An Econometric Analysis of the Illicit South African Tobacco Market

by

Cassandra Pengelly

Stellenbosch
UNIVERSITY
IYUNIVESITHI
UNIVERSITEIT

Thesis presented in partial fulfilment of the requirements for the degree of Master in Economics in the Faculty of Economic and Management Sciences at Stellenbosch University

Supervisor: Prof. W. H. Boshoff

August 2023

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: ....

Abstract

An Econometric Analysis of the Illicit South African Tobacco Market

C. Pengelly

Thesis: MCom (Economics)

August 2023

Using monthly data This paper investigates the constraint imposed on the legal tobacco market by the presence of illicit cigarettes.

This is the second paragraph of my English abstract.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This study examines the relationship between the legal and illegal tobacco markets in South Africa. Section 3.2 discusses the data used and how it was cleaned. Section 3 explains the methodology, where a VECM model is presented. The final section details discussion points (4). The appendix contains the full model outputs.

Chapter 2

The Illicit Tobacco Market

The health problems associated with tobacco consumption are significant and have been well-documented (Tingum, Mukong and Mdege, 2020:62). Worldwide, smoking is one of the leading causes of preventable deaths; and in South Africa 23% of the total deaths in 2018 were smoking-related (Statistics South Africa). Given the large and negative impact of smoking, the tobacco market has come under increasing scrutiny by governments and health organizations. A significant component of the tobacco market is the illicit cigarette trade, which the Financial Action Task Force (2012:7) defines as

> "*. . . the supply, distribution and sale of smuggled genuine, counterfeit or cheap white tobacco products.*"

The illicit tobacco market gives rise to concern on three main fronts: public health, tax evasion, and criminal activity. From a public health perspective, illicit cigarettes make smoking more affordable and thereby increases tobacco access (van der Zee, van Walbeek and Magadla, 2020:242). Cheaper cigarettes may induce non-smokers to smoke, increase the volume of cigarettes consumed by smokers, and decrease the likelihood that smokers will quit smoking (**?**:pagenumber; van der Zee *et al.*, 2020:242).

According to the International Agency for Research on Cancer, tobacco use is more prevalent among low socio-economic groups, and the poor are more sensitive to cigarette prices (IARC, 2011:276). Consequently, the health problems linked to smoking are disproportionately higher among the poor, which results in a greater burden on the public healthcare sector (). In South Africa, the healthcare costs associated with smoking amounted to R14.48 billion in 2016, which accounted for 4.1% of health spending (). In addition to the healthcare costs, there are indirect costs linked to smoking such as the loss of productive lives and loss of productive days due to illness (Boachie, Rossouw and Ross, 2020:pg). In 2016, the total cost of smoking to South Africa was R42.32 billion (0.97% of GDP).

Evidently, the public health and economic costs of tobacco use are extensive. In an effort to reduce tobacco use, many countries have implemented tax and price policies on tobacco products (Chaloupka, Straif, Leon). For a middle-income country, South Africa was considered to be at the forefront of tobacco control policies, including excise duties, for many years (Vellios, van Walbeek and Ross, 2022). From a fiscal perspective, excise tax on cigarettes is a source of government revenue, which is undermined by the illicit tobacco industry. **Blah ble** estimates the loss of income due to the illicit market to be R123123 The loss of income

Chapter 3

Methodology

3.1 Theoretical Model

This paper makes use of the demand model to estimate the effect of a relative price change, between licit and illicit cigarette price,on cigarette consumption in South Africa from January 2012 to March 2020. The relationship between cigarette consumption, price and income is well-documented (for example, see Boshoff (2008); Tingum *et al.* (2020); Organization *et al.*, 2010) and is typically expressed as:

$$Q_t = f(P_t, Y_t, D_t) \tag{3.1.1}$$

3.1.1 Where Q_t represents cigarette consumption volume in period t, P_t is the price of cigarettes adjusted for inflation, Y_t is disposable income, and D_t is is the dummy variable for the change in market structure that occurred in 2018.

Two sets of prices are used in this study; the first is the legal cigarette price, which is the price of cigarettes for which duties have been paid. The second is the illicit cigarette price, which is the price of cigarettes for which duties have been paid. For this study, cigarette consumption, as represented by Q_t, is the quantity of legal cigarettes consumed (cigarettes for which duties have been paid).

3.2 Data

The sample period for this study runs from January 2012 to March 2020. Monthly data is used such that there are 99 observation points for each variable in the data set. One of the advantages of using monthly data rather than annual data is that it allows for more degrees of freedom. The data used in this study include figures for South Africa for the average prices of legal cigarettes, the average prices of illicit cigarettes, volume of legal cigarettes consumed, tobacco excise duties, Value Added Tax (VAT), and real disposable income. The data for legal and illicit prices, and legal cigarette consumption are provided by a prominent South African cigarette manufacturer. The other data are extracted from the South African Reserve Bank (SARB) and Statistics South Africa (StatSA).

To prepare the legal price data for analysis, the most popular price category (MPPC) is identified as the 20-cigarette pack. A weighted average of before-tax 20-pack prices is used as a base price. The excise duty per 20's pack and VAT are then added to the base price to calculate the price of legal cigarettes. The legal and illicit prices, and disposable income values, are adjusted for inflation, taking December 2016 as the base month and year, respectively. For the statistical analysis in sections 3.3 and 3.4, all of the variables have been transformed into log form to reduce variability.

For the model specification, the two price series are combined to form a single price-ratio series. The price ratio is calculated as the real legal price divided by the real illicit cigarette price ($P_{legal}/P_{illicit}$). Consumer theory suggests that individuals base decisions on relative price changes rather than absolute price changes **ADD REFERENCES**. Thus, it makes intuitive sense to model the effect of a relative change in prices on the quantity of legal cigarettes

The series are plotted in levels in figure 3.1 below. The general downwards trend of the graph depicting the real price of legal cigarettes suggests that inflation has been outstripping nominal cigarette prices. Real legal prices decreased slightly over the period 2012 - 2018. On the other hand, real prices for illicit cigarettes remained fairly constant from 2012 until late 2014. Following a small and sharp uptick, illicit prices slowly decreased from late 2014 until November 2018 when prices increase rapidly. The real price ratio is generally flat from 2012 - 2016, and then trends upwards slightly, implying that real legal prices were decreasing more slowly than real illicit prices were decreasing over the same period. There is a sudden drop in the price ratio in November 2018, reflecting the spike in real illicit prices. The graphs of the respective logged series (shown in appendix A.1) illustrate similar patterns.

The plot of the real price ratio suggests that there is a structural break in the series, in late 2018. To test for this[1], the Chow structural test (based on the work by Chow, 1960) is employed for November 2018; the results are presented in table A.1. The p-value is less than 0.05; thus, the null hypothesis is rejected and there is sufficient evidence to conclude that the data contains a structural breakpoint. To control for this in the model, a dummy variable is used to capture the change in the illicit price (and therefore the price ratio) that occurred in November 2018. The variable is labelled '*Structural change*' and is coded 0 for periods before November 2018, and 1 otherwise.

The graph for the quantity of legal cigarettes in figure 3.1 shows a general downwards trend from 2012 to March 2020, and the series exhibits strong signs of seasonality. The series is decomposed and its separate components are graphed in appendix A.2, which confirm the presence of seasonality. While the series could be seasonally adjusted to account for the seasonality, this may change the outcome of conventional unit root tests, according to Ghysels, 1990, and lead to bias in the estimation of relationships between series (Bell and Hillmer, 2002:110). Therefore, this study uses the unaltered data for this consumption variable in the model estimation.

The overall trends and shapes of the graphs suggest that the series are not stationary, which is formally tested below in section 3.3.

[1]The test was conducted using the programming language *'R'*, using the *'strucchange'* package by Zeileis, Leisch, Hornik and Kleiber (2002) and Zeileis, Kleiber, Krämer and Hornik (2003)

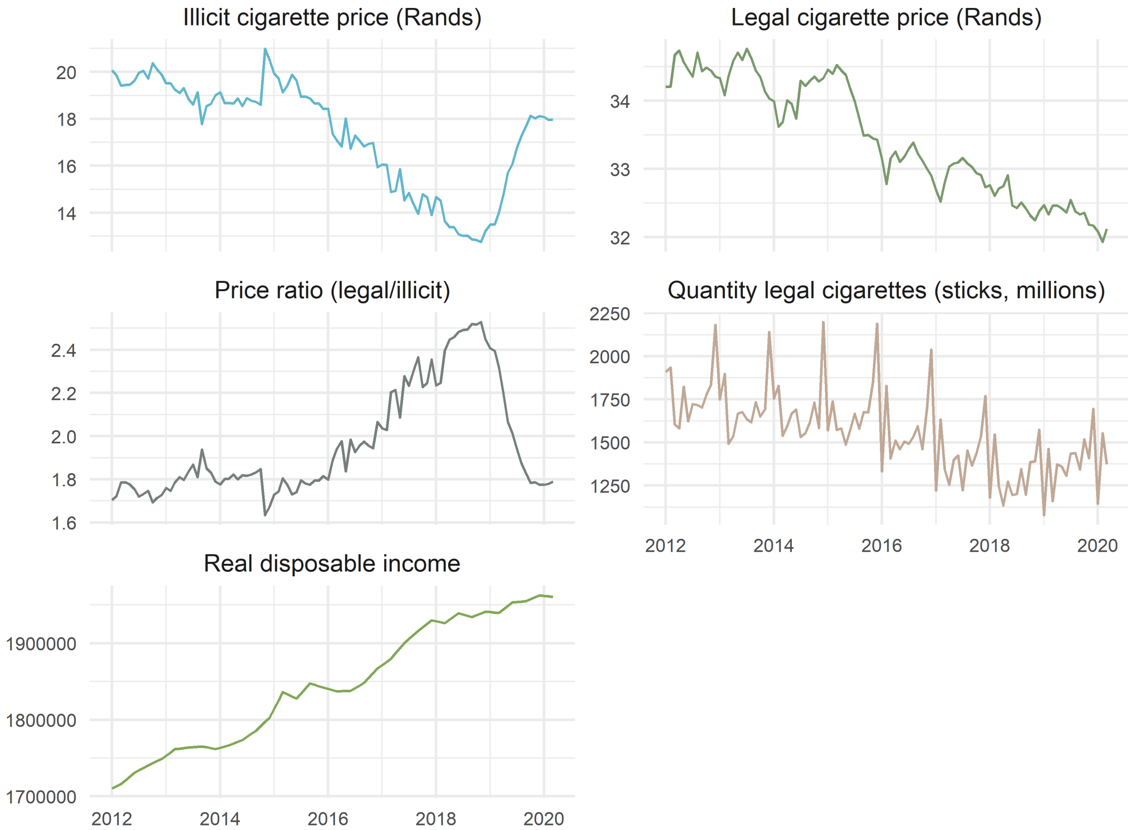

Figure 3.1: Plot for variables at levels

3.3 Stationarity and Cointegration

To test whether the series are stationary, three tests were employed: the Augmented Dickey-Fuller[2] (ADF) and Phillip-Perron[3] (PP) unit root tests, and the Kwiatkowski–Phillips–Schmidt–Shin[4] (KPSS) stationarity test. The results are presented in appendix B.

The results of the ADF tests in table B.1 show that the log of cigarette consumption, log of the real price ratio, and log of real disposable income all contain a unit root since the null hypothesis cannot be rejected, even at a 10% signifiance level. The KPSS stationarity tests (B.3) confirm these results, where the null hypothesis of stationarity is rejected at a 1% level of signifiance for each series. The results of the PP tests are presented in table B.2, which

[2]Based on the work of Dickey and Fuller (1979), using the '*urca*' package by Pfaff (2008)

[3]Based on the work of Phillips and Perron (1988) and using the '*tseries*' package by Trapletti and Hornik (2018)

[4]Based on the work of Kwiatkowski, Phillips, Schmidt and Shin (1992), using the '*tseries*' package by Trapletti and Hornik (2018)

shows that the log of cigarette consumption does not contain a unit root at 1% significance level but that the other two series do contain unit roots.

Given that there is a structural break point in the price ratio series, standard unit root testing may not be appropriate for this series. Consequently, the Zivot-Andrews Unit Root test is used as a robustness check for the stationarity of the price ratio. The results are presented in table B.4 and show that the null of a unit root cannot be rejected, at even 10% level of significance. Thus confirming the original test results.

Based on the unit root test results, the variables are assumed to be non-stationary in levels.

After first differencing each of the series, all tests indicate that the variables are stationary at a 5% level of significance. These results are presented in tables B.5, B.6, B.7, and B.8 for the ADF, PP, KPSS and Zivot-Andrews tests respectively. Given that the series are non-stationary in levels but are stationary in first differences, a cointegration approach can be used to evaluate whether a long-run relationship exists among the variables.

Cointegration

To test for conintegrating relationships, two tests were employed: the Johansen Trace test for Cointegration, and the Johansen Maximum Eigenvalue test for Cointegration based on the work of Johansen and Juselius, 1990.

The Johansen Trace Test for Cointegration indicates that there is one cointegrating vector. As shown in table C.1, the test statistic of 80.8 is larger than 60.2 (the critical value at a 1% level of significance). Thus, the null hypothesis is rejected, where the null hypothesis is that there are zero cointegrating vectors. However, the test statistic 41 is lower than 41.1 (the critical value at a 1% level of significance). This means that the null hypothesis cannot be rejected even at a 1% signicance level, where the null hypothesis is that there is 1 or fewer cointegrating vectors.

The results of the Johansen Maximum Eigenvalue test are presented in table C.2. Similar to the Trace test, the Maximum Eigenvalue test shows that there is a single cointegrating vector. Both cointegration tests were run with a lag length of 3, indicating that the appropriate number of lags to use in the Vector Error Correction Model is 2 lags. The presence of a cointegrating vector indicates that a long-run relationship exist among the variables. A Vector Error Correction Model is used to combine the long and short-run dynamics. This model is specified and estimated in the following section (3.4).

3.4 Vector Error Correction Model

The equation representing the long-run equilibrium model is a double-logarithmic demand equation, given by:

$$lnQ_t = \alpha_0 + \alpha_1 lnP_t + \alpha_2 lnY_t + \alpha_3 lnD_t + \mu_t \tag{3.4.1}$$

where $i = 0, 1, 2, 3,$
lnQ_t is the log of legal cigarette consumption,
lnP_t is the log of real price ratio for cigarettes,
lnY_t is the log of real disposable income,
D_t is the dummy variable for the structural break
μ_t is the random error term, and t is measured in months

The Vector Error Correction model specification is given by:

$$\begin{aligned} dLnQ_t =& \beta_0 + \beta_1 \sum_{i=1}^{n-1} dLnQ_{t-i} + \beta_2 \sum_{i=1}^{n-1} dLnP_{t-i} + \beta_3 \sum_{i=1}^{n-1} dLnY_{t-i} && (3.4.2) \\ &+ \beta_4 D_t + \lambda_i ECT_{t-1} + \mu_t && (3.4.3) \end{aligned}$$

where $j - 1$ is the lag length, reduced by 1 due to the loss of a lag when differencing a Vector Autoregression (VAR),
d is the difference operator,
β_i represents model's short-run coefficients,
λ_i is the speed of adjustment parameter,
ECT_{t-1} is the error correction term, and
μ_t is the random error term, and t is measured in months

Chapter 4

Analysis

The VECM was with 1 cointegrating vector and at lag length of 2. The results are presented in table the appendix (D). The price ratio captures a relative change of the legal price compared to the illicit price. If the ratio shrinks, it indicates that the cost of legal cigarettes decreased relative to the cost of illicit cigarettes. There should be a negative long run relationship between the price ratio and the quantity of legal cigarettes. There should be a positive sign for the cointegrating relationship, which there is. The coefficient for the real disposable income variable should be negative for the long-run relationship (so that the reverse sign is positive); whereas the sign is positive in the Vecm results below.

The VECM was estimated at lag length of 1 with 1 cointegrating vector, and the results are summarized in Table 2. The demand equation for annual cigarette consumption obtained from the VECM and 2SLS estimations is reported in Tables 2. Table 3 presents the rst stage estimates of the 2SLS method. Appendix 6 shows estimated results of the demand equation using specic dummies for the existing legislative acts in South Africa. The relevant legislation is the Tobacco Product Control Act, no 83 in 1993, that was amended in 1999, 2007 and 2008. The VECM uses stationary data at rst dierences and includes the lagged residuals of the long-run relationship as an explanatory variable. Coe¢ cients from ECM represent the relationship in the short run, and the coe¢ cient of the lagged residuals measures the speed of convergence to the long-run equilibrium. The error correction term is the speed of adjustment in the direction of long-term equilibrium after any deviation from the steady state. The error correction terms in Table 2 have the correct sign and are signicant. This indicates that per capita cigarette demand converges to steady state equilibrium at the speed of 15% in the restricted model (see column 1) and 26% in the unrestricted model (see column 2). The estimate

If the real disposable income variable is excluded then the Vecm results are as follows. The coefficient for

The previous month's deviation from the long run equilibrium is corrected at a speed of

The results of the diagnostic tests are shown in Appendix E. Table E.1 shows the results of the serial autocorrelation test. Table E.2 shows the arch results. Table E.3 shows the normal results.

Figure E.1 shows the residuals.

Chapter 5

Conclusion

Appendices

Appendix A

Time Series Data

A.1 Time series logged graphs

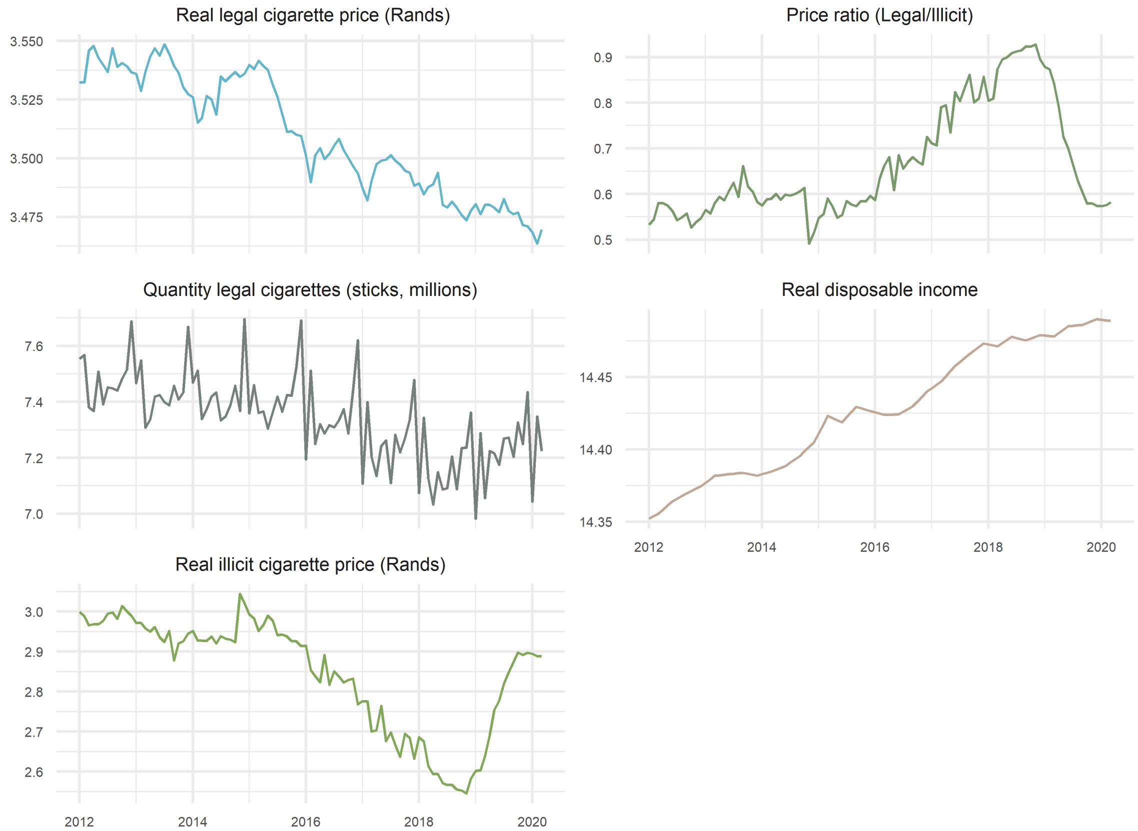

Figure A.1: Plot for variables at levels

A.2 Seasonality decomposition

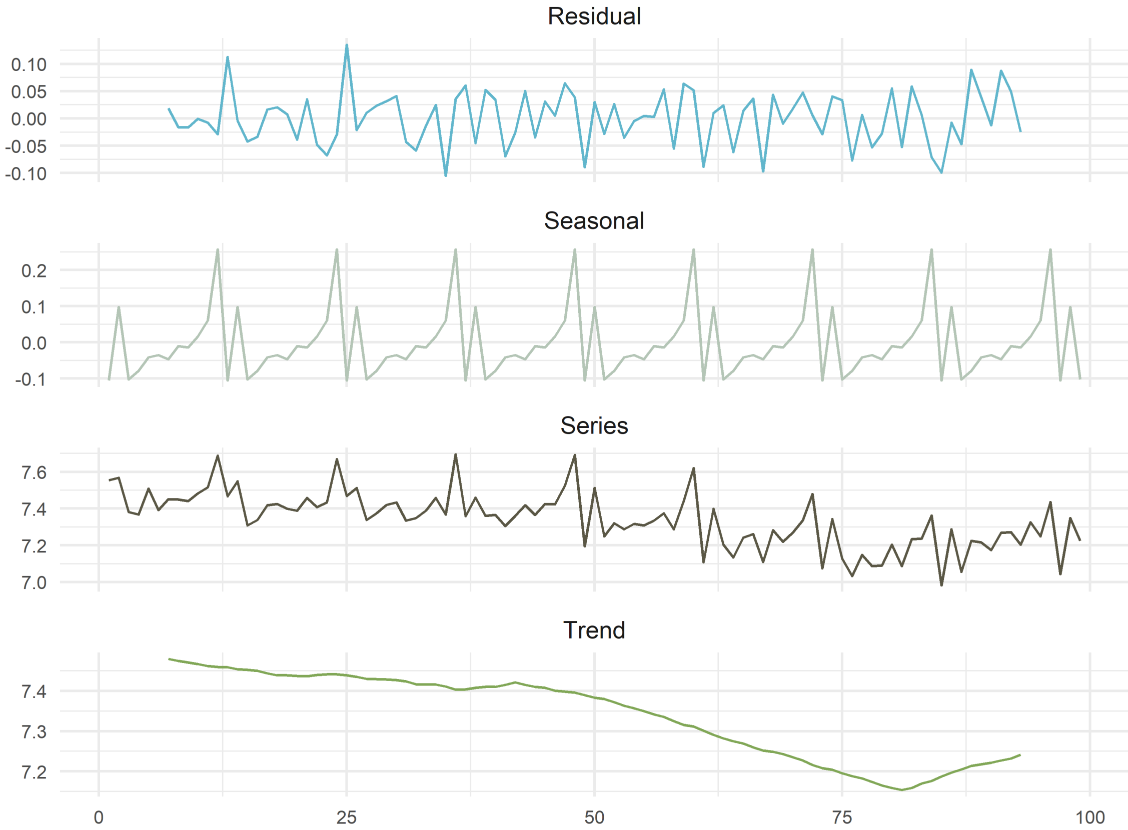

Figure A.2: Decomposisiton and seasonality for quantity of legal cigarettes series

A.3 Structural breakpoint test

Table A.1: Chow structural break test for Real Price Ratio time series

	Statistic	**p-value**	**Alternative**
Chow test	67.13	0	Structural breakpoint

Appendix B

Stationarity Testing

Table B.1: Augmented Dickey Fuller Tests

	Test statistic	**1%**	**5%**	**10%**	**Alternative**
Quantity legal cigarettes	-0.445	-2.6	-1.95	-1.61	stationary
Price ratio	-0.0835	-2.6	-1.95	-1.61	stationary
Disposable income	2.63	-2.6	-1.95	-1.61	stationary

Table B.2: Phillips-Perron Unit Root Test

	Statistic	**p-value**	**Alternative**
Quantity legal cigarettes	-119	0.01	stationary
Price ratio	-2.54	0.952	stationary
Disposable income	-6.73	0.726	stationary

Table B.3: KPSS Unit Root Test

	Statistic	**p-value**	**Alternative**
Quantity legal cigarettes	1.89	0.01	non-stationary
Price ratio	1.35	0.01	non-stationary
Disposable income	2.55	0.01	non-stationary

Table B.4: Zivot-Andrews Unit Root Test

	Test statistic	**1%**	**5%**	**10%**	**Alternative**
Price ratio	-3.64	-5.57	-5.08	-4.82	stationary

Table B.5: Augmented Dickey Fuller Tests

	Test statistic	**1%**	**5%**	**10%**	**Alternative**
Quantity legal cigarettes	-11.8	-2.6	-1.95	-1.61	stationary
Price ratio	-7.7	-2.6	-1.95	-1.61	stationary
Disposable income	-3.47	-2.6	-1.95	-1.61	stationary

Table B.6: Phillips-Perron Unit Root Test for Differenced Series

	Statistic	**p-value**	**Alternative**
Quantity legal cigarettes	-153	0.01	stationary
Price ratio	-120	0.01	stationary
Disposable income	-33	0.01	stationary

Table B.7: KPSS Stationarity Test for Differenced Series

	Statistic	**p-value**	**Alternative**
Quantity legal cigarettes	0.0229	0.1	non-stationary
Price ratio	0.275	0.1	non-stationary
Disposable income	0.149	0.1	non-stationary

Table B.8: Zivot-Andrews Unit Root Test

	Test statistic	**1%**	**5%**	**10%**	**Alternative**
Price ratio	-13.2	-5.57	-5.08	-4.82	stationary

Appendix C

Cointegration Tests

Table C.1: Johansen Trace Test

	Statistic	**10%**	**5%**	**1%**
r<=3	4.86	7.52	9.24	13
r<=2	19.2	17.9	20	24.6
r<=1	41	32	34.9	41.1
r=0	80.8	49.6	53.1	60.2

Table C.2: Johansen Maximum Eigenvalue Test

	Statistic	**10%**	**5%**	**1%**
r<=3	4.86	7.52	9.24	13
r<=2	14.3	13.8	15.7	20.2
r<=1	21.8	19.8	22	26.8
r=0	39.9	25.6	28.1	33.2

Appendix D

Vector Error Correction Results

Table D.1: Vector Error Correction Model

	Long run coefficient
Quantity cigarettes	1
Price ratio	0.885
Real income	-0.679
Structural change	0.354
Constant	1.79

Table D.2: Vector Error Correction Model Short run

	Quantity Cigarettes	Price ratio	Real income	Structural Change
ECT	-0.488 ***	-0.149 ***	-0.002	-0.042
	(0.114)	(0.035)	(0.002)	(0.124)
Cigarette Quantity (lag 1)	-1.068 ***	-0.081 *	-0.001	0.035
	(0.101)	(0.031)	(0.001)	(0.111)
Price ratio (lag 1)	-2.156 ***	-0.334 **	-0.003	-0.105
	(0.339)	(0.105)	(0.005)	(0.370)
Real Income (lag 1)	2.502	-0.195	0.896 ***	11.068
	(7.578)	(2.351)	(0.106)	(8.270)
Structural change (lag 1)	0.176	-0.026	0.000	0.001
	(0.097)	(0.030)	(0.001)	(0.106)
Cigarette Quantity (lag 2)	-0.649 ***	-0.154 ***	-0.001	-0.042
	(0.123)	(0.038)	(0.002)	(0.134)
Price ratio (lag 2)	-1.465 ***	-0.330 **	-0.007	0.102
	(0.397)	(0.123)	(0.006)	(0.433)
Structural change (lag 2)	-0.290 **	-0.012	-0.001	-0.007
	(0.098)	(0.030)	(0.001)	(0.107)
N	96	96	96	96
R2	0.694	0.224	0.657	0.035

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Appendix E

Diagnostic Tests

Table E.1: Autocorrelation Test

	Statistic	Parameter	p-value	Lags
Portmanteau Test (asymptotic)	265	196	0	15
Portmanteau Test (asymptotic)	468	436	0.14	30

Table E.2: ARCH test

	Statistic	Parameter	p-value	Lags
ARCH (multivariate)	810	1500	1	15

Table E.3: Normality test

	Statistic	Parameter	p-value
JB-Test (multivariate)	29175.0802324414	8	0

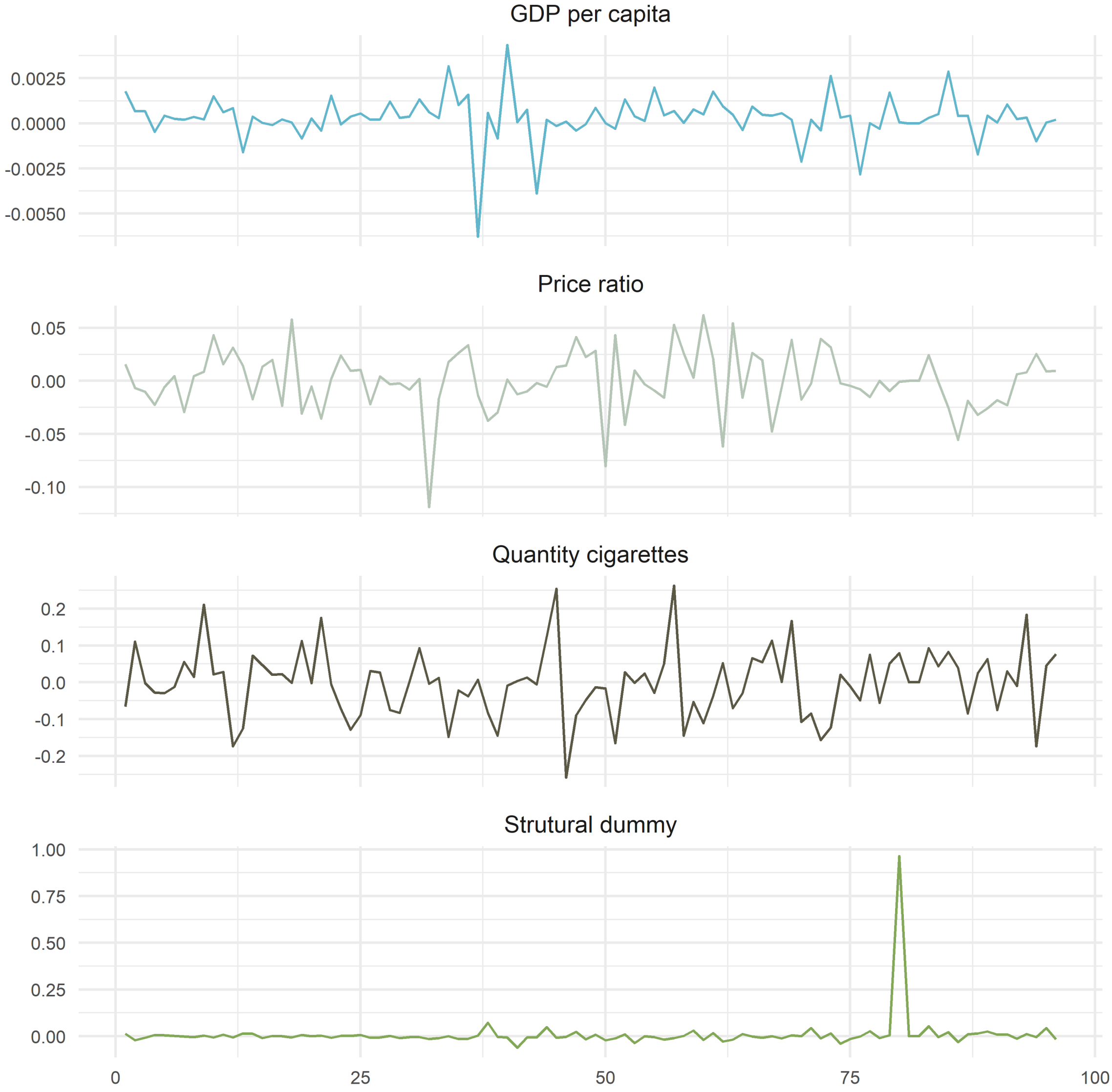

Figure E.1: Plot of residuals

List of References

Bell, W.R. and Hillmer, S.C. (2002). Issues involved with the seasonal adjustment of economic time series. *Journal of Business & Economic Statistics*, vol. 20, no. 1, pp. 98–127. ISSN 07350015.
Available at: http://www.jstor.org/stable/1392153

Boachie, M.K., Rossouw, L. and Ross, H. (2020). The Economic Cost of Smoking in South Africa, 2016. *Nicotine & Tobacco Research*, vol. 23, no. 2, pp. 286–293. ISSN 1462-2203.

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*, vol. 76, no. 1, pp. 118–131. ISSN 1813-6982.

Chow, G.C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica: Journal of the Econometric Society*, pp. 591–605.

Dickey, D.A. and Fuller, W.A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, vol. 74, no. 366a, pp. 427–431.

Ghysels, E. (1990). Unit-root tests and the statistical pitfalls of seasonal adjustment: The case of u.s. postwar real gross national product. *Journal of Business & Economic Statistics*, vol. 8, no. 2, pp. 145–152. ISSN 07350015.
Available at: http://www.jstor.org/stable/1391976

Johansen, S. and Juselius, K. (1990). Maximum likelihood estimation and inference on cointegration—with appucations to the demand for money. *Oxford Bulletin of Economics and statistics*, vol. 52, no. 2, pp. 169–210.

Kwiatkowski, D., Phillips, P.C., Schmidt, P. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics*, vol. 54, no. 1-3, pp. 159–178.

Organization, W.H., Initiative, W.T.F. *et al.* (2010). Economics of tobacco toolkit: economic analysis of demand using data from the global adult tobacco survey (gats).

Pfaff, B. (2008). *Analysis of Integrated and Cointegrated Time Series with R*. 2nd edn. Springer, New York. ISBN 0-387-27960-1.
Available at: https://www.pfaffikus.de

Phillips, P.C. and Perron, P. (1988). Testing for a unit root in time series regression. *biometrika*, vol. 75, no. 2, pp. 335–346.

Tingum, E.N., Mukong, A.K. and Mdege, N. (2020). The effects of price and non-price policies on cigarette consumption in South Africa. *Tobacco Induced Diseases*, vol. 18, p. 62. ISSN 2070-7266.

Trapletti, A. and Hornik, K. (2018). *tseries: Time Series Analysis and Computational Finance*. R package version 0.10-43.
Available at: https://CRAN.R-project.org/package=tseries

van der Zee, K., van Walbeek, C. and Magadla, S. (2020). Illicit/cheap cigarettes in South Africa. *Trends in Organized Crime*, vol. 23, no. 3, pp. 242–262. ISSN 1936-4830.

Vellios, N., van Walbeek, C. and Ross, H. (2022 July). Measuring the illicit cigarette market in the absence of pack security features: A case study of South Africa. *Tobacco Control*, vol. 31, no. 4, pp. 580–585. ISSN 0964-4563, 1468-3318.

Zeileis, A., Kleiber, C., Krämer, W. and Hornik, K. (2003). Testing and dating of structural changes in practice. *Computational Statistics & Data Analysis*, vol. 44, no. 1–2, pp. 109–123.

Zeileis, A., Leisch, F., Hornik, K. and Kleiber, C. (2002). strucchange: An r package for testing for structural change in linear regression models. *Journal of Statistical Software*, vol. 7, no. 2, pp. 1–38.