

Modelling Tobacco Demand: How the Illicit Cigarette Market Constrains the Legal Market

Cassandra Pengelly^a

^a*Stellenbosch University*

1. Introduction

This thesis draft lays out the method and model used to study the relationship between the legal and illegal tobacco market in South Africa. Section 2 discusses the data used and how it was cleaned. Section 3 explains the methodology, where a VECM and a VAR model are presented. The final section details discussion points (4). The appendix 4 contains the full model outputs.

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 includes figures for the prices and volumes of cigarettes in South Africa, tobacco excise duties, VAT, and disposable income. To prepare the data for analysis the most popular price category (MPPC) was identified as the 20-cigarette pack. Then a weighted average of before-tax 20-pack prices was used as a base price. The excise duty per 20's pack and VAT were then added to the base price to calculate the price of licit cigarettes. The licit, illicit and disposable income amounts were adjusted for inflation, taking December 2016 as the base month and year. All of the variables have been transformed into log form.

The figure below 2.1 plots the time series of the logged variables. The graphs show that the data could be trending, which is formally tested in section 3.

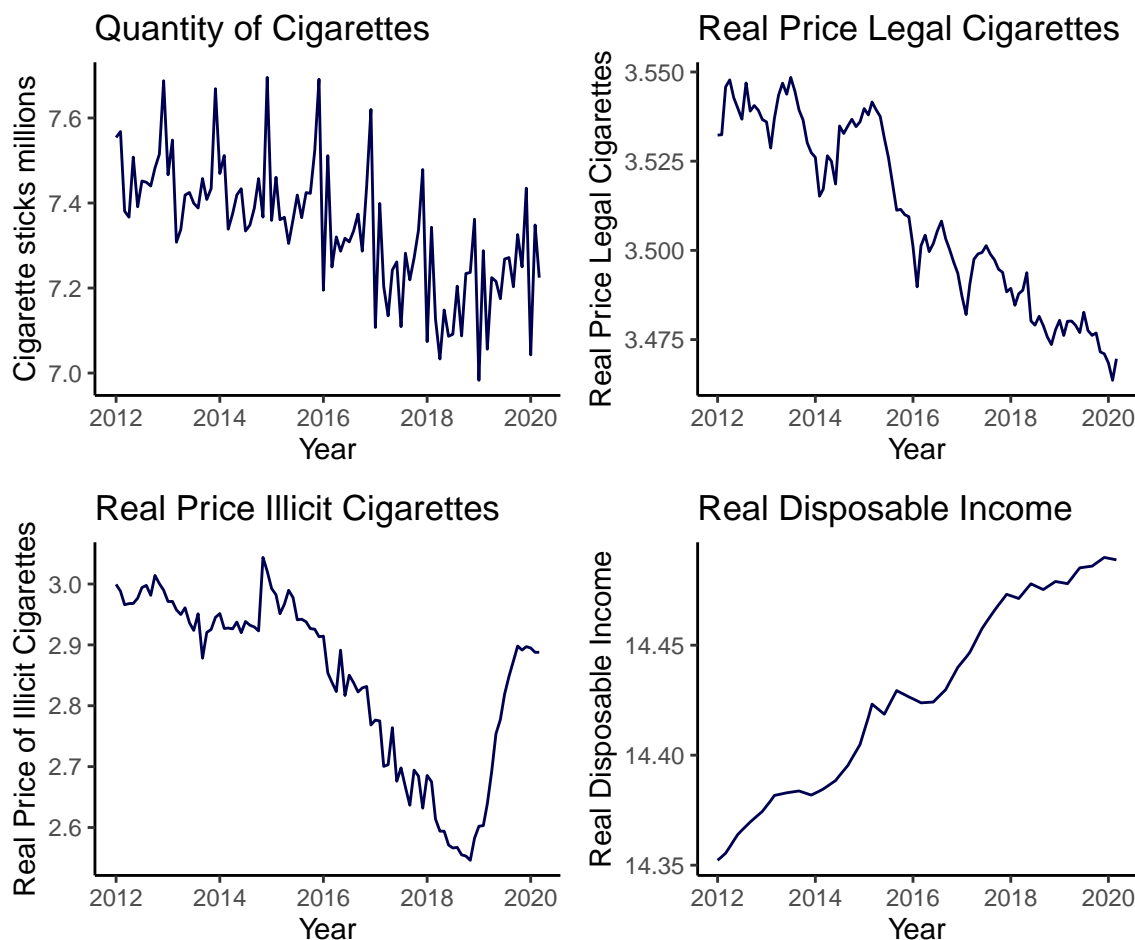


Figure 2.1: Time Series Plot

3. Methodology

To check whether the data is stationary, a number of tests is employed. First the autocorrelation functions are plotted below 3.1. They indicate that all four series are persistent; this is confirmed by the Ljung-Box tests in table 3.1. The Ljung-Box test for independence assesses whether there is significant evidence for non-zero correlations at a given lag, with the null hypothesis that there is independence in a given time series. A low p-value indicates a signal of non-stationarity. The augmented Dickey Fuller test (3.2) suggests that all four of the series contain a unit root (using the number of lags as 10¹). This further suggests that the series are non-stationary.

¹Some of the series test as stationary when the number of lags is reduced

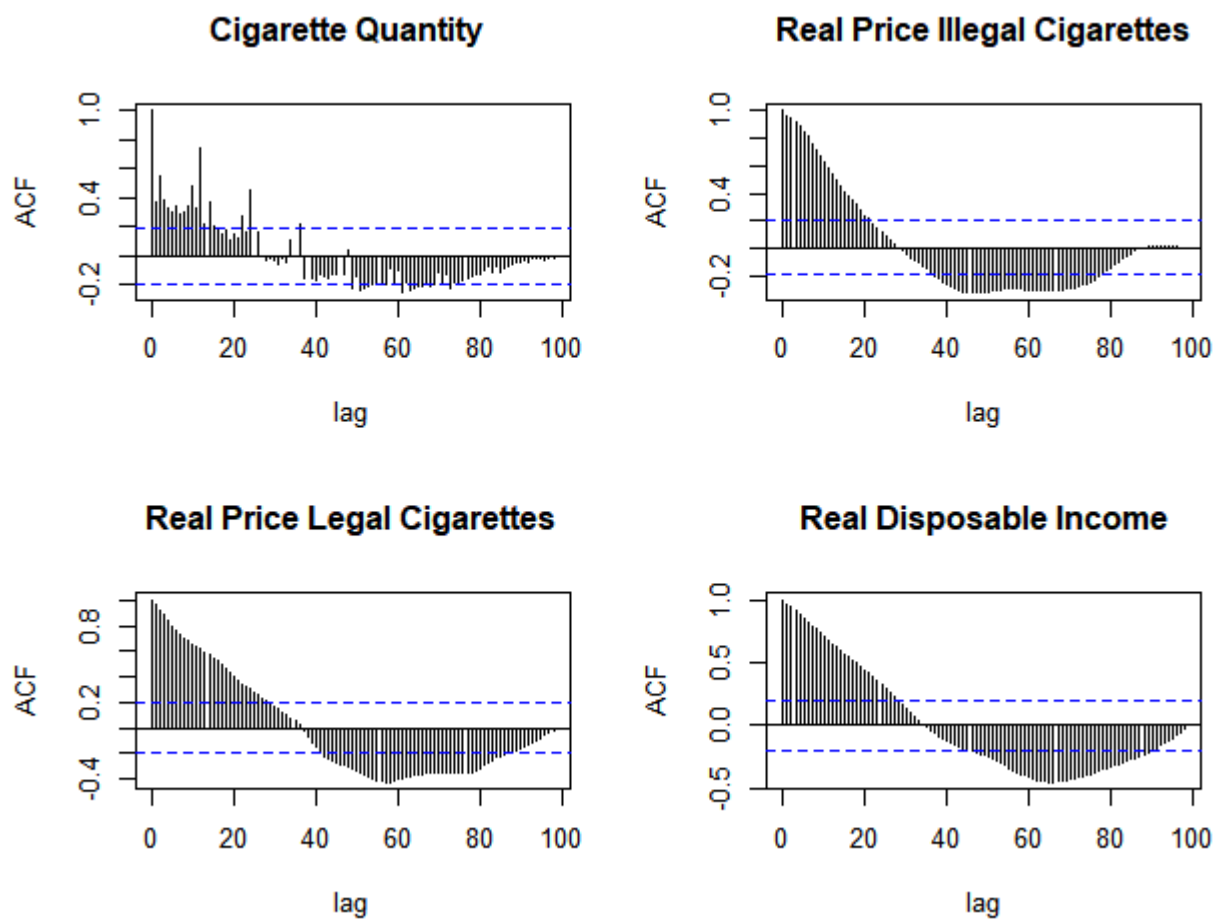


Figure 3.1: ACF Plots

Time Series	p value	Test Result	Interpretation
Cigarette Quantity	0.00	Reject Null	Non-stationary
Real Price Legal	0.00	Reject Null	Non-stationary
Real Price Illegal	0.00	Reject Null	Non-stationary
Real Disposable Income	0.00	Reject Null	Non-stationary

Table 3.1: Ljung-Box Test

Time Series	p value	Test Result	Interpretation
Cigarette Quantity	0.61	Fail to reject Null	Non-stationary
Real Price Legal	0.22	Fail to reject Null	Non-stationary
Real Price Illegal	0.38	Fail to reject Null	Non-stationary
Real Disposable Income	0.06	Fail to reject Null	Non-stationary

Table 3.2: Augmented Dickey-Fuller Test

To assess whether a long-run relationship between the variables exists, the Johansen test is employed. According to Akaike's information criterion (AIC), the appropriate maximum number of lags to use is 10 (3.3). A lag order of 9 is used for the test, since the test requires a lag order of $N - 1 = 10 - 1 = 9$. Two Johansen tests are used: the Trace and the Maximum Eigenvalue tests. The Trace statistic test (3.4) shows we reject the null hypothesis that there are zero cointegrating relationships: the test statistic 85.85 is greater than the 1% significance level of 55.43. The test results indicate that there is 1 cointegrating relationship. Similarly, the Maximum Eigenvalue test rejects that there are zero cointegrating relationships, and fails to reject that there is at most 1 cointegrating relationship. The presence of a cointegrating vector amongst the variables suggests a Vector Error Correction Model is appropriate to analyse the variable dynamics.

AIC(n)	HQ(n)	SC(n)	FPE(n)
10	5	2	10

Table 3.3: Optimal Lag Selection

	Test Statistic	10%	5%	1%
$r \leq 3$	1.21	6.50	8.18	11.65
$r \leq 2$	6.04	15.66	17.95	23.52
$r \leq 1$	26.89	28.71	31.52	37.22
$r = 0$	85.85	45.23	48.28	55.43

Table 3.4: Johansen Trace Test for Cointegration Results

	Test Statistic	10%	5%	1%
$r \leq 3$	1.21	6.50	8.18	11.65
$r \leq 2$	4.83	12.91	14.90	19.19
$r \leq 1$	20.85	18.90	21.07	25.75
$r = 0$	58.95	24.78	27.14	32.14

Table 3.5: Johansen Eigenvalue Test for Cointegration Results

3.1. Vector Error Correction Model

A summary of the VECM results is given in 3.6. The full model output with the 9 lags is given in the appendix 4. The results show that the error correction terms are not significant, with the exception of the cigarette quantity term, and very few of the first lag coefficients are significant.

	ECT	Intercept
QDP	-1.9701(0.5869)**	0.0054(0.0238)
PREALWAPDP	0.0598(0.0301).	-0.0022(0.0012).
PREALWAPDNP	-0.1659(0.1979)	0.0100(0.0080)
YDISPREAL	0.0078(0.0081)	0.0001(0.0003)

Table 3.6: Vector Error Correction Model 2OLS

	QDP-1	PREALWAPDP-1	PREALWAPDNP-1	YDISPREAL-1
QDP	0.7123(0.5286)	-6.5487(2.6841)*	0.5219(0.5503)	12.7417(10.2110)
PREALWAPDP	-0.0565(0.0271)*	0.0007(0.1375)	0.0567(0.0282)*	0.9735(0.5229).
PREALWAPDNP	0.1705(0.1783)	0.5743(0.9052)	-0.2648(0.1856)	-1.8064(3.4435)
YDISPREAL	-0.0078(0.0073)	0.0229(0.0369)	0.0093(0.0076)	0.8393(0.1406)***

Table 3.7: Vector Error Correction Model 2OLS

To test the accuracy of the model, a number of diagnostic tests were run. The majority of specification tests are modified for a Vector Autoregression model (VAR) so first the results of the VECM were converted to a VAR format. The Portmanteau serial correlation test shows that the null hypothesis of no autocorrelation is rejected at a 1% level of significance since the p-value is 0. The ARCH test shows that there are no ARCH effects (the test fails to reject the null). Finally, the normality test indicates the the residuals of the VECM model are not normally distributed. The serial correlation and normality test indicate that the model could be better specified. The next section explores an unstructured VAR model as an alternative.

Test	p value	Test Result	Interpretation
Portmanteau	0.00	Reject Null	Autorrelation

Table 3.8: Serial Correlcation Test

Test	p value	Test Result	Interpretation
ARCH	1.00	Fail to reject Null	No ARCH effects

Table 3.9: ARCH Tests

Test	p value	Test Result	Interpretation
Normality	0.00	Reject Null	Residuals not normally distributed

Table 3.10: Normality Tests

3.2. Unstructured VAR

If we proceed with using an unstructured VAR model despite the non-stationarity of the time series, we get the base model below (3.1):

$$Q_t = \mu + \sum_{i=1}^n \beta_i Q_{t-i} + \sum_{i=1}^n \gamma_i P_{t-i} + \sum_{i=1}^n \theta_i Y_{t-i} + \sum_{i=1}^n \phi_i I_{t-i} \quad (3.1)$$

where Q_t is the log of cigarette consumption,

P_t is the log of real cigarette price,

Y_t is the log of real disposable income,

I_t is the log of real illicit cigarette price,

n is the number of lags and t is measured in months

The results of the VAR are presented in table 3.11 below. The diagnostic tests are summarised in the tables under the regression output in 3.12. The VAR fairs better in the specification tests, although there is stil serial correlation present.

Table 3.11: Vector Autoregression

	<i>Dependent variable:</i>			
	y			
	(1)	(2)	(3)	(4)
Quantity.l1	−0.424*** (0.132)	−0.004 (0.007)	0.003 (0.042)	−0.001 (0.002)
Price.Duties.Paid.l1	−5.153* (2.681)	0.770*** (0.132)	0.649 (0.860)	−0.0005 (0.034)
Price.Duties.not.Paid.l1	1.418*** (0.450)	0.019 (0.022)	0.751*** (0.144)	0.004 (0.006)
Disposable.Income.l1	5.829 (10.816)	0.544 (0.532)	−2.093 (3.472)	1.737*** (0.138)
Quantity.l2	−0.122 (0.138)	−0.026*** (0.007)	−0.010 (0.044)	−0.004** (0.002)
Price.Duties.Paid.l2	3.291 (3.203)	−0.094 (0.158)	−0.763 (1.028)	−0.058 (0.041)
Price.Duties.not.Paid.l2	0.088 (0.586)	−0.021 (0.029)	0.306 (0.188)	0.008 (0.008)
Disposable.Income.l2	6.112 (21.486)	−0.940 (1.057)	−0.751 (6.896)	−0.584** (0.275)
Quantity.l3	−0.403*** (0.128)	0.005 (0.006)	−0.098** (0.041)	−0.001 (0.002)
Price.Duties.Paid.l3	−1.302 (3.132)	0.340** (0.154)	0.096 (1.005)	0.004 (0.040)
Price.Duties.not.Paid.l3	−1.099* (0.584)	0.054* (0.029)	0.138 (0.187)	0.002 (0.007)
Disposable.Income.l3	−28.207 (22.460)	0.895 (1.105)	9.412 (7.209)	−0.813*** (0.287)
Quantity.l4	−0.178 (0.141)	0.015** (0.007)	−0.126*** (0.045)	0.0004 (0.002)
Price.Duties.Paid.l4	2.948 (3.139)	−0.181 (0.154)	1.188 (1.007)	0.029 (0.040)
Price.Duties.not.Paid.l4	0.372 (0.562)	−0.027 (0.028)	0.041 (0.180)	−0.007 (0.007)
Disposable.Income.l4	17.786 (24.173)	−0.334 (1.190)	−10.458 (7.759)	1.150*** (0.309)
Quantity.l5	−0.132 (0.157)	0.007 (0.008)	0.145*** (0.051)	−0.0001 (0.002)
Price.Duties.Paid.l5	−0.123 (3.180)	−0.062 (0.157)	−0.587 (1.021)	0.010 (0.041)
Price.Duties.not.Paid.l5	−0.716 (0.564)	−0.045 (0.028)	−0.038 (0.181)	−0.017** (0.007)
Disposable.Income.l5	3.474 (25.876)	−0.383 (1.274)	−1.791 (8.305)	−0.509 (0.331)
Quantity.l6	−0.356** (0.159)	−0.015* (0.008)	0.069 (0.051)	0.001 (0.002)
Price.Duties.Paid.l6	−3.564 (3.042)	−0.072 (0.150)	−0.559 (0.976)	0.028 (0.039)
Price.Duties.not.Paid.l6	0.032 (0.482)	0.009 (0.024)	−0.163 (0.155)	0.006 (0.006)
Disposable.Income.l6	−17.569 (23.169)	0.052 (1.140)	9.746 (7.436)	0.049 (0.297)
Quantity.l7	−0.459*** (0.166)	0.009 (0.008)	−0.069 (0.053)	0.001 (0.002)
Price.Duties.Paid.l7	2.310 (2.887)	−0.056 (0.142)	1.381 (0.927)	−0.045 (0.037)
Price.Duties.not.Paid.l7	1.657*** (0.473)	0.010 (0.023)	−0.304* (0.152)	0.005 (0.006)
Disposable.Income.l7	8.108 (22.086)	−0.347 (1.087)	−8.990 (7.089)	−0.027 (0.283)
Quantity.l8	−0.375** (0.163)	−0.001 (0.008)	−0.060 (0.052)	−0.001 (0.002)
Price.Duties.Paid.l8	−1.022 (2.827)	0.093 (0.139)	−0.947 (0.907)	−0.033 (0.036)
Price.Duties.not.Paid.l8	−0.571 (0.520)	−0.014 (0.026)	0.303* (0.167)	−0.003 (0.007)
Disposable.Income.l8	1.136 (20.760)	1.055 (1.022)	5.354 (6.663)	−0.095 (0.266)
Quantity.l9	−0.316** (0.151)	−0.010 (0.007)	−0.040 (0.048)	−0.003 (0.002)
Price.Duties.Paid.l9	3.185 (2.367)	0.049 (0.116)	−0.384 (0.760)	0.007 (0.030)
Price.Duties.not.Paid.l9	0.546 (0.459)	0.018 (0.023)	−0.012 (0.147)	0.007 (0.006)
Disposable.Income.l9	−0.681 (10.231)	−0.715 (0.504)	−0.717 (3.284)	0.052 (0.131)
const	78.609*** (25.524)	3.378*** (1.256)	5.199 (8.192)	0.807** (0.327)
Observations	90	90	90	90
R ²	0.826	0.984	0.979	1.000
Adjusted R ²	0.708	0.973	0.965	0.999
Residual Std. Error (df = 53)	0.082	0.004	0.026	0.001
F Statistic (df = 36; 53)	7.007***	90.623***	68.995***	3,434.504***

Note:

*p<0.1; **p<0.05; ***p<0.01

Test	p value	Test Result	Interpretation
Portmanteau	0.00	Reject Null	Autorelation

Table 3.12: Serial Correlcation Test

Test	p value	Test Result	Interpretation
ARCH	1.00	Fail to reject Null	No ARCH effects

Table 3.13: ARCH Tests

Test	p value	Test Result	Interpretation
Normality	0.09	Fail to reject Null	Residuals normally distributed

Table 3.14: Normality Tests

The results of the VAR are then used to calculate elasticities according to the formula below:

$$\hat{E}_d = \frac{\sum_{t=1}^9 \hat{\gamma}_t}{1 - \sum_{t=1}^9 \hat{\beta}_t} \quad (3.2)$$

The elasticities are presented in the table below 3.15. The income elasticity is expected to be positive; however, it is unusual that both price elasticities are positive.

Illicit price elasticity of cigarette demand	0.27
Income elasticity	0.26
Price elasticity	0.21

Table 3.15: Elasticities

4. Discussion Points

There are several concerns that I have regarding the data work:

- How do I choose the appropriate maximum number of lags for the stationarity tests? Changing the max lags impacted the optimal lag selection results, which then changed some results from nonstationary to stationary.

- If some of the time series are stationary and some are nonstationary, should I use a VECM or a VAR?
- How do I interpret the coefficients of the VAR (and VECM) for the elasticities? I think this is where I am going wrong in my final calculations.

Appendix

The full model output for the VECM is given below (for all 9 lags):

```
##                               ECT           Intercept
## Equation log_QDP              -1.9701(0.5869)**  0.0054(0.0238)
## Equation log_PREALWAPDP        0.0598(0.0301).   -0.0022(0.0012).
## Equation log_PREALWAPDNP -0.1659(0.1979)         0.0100(0.0080)
## Equation log_YDISPREAL         0.0078(0.0081)     0.0001(0.0003)
##                               log_QDP -1  log_PREALWAPDP -1
## Equation log_QDP              0.7123(0.5286)     -6.5487(2.6841)*
## Equation log_PREALWAPDP       -0.0565(0.0271)*    0.0007(0.1375)
## Equation log_PREALWAPDNP      0.1705(0.1783)       0.5743(0.9052)
## Equation log_YDISPREAL       -0.0078(0.0073)       0.0229(0.0369)
##                               log_PREALWAPDNP -1  log_YDISPREAL -1
## Equation log_QDP              0.5219(0.5503)      12.7417(10.2110)
## Equation log_PREALWAPDP        0.0567(0.0282)*      0.9735(0.5229).
## Equation log_PREALWAPDNP -0.2648(0.1856)           -1.8064(3.4435)
## Equation log_YDISPREAL         0.0093(0.0076)        0.8393(0.1406)***
##                               log_QDP -2  log_PREALWAPDP -2
## Equation log_QDP              0.8721(0.4652).     -0.6078(2.7827)
## Equation log_PREALWAPDP       -0.0746(0.0238)**   -0.1448(0.1425)
## Equation log_PREALWAPDNP      0.1576(0.1569)       -0.7516(0.9384)
## Equation log_YDISPREAL       -0.0088(0.0064)       -0.0246(0.0383)
##                               log_PREALWAPDNP -2  log_YDISPREAL -2
## Equation log_QDP              -0.2354(0.5326)      5.1050(13.0520)
## Equation log_PREALWAPDP        0.0185(0.0273)      -0.6844(0.6684)
## Equation log_PREALWAPDNP      0.1521(0.1796)       -1.7708(4.4016)
## Equation log_YDISPREAL         0.0144(0.0073).     0.1624(0.1797)
##                               log_QDP -3  log_PREALWAPDP -3
## Equation log_QDP              0.6822(0.4178)       -2.4916(2.5640)
## Equation log_PREALWAPDP       -0.0543(0.0214)*      0.2848(0.1313)*
## Equation log_PREALWAPDNP      0.0662(0.1409)       -0.3331(0.8647)
## Equation log_YDISPREAL       -0.0082(0.0058)       -0.0257(0.0353)
##                               log_PREALWAPDNP -3  log_YDISPREAL -3
## Equation log_QDP              -1.0656(0.4827)*   -22.5035(13.0477).
## Equation log_PREALWAPDP        0.0716(0.0247)**     0.4249(0.6682)
## Equation log_PREALWAPDNP      0.1933(0.1628)        7.6662(4.4001).
```

```

## Equation log_YDISPREAL      0.0120(0.0066).      -0.6690(0.1796)***
##                               log_QDP -4      log_PREALWAPDP -4
## Equation log_QDP              0.6858(0.3822).      -1.2043(2.5294)
## Equation log_PREALWAPDP      -0.0315(0.0196)      0.1188(0.1295)
## Equation log_PREALWAPDNP     -0.0563(0.1289)      1.3067(0.8530)
## Equation log_YDISPREAL      -0.0063(0.0053)      0.0114(0.0348)
##                               log_PREALWAPDNP -4      log_YDISPREAL -4
## Equation log_QDP             -0.4782(0.4810)      2.7420(14.7493)
## Equation log_PREALWAPDP      0.0192(0.0246)      0.3427(0.7553)
## Equation log_PREALWAPDNP     0.1161(0.1622)      -4.4195(4.9739)
## Equation log_YDISPREAL      0.0040(0.0066)      0.4683(0.2030)*
##                               log_QDP -5      log_PREALWAPDP -5
## Equation log_QDP             0.6104(0.3526).      -2.1442(2.4145)
## Equation log_PREALWAPDP     -0.0198(0.0181)      -0.0097(0.1236)
## Equation log_PREALWAPDNP     0.1164(0.1189)      0.7490(0.8142)
## Equation log_YDISPREAL     -0.0050(0.0049)      0.0139(0.0332)
##                               log_PREALWAPDNP -5      log_YDISPREAL -5
## Equation log_QDP            -1.1575(0.4806)*      8.4217(15.4113)
## Equation log_PREALWAPDP     -0.0294(0.0246)      -0.6263(0.7892)
## Equation log_PREALWAPDNP     0.0564(0.1621)      -6.3966(5.1972)
## Equation log_YDISPREAL     -0.0116(0.0066).      0.0101(0.2121)
##                               log_QDP -6      log_PREALWAPDP -6
## Equation log_QDP            0.2154(0.3186)      -3.0998(2.3458)
## Equation log_PREALWAPDP     -0.0275(0.0163).      -0.0928(0.1201)
## Equation log_PREALWAPDNP     0.2134(0.1074).      -0.2221(0.7911)
## Equation log_YDISPREAL     -0.0043(0.0044)      0.0570(0.0323).
##                               log_PREALWAPDNP -6      log_YDISPREAL -6
## Equation log_QDP            -0.4884(0.4995)      -7.0755(14.3280)
## Equation log_PREALWAPDP      0.0224(0.0256)      0.1192(0.7338)
## Equation log_PREALWAPDNP    -0.1171(0.1684)      4.9165(4.8319)
## Equation log_YDISPREAL      0.0010(0.0069)      0.1489(0.1972)
##                               log_QDP -7      log_PREALWAPDP -7
## Equation log_QDP            -0.0028(0.2846)      -0.8281(2.4960)
## Equation log_PREALWAPDP     -0.0136(0.0146)      -0.1651(0.1278)
## Equation log_PREALWAPDNP     0.1221(0.0960)      0.8960(0.8417)
## Equation log_YDISPREAL     -0.0028(0.0039)      -0.0009(0.0344)
##                               log_PREALWAPDNP -7      log_YDISPREAL -7
## Equation log_QDP            0.7909(0.4494).      -9.3025(12.6131)
## Equation log_PREALWAPDP     0.0218(0.0230)      -0.4428(0.6459)

```

```

## Equation log_PREALWAPDNP -0.3807(0.1516)* -4.6557(4.2536)
## Equation log_YDISPREAL 0.0061(0.0062) -0.0462(0.1736)
## log_QDP -8 log_PREALWAPDP -8
## Equation log_QDP -0.0038(0.2306) -0.9364(2.3362)
## Equation log_PREALWAPDP -0.0157(0.0118) 0.0052(0.1196)
## Equation log_PREALWAPDNP 0.0144(0.0778) 0.1434(0.7878)
## Equation log_YDISPREAL -0.0019(0.0032) -0.0085(0.0322)
## log_PREALWAPDNP -8 log_YDISPREAL -8
## Equation log_QDP -0.5773(0.4488) -2.5085(12.5567)
## Equation log_PREALWAPDP -0.0024(0.0230) 0.4481(0.6431)
## Equation log_PREALWAPDNP -0.0005(0.1514) 0.2502(4.2345)
## Equation log_YDISPREAL -0.0017(0.0062) -0.0958(0.1728)
## log_QDP -9 log_PREALWAPDP -9
## Equation log_QDP -0.1666(0.1414) 5.0666(2.2707)*
## Equation log_PREALWAPDP -0.0160(0.0072)* 0.1370(0.1163)
## Equation log_PREALWAPDNP -0.0262(0.0477) -1.1288(0.7657)
## Equation log_YDISPREAL -0.0030(0.0019) -0.0382(0.0313)
## log_PREALWAPDNP -9 log_YDISPREAL -9
## Equation log_QDP -1.1384(0.4241)** 2.1330(9.8460)
## Equation log_PREALWAPDP 0.0234(0.0217) 0.1289(0.5042)
## Equation log_PREALWAPDNP 0.1289(0.1430) 0.0685(3.3204)
## Equation log_YDISPREAL -0.0039(0.0058) 0.0162(0.1355)

```