55	32.7***
	VIF	SQRT VIF	Tolerance	R-Squared	
sex	1.01	1.00	0.99	0.01	
income	1.12	1.06	0.89	0.11	
education	1.13	1.06	0.88	0.12	
age_at_recruitment	1.01	1.01	0.99	0.01	

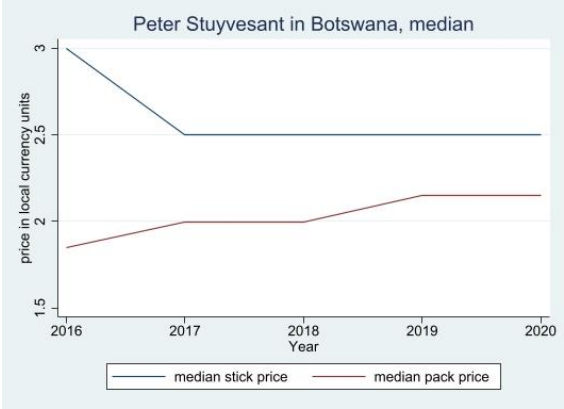
Appendix C: Price Data for all countries in the REEP dataset

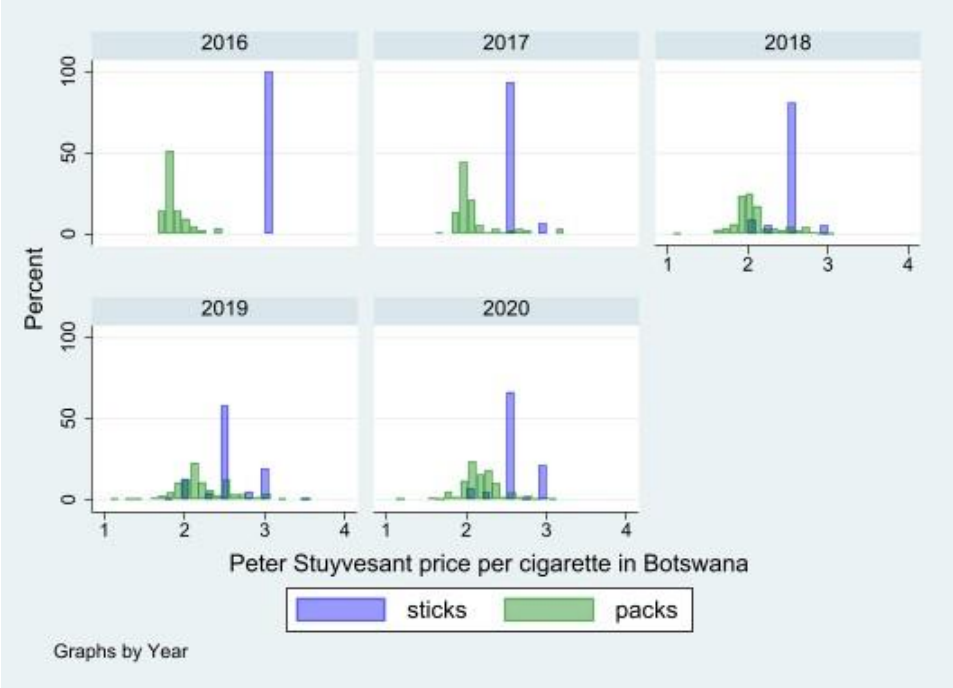
This appendix illustrates, for each country in the REEP dataset for which a sample size of over 2,000 observations is available (excluding Malawi, for which stick prices represent less than 10% of the observations for the main brand):

1) The trends in nominal prices (per cigarette) of cigarettes sold as sticks and packs, averaged over the price of every brand that makes up at least 2% of the total prices recorded.
2) the trends in prices (per cigarette) of cigarettes sold as sticks and packs of the dominant brand in the country that is sold as both stick and packs (dominant brand defined as the brand that has the highest peak market share at any year for which data is available), both averaged across all price points and using only the median price.
3) the distribution/histogram of the prices (per cigarette) of cigarettes sold as sticks and packs of the dominant brand over time.
4) The average mark-up applied to loose cigarettes, expressed as share of the price of a packed one, as well as its standard deviation and the share of brand-store pairs for which the mark-up is negative (i.e., selling sticks at a lower price than packed cigarettes).

Botswana

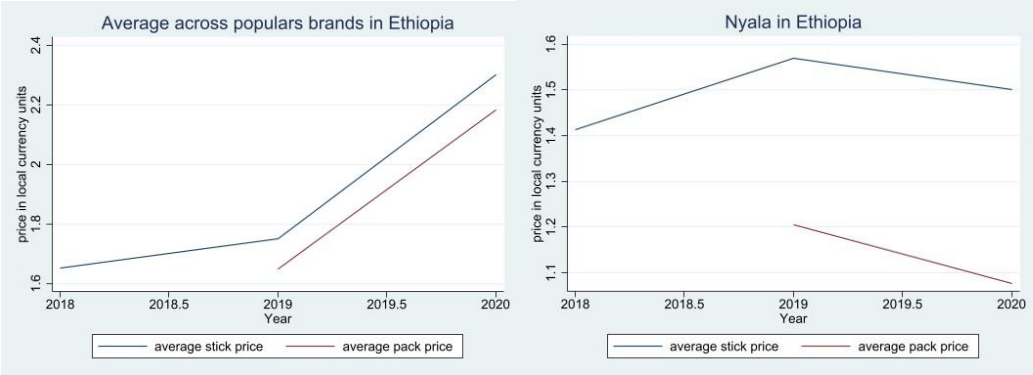


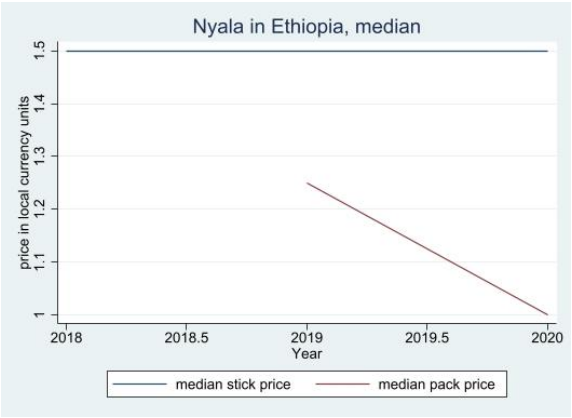


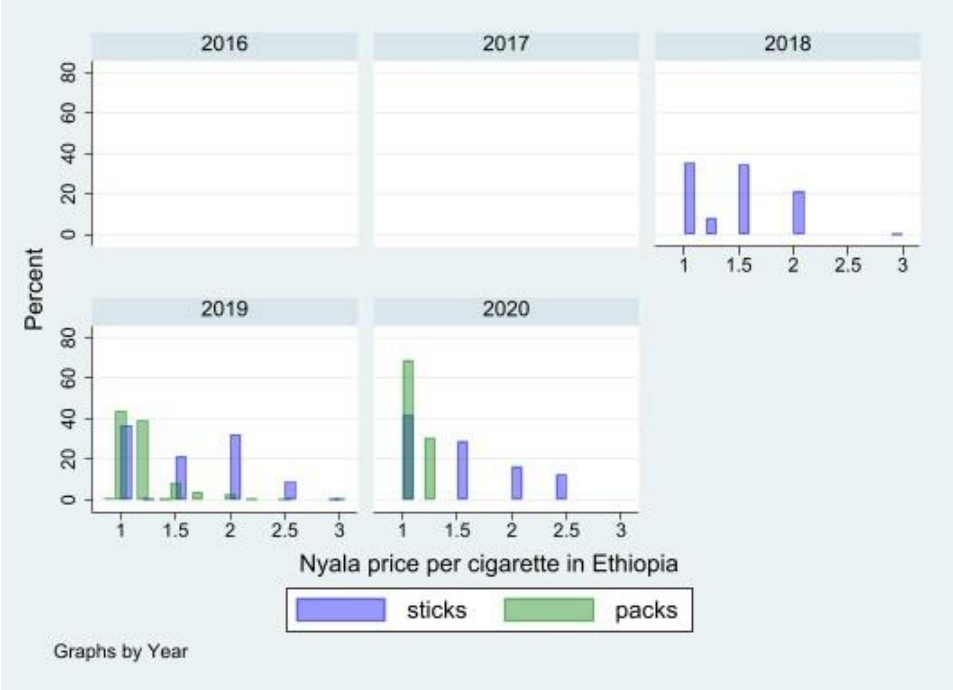


Mark up on loose as a share of packed cigarette price	Average	Standard Deviation	Negative Share
	6.5%	14.2%	0.0%

Ethiopia

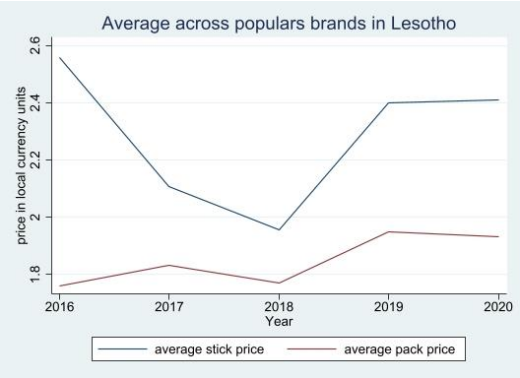

Average across populars brands in Ethiopia
price in local currency units
1.6
1.8
2
2.2
2.4
2018
2018.5
2019
2019.5
2020
Year
average stick price
average pack price
Nyala in Ethiopia
price in local currency units
1.1
1.2
1.3
1.4
1.5
1.6
2018
2018.5
2019
2019.5
2020
Year
average stick price
average pack price

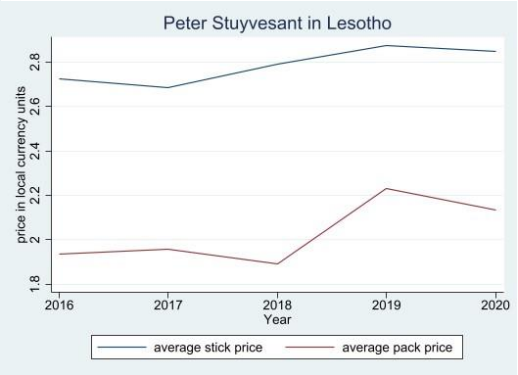

Nyala in Ethiopia, median
price in local currency units
1
1.1
1.2
1.3
1.4
1.5
2018
2018.5
2019
2019.5
2020
Year
median stick price
median pack price

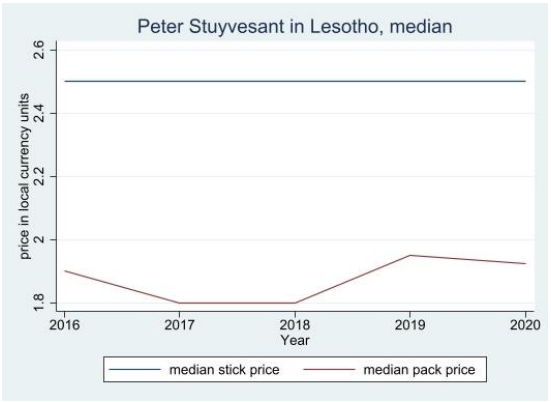

2016
2017
2018
2019
2020
Percent
0
20
40
60
80
1
1.5
2
2.5
3
Nyala price per cigarette in Ethiopia
sticks
packs
Graphs by Year

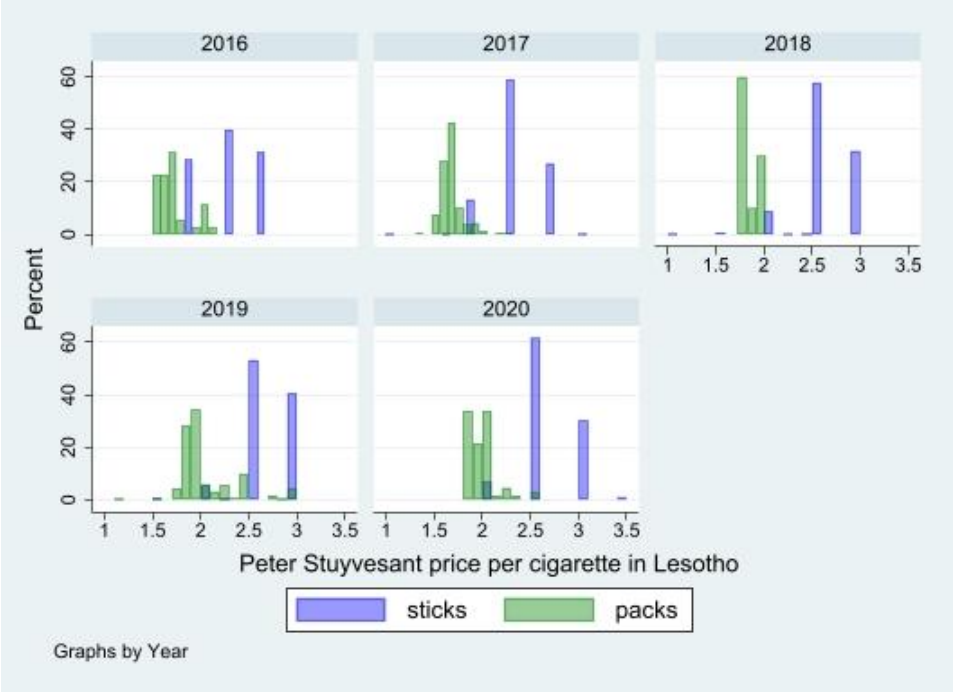
Mark up on loose as a share of packed cigarette price	Average	Standard Deviation	Negative Share
	18.3%	36.6%	10.8%

Lesotho



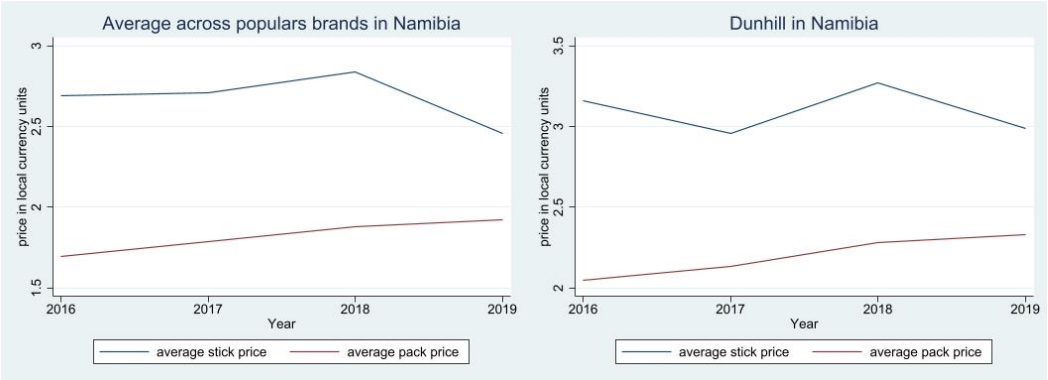


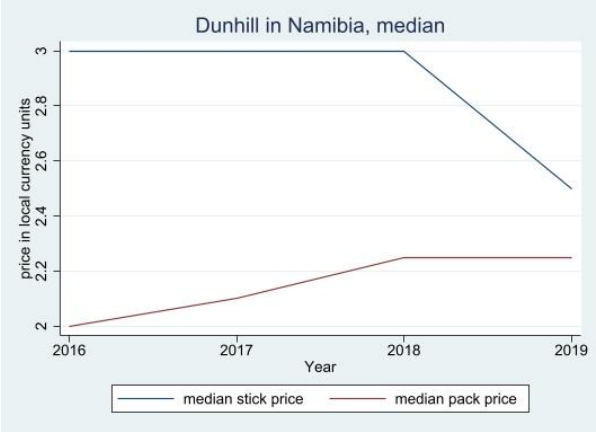


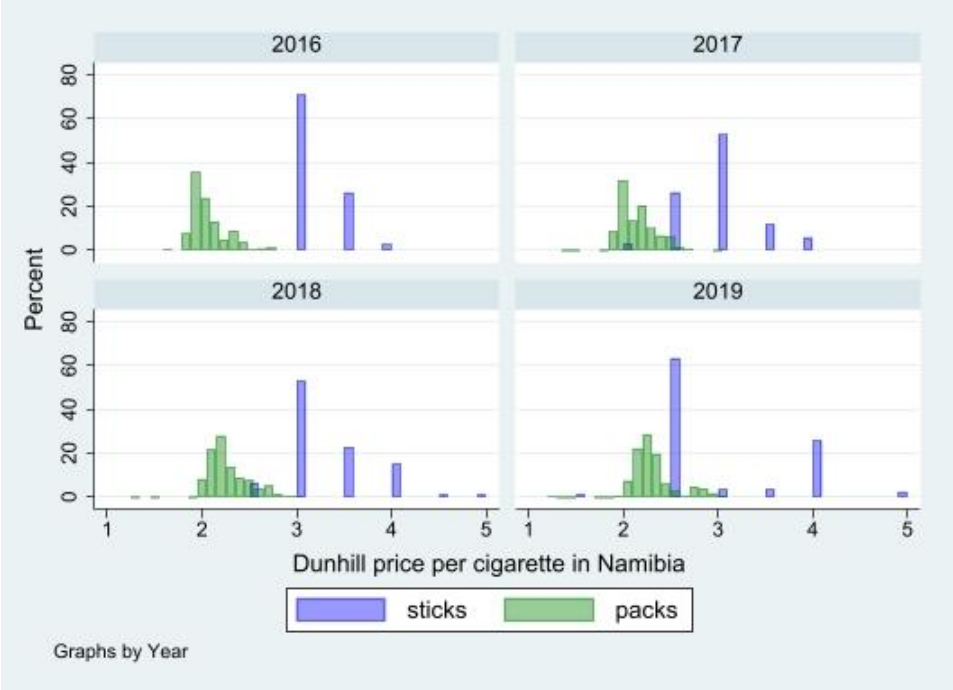


Mark up on loose as a share of packed cigarette price	Average	Standard Deviation	Negative Share
	27.4%	22.7%	1.0%

Namibia

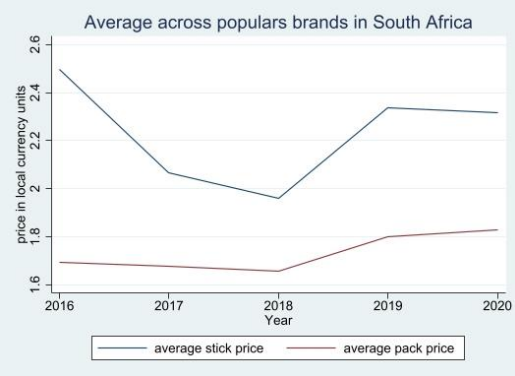


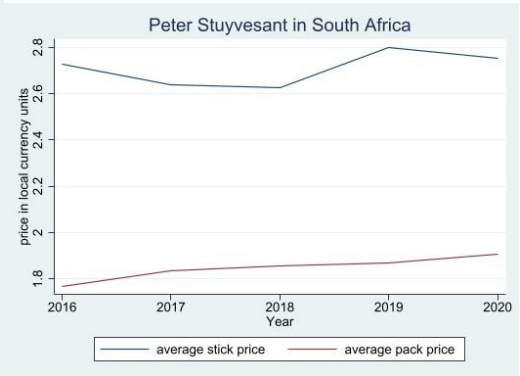


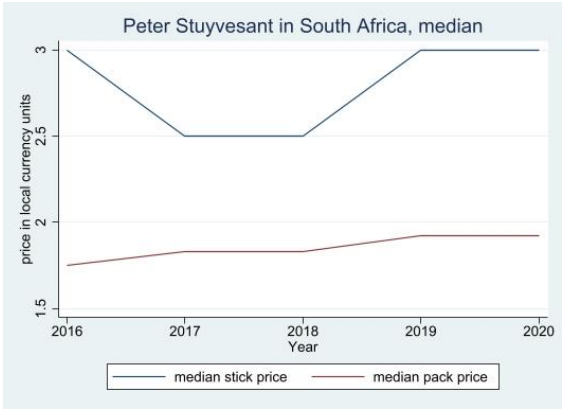


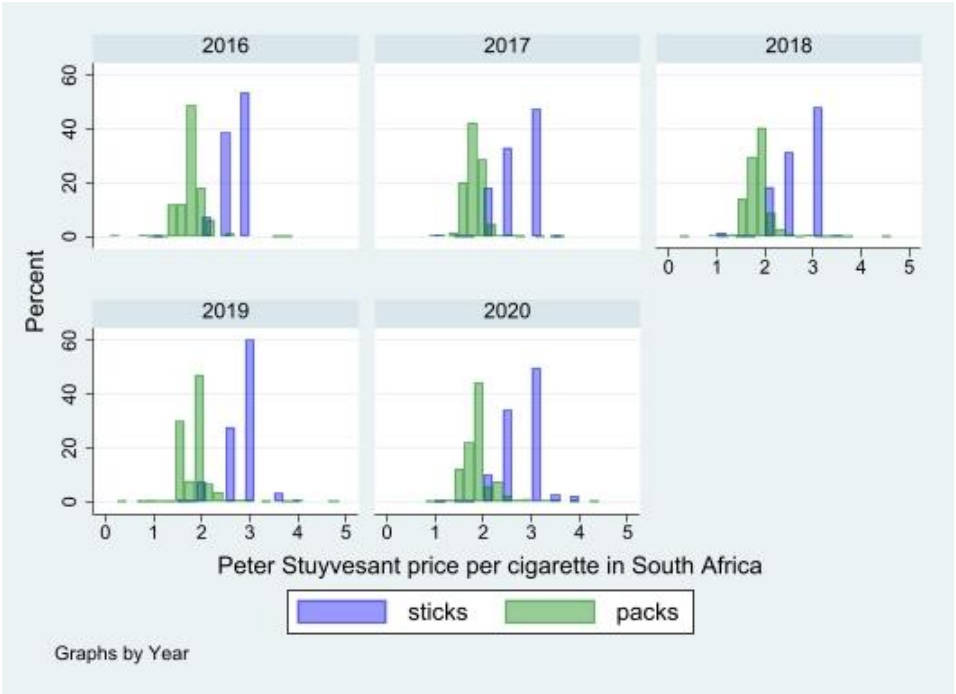
Mark up on loose as a share of packed cigarette price	Average	Standard Deviation	Negative Share
	54.3%	46.9%	0.0%

South Africa



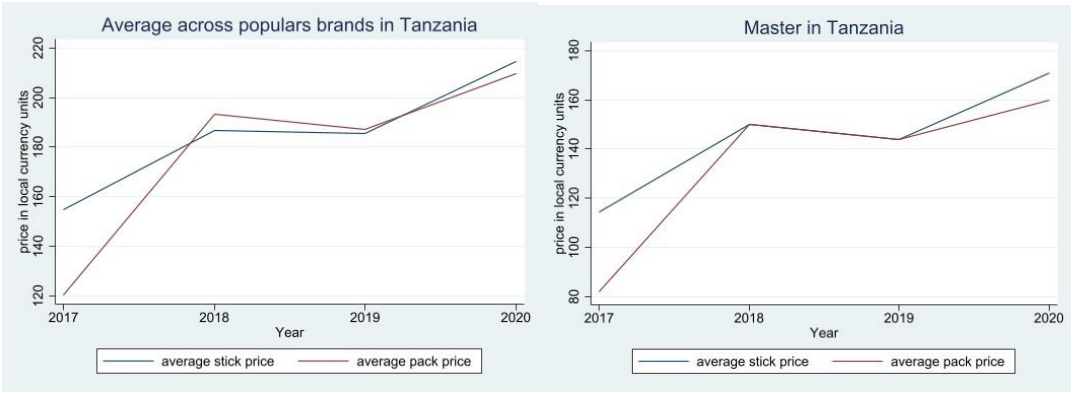




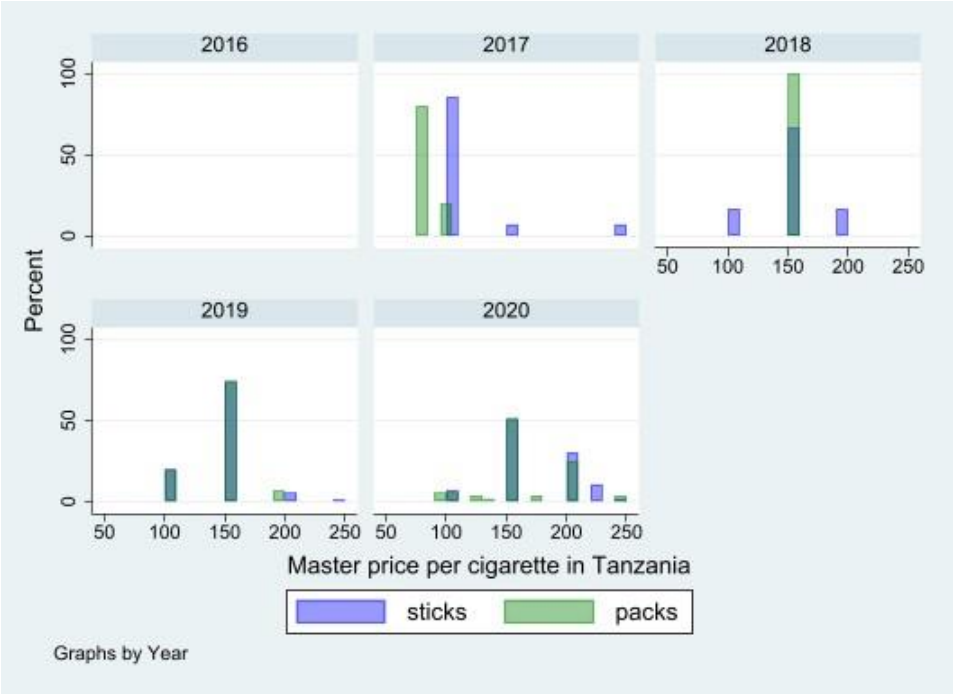


Mark up on loose as a share of packed cigarette price	Average	Standard Deviation	Negative Share
	28.6%	27.0%	3.0%

Tanzania

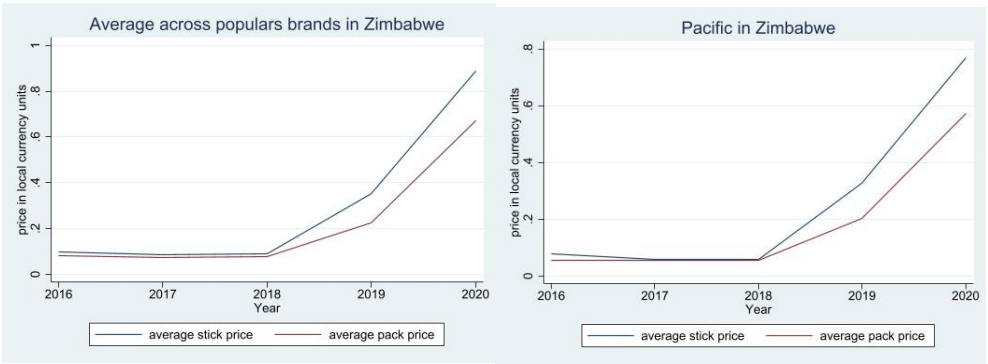

Average across populars brands in Tanzania
price in local currency units
120
140
160
180
200
220
2017
2018
2019
2020
Year
average stick price
average pack price
Master in Tanzania
price in local currency units
80
100
120
140
160
180
2017
2018
2019
2020
Year
average stick price
average pack price

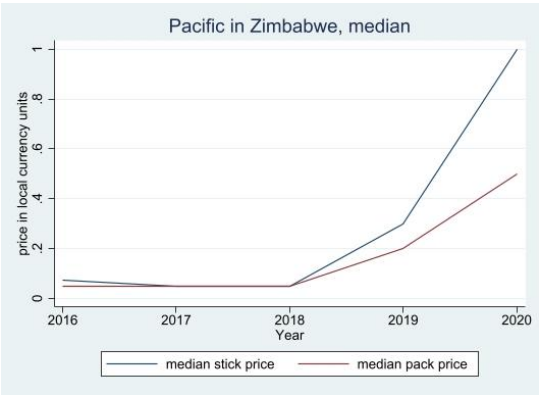

Master in Tanzania, median
price in local currency units
80
100
120
140
160
2017
2018
2019
2020
Year
median stick price
median pack price

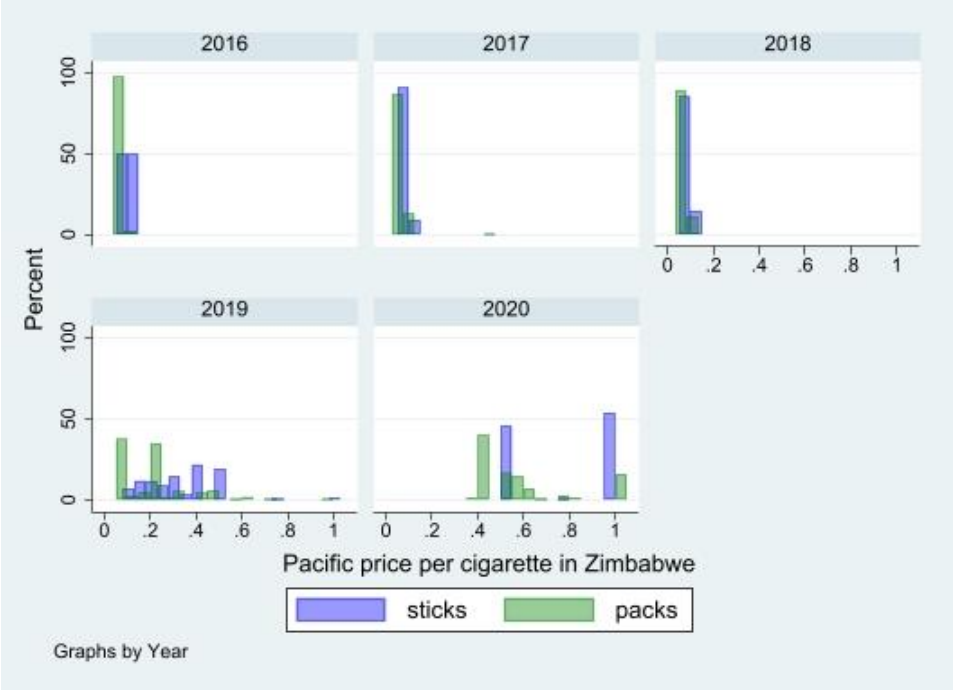

2016
2017
2018
2019
2020
Percent
0
50
100
50
100
150
200
250
Master price per cigarette in Tanzania
sticks
packs
Graphs by Year

Mark up on loose as a share of packed cigarette price	Average	Standard Deviation	Negative Share
	5.1%	20.9%	5.6%

Zimbabwe







Mark up on loose as a share of packed cigarette price	Average	Standard Deviation	Negative Share
	17.4%	29.8%	6.2%

Appendix D: Summary statistics on switchers

Table D1: Frequency and share of people who smoke and have switched between loose and packed cigarettes, by country.

	Packs to Loose		Loose to Packs	
	Frequency	Share	Frequency	Share
Bangladesh	447	5.3%	729	8.7%
India	132	4.3%	155	5.1%
Kenya	16	0.9%	49	2.8%
Thailand	602	7.4%	27	0.3%
Zambia	7	0.9%	15	1.9%

Source: Authors' elaboration on ITC data, switching is defined as a difference in the type of tobacco last acquire by a respondent in 2 consecutive waves.