Time Series Research Assignment

Jacques Rossouw - 21159793[a]

[a] *Stellenbosch, Western Cape, South Africa*

Abstract

Abstract to be written here.

Email address: `gerardrossouw@gmail.com` (Jacques Rossouw - 21159793)

Table of Contents

1. Part I

2. Introduction

The aim of this paper is to replicate the work by MacDonald & Ricci (2004). They analyse the determinants of the Real Effective Exchange Rate (REER) based on a collection of variables as determinants, using a Vector Error Correction Model. The purpose of their study was to use the most current research to investigate a set of variables with explanatory power in determining the long term behaviour of the REER in South Africa. They used the period starting from the first quarter of 1970 until the first quarter of the year 2002. Towards this end, they made use of the Maximum Likelihood method of estimating the VECM, developed in Johansen (1995). Their choice of variables is based on the most recent research at the time, of the determinants of the real exchange rate in developing economies such as South Africa. This paper serves to replicate the steps taken by MacDonald & Ricci (2004) and reach a conclusion independently, and thus will critically evaluate the logical flow towards estimating the final model. Therefore, additional tests and evaluations are made to reassess the robustness of the final mode.

The choice of the model vector introduced later, is based on several developments prior to 2004. Briefly, the variables have been found to serve as feasible explanatory variables for the REER includes: productivity, real interest rates relative to mostly traded with, the relative openness of the selected economy to trade, and the magnitude of the fiscal balance and net foreign assets (MacDonald & Ricci, 2004).

The Purchasing Power Parity (PPP) points to the equality between the price levels between two countries if they were quoted in the same currencies. When the PPP holds, the real exchange rate must not vary (Sarno & Taylor, 2002). The VECM for the REER is thus an attempt to elucidate the nature of deviations from the PPP. A comprehensive and accurate model for the deviations from the PPP that is explainable by real factors would provide an appropriate framework for policy-makers to respond with ideal policy (Sarno & Taylor, 2002). In light of this, MacDonald & Ricci (2004) attempt to bring various explanatory variables together in a broad model that can appropriately explain these deviations.

Part II: Replication

3. Importing and Cleaning the Data

4. Plotting the Model Variables (STEP 0.)

Figures 4.1 & 4.2 displays each variable as they as they were before their natural logarithm transformation.

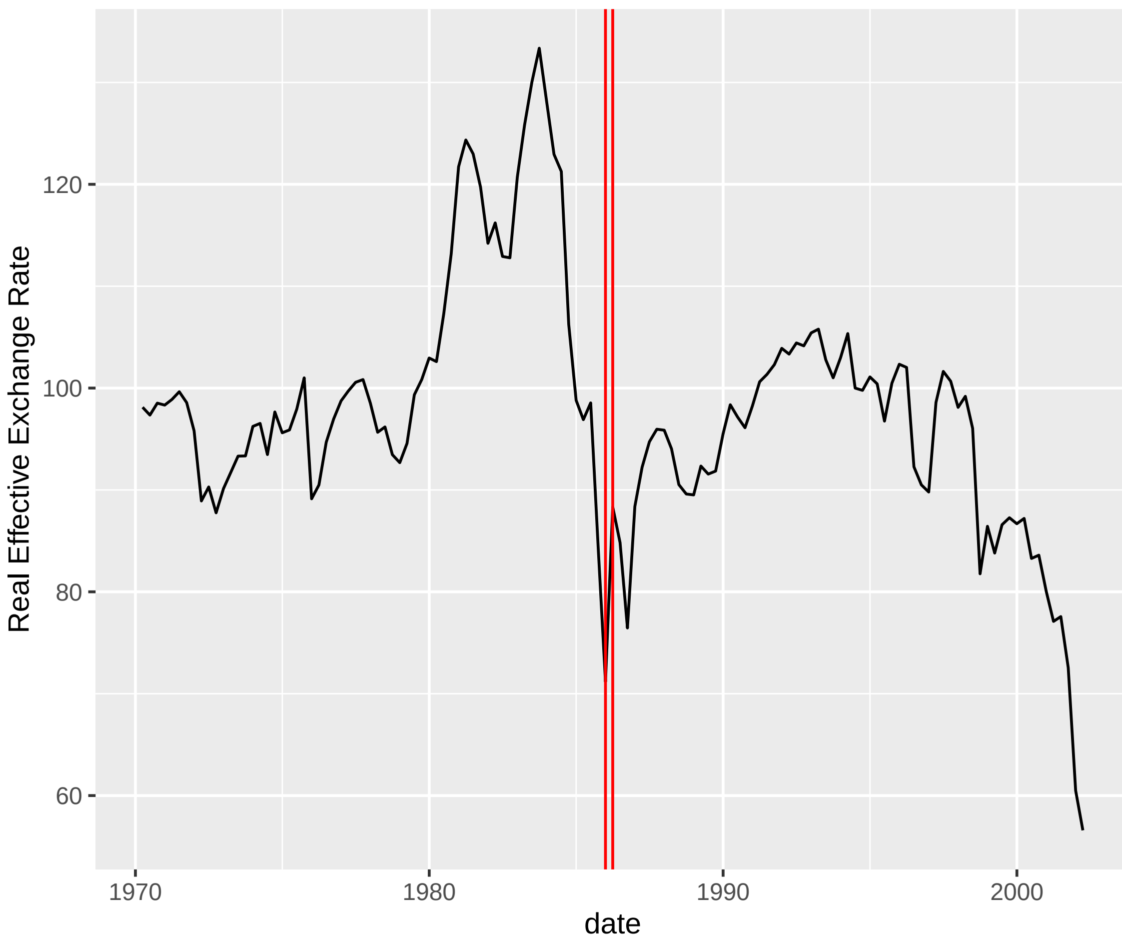

Figure 4.1: The South Africa Real Exchange Rate

might have to rethink this: The regime changes might be accounted for/proxied by including the openness, real commodity prices and net foreign assets??

The red lines indicate the dates at which dummy variables for possible outliers will be included as an alternative model specification. This alternative specification aims to account for the large changes in the real exchange rate regime over the period. There was little consistency in the manner in which the real exchange rate was determined over the period from 1970 to 2002. Most significantly, intermittent changes in the degree and nature of intervention from the South African Reserve Bank (SARB) suggests that the explanatory factors determining the REER will necessary be inconsistent over this period (Aron, Elbadawi & Kahn, 1998). The period from 1979 to 1988 from figure 4.1 for example, is visually unique from the rest of the time series. The coincidence seems unlikely with some research suggesting that intervention during this period was based on maintaining the real price of gold in rand (Aron *et al.*, 1998). This suggests that a model that can account for some structural deviations might be more appropriate. The VECM developed by MacDonald & Ricci (2004) to explain the equilibrium deviations of the REER might achieve more preciseness by accounting for these periods of deviation from free market behaviour.

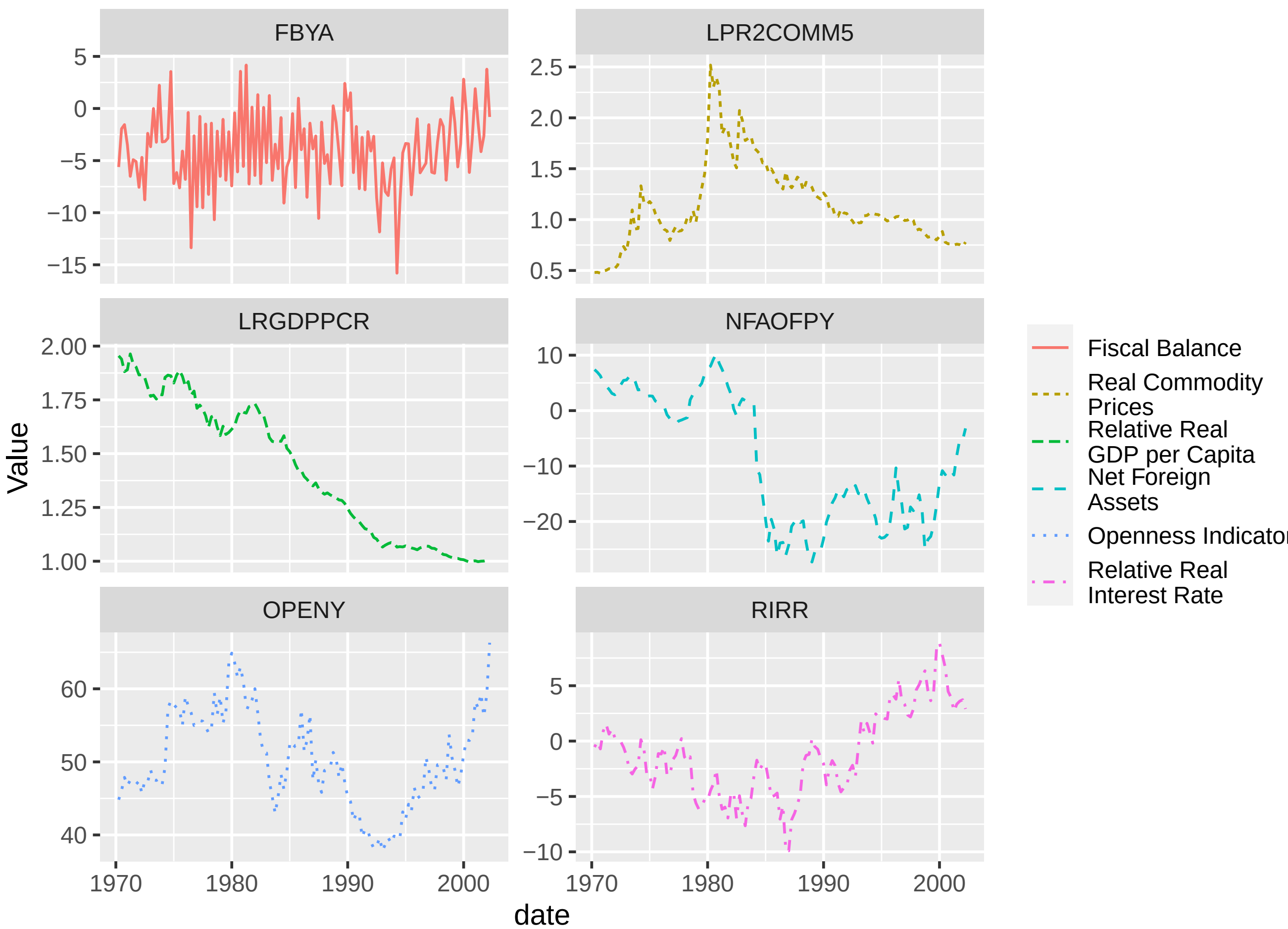

Figure 4.2: Determinants of the Real Exchange Rate for South Africa

Figure 4.2 above might also suggest that there is a possible structural break in the time series around the year 1985. This is mostly evident in the behaviour of the variable `NFAOFPY`, the Net Foreign Assets proxy. The openness indicator, Relative Real Interest Rate, Relative Real GDP, and Real Commodity prices slightly correspond to this theory as well. A test for this is therefore necessary.

To illustrate the co-movement of the system, figure ?? below plots them jointly. Note that the Openness Indicator has also been logged to make the visual comparison easier.

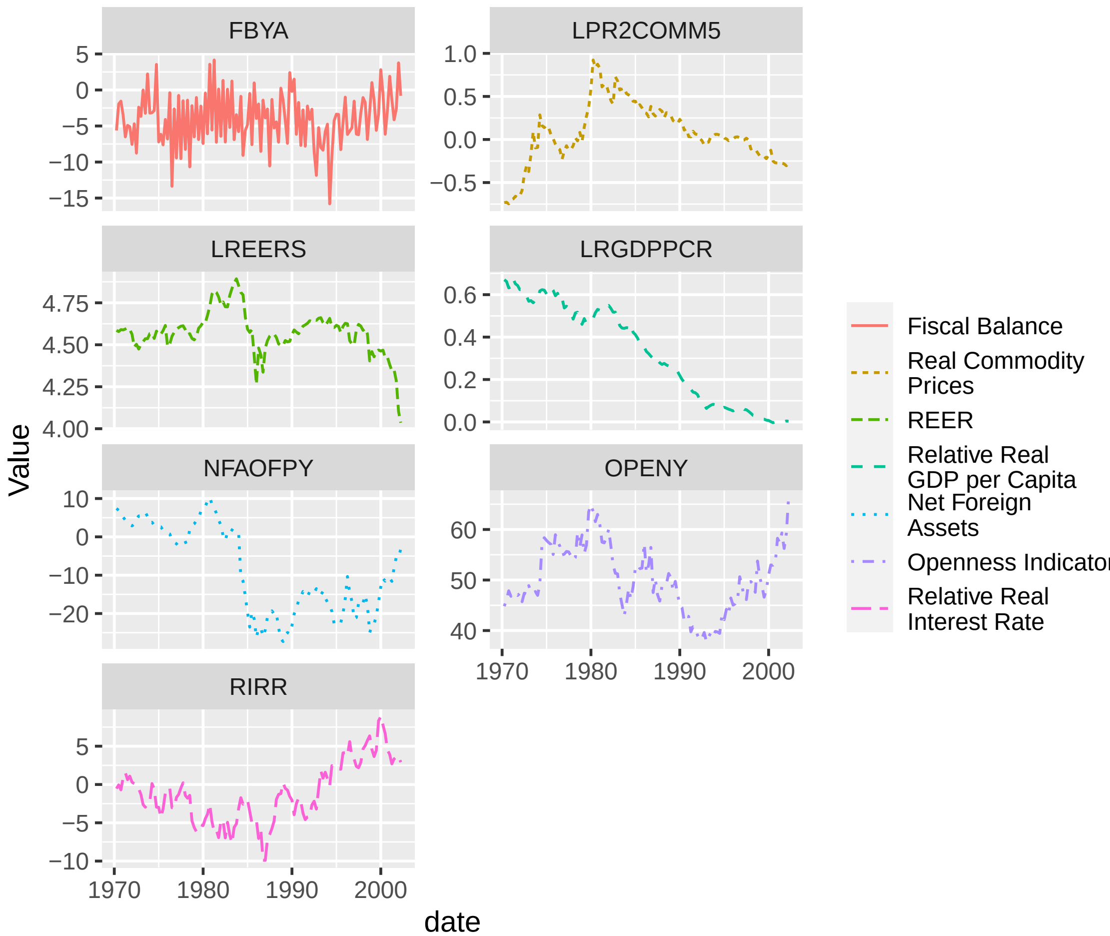

Figure 4.3: The Joint plot of the Model Varibales (Logged where necessary)

A relative degree of co-movement is noticeable in each of these time series altogether. The Log of Real GDP, however, requires one to take into account the declining trend, after which the co-movement seems more evident. Also worth noting, is that none of these series seem to be stationary.

5. The Johansen Method in Theory

explain johansen method

The Johansen (1995) method requires six steps in the estimation process, which follows as:

1. Choosing a specification for the deterministic parts of the model
2. Pre-testing the variables in the system to ensure that they are likely integrated of order one (i.e. $x_t \sim I(1)$$)
3. Estimating the unrestricted Vector Autoregressive model in levels and checking this models adequacy
4. Estimating the VECM form and determining the cointegration rank (i.e. 'r')
5. etc.

6. 1. Plotting the Diffirenced Variables

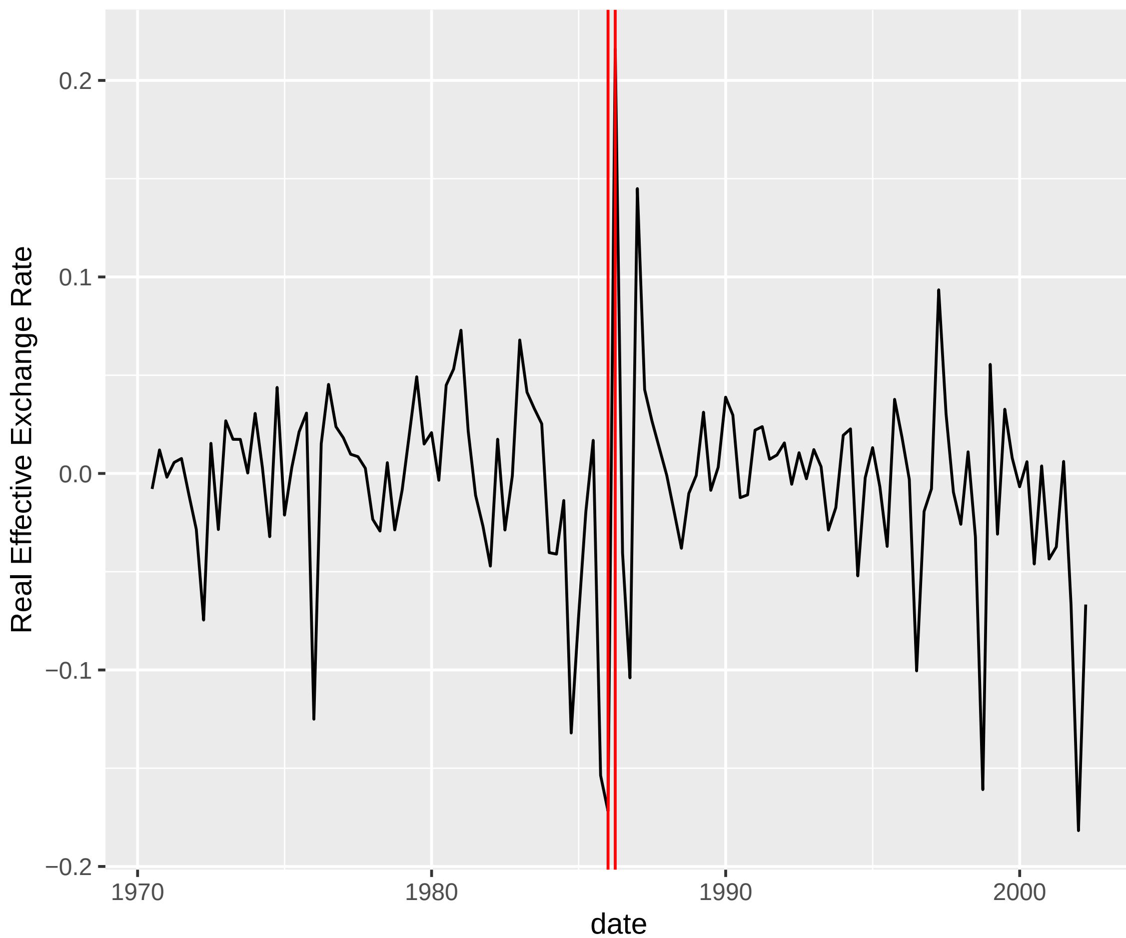

Figure 6.1: The First Difference of the LREERS

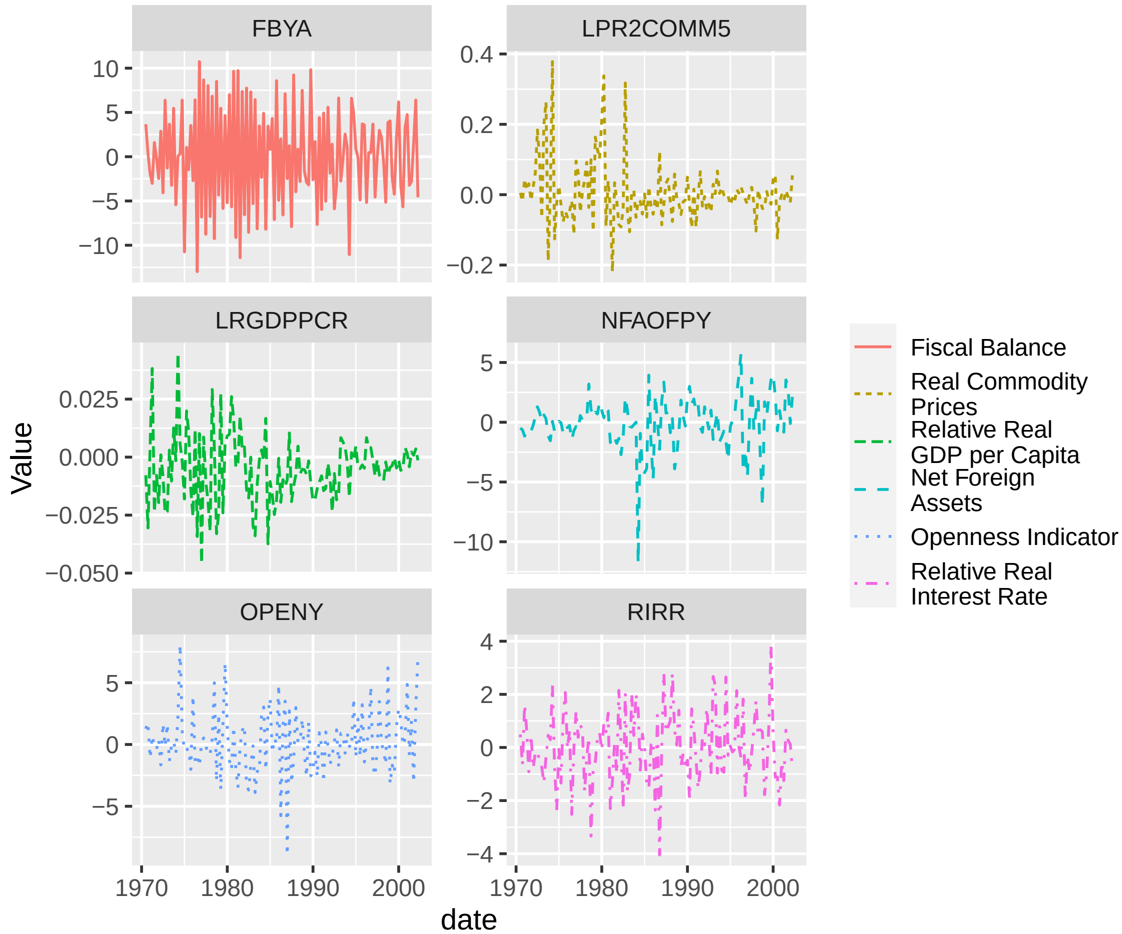

Figure 6.2: The First Difference of the Determinants of the Real Exchange Rate

figures 6.2 and 6.1 above suggest that the REER time series and its candidate explanatory variables are possibly integrated of the first order (i.e. $REER \sim I(1)$). The deterministic component is most likely a constant term. Each of these variables seem somewhat stationary after their first difference is been determined.

The Augmented Dicky-Fuller tests for the differenced time series values of all the variables in the system are displayed in figure **??** below.

Table 6.1: ADF Test on the Model Time Series Variables

	Statistic	Lags	Null_Hypothesis	p_value
RIRR	−2.385	5	stationary	0.417
OPENY	−1.415	5	stationary	0.820
FBYA	−2.785	5	stationary	0.250
NFAOFPY	−0.727	5	stationary	0.966
LREERS	−1.160	5	stationary	0.910
LRGDPPCR	−2.140	5	stationary	0.518
LPR2COMM5	−2.755	5	stationary	0.263

Table 6.2: ADF Test on the Model Time Series Variables

	Statistic	Lags	Null_Hypothesis	p_value
D_RIRR	−5.101	5	stationary	0.010
D_OPENY	−3.767	5	stationary	0.023
D_FBYA	−6.184	5	stationary	0.010
D_NFAOFPY	−4.901	5	stationary	0.010
D_LREERS	−4.718	5	stationary	0.010
D_LRGDPPCR	−3.857	5	stationary	0.018
D_LPR2COMM5	−4.491	5	stationary	0.010

To test this hypothesis, that each of these series are only stationary after the first difference, the Augmented Dickey Fuller test is employed. The Augmented Dicky-Fuller tests for the differenced time series values of all the variables in the system, as well as the first difference of each of these series are displayed in figures **??** and **??** above. The figures confirm that none of the variables are stationary, and that all but `NFAOFPY` i.e. the net foreign assets are stationary after the first difference. For those that are stationary, the null hypothesis was rejected at the 1% significance level, barring the openness indicator for which the p-value is less than 0.05. This might be due to the large deviations created by the apartheid era sanctions (MacDonald & Ricci, 2004). The problem with only using the Dicky-Fuller test in this case is that this test has a lower size when there exists structural breaks in the data. To mitigate this problem the ADF-GLS test is implemented to confirm whether the series are non-stationary before the first difference.

7. 2.A: Estimating the Unrestricted VAR in Levels

```
options(scipen=999)

var_sel <- endog_df %>% dplyr::select(2:8) %>%
    VARselect(., type = "both", lag.max = 8, season = 4)
var_sel$selection
```

```
## AIC(n)  HQ(n)  SC(n) FPE(n)
##      2      1      1      2
```

```
var_sel2 <- endog_df %>% dplyr::select(2:8) %>%
    VARselect(., type = "both", lag.max = 8, season = 4,
          exogen = exog_df[,5:8])
var_sel2$selection
```

```
## AIC(n)  HQ(n)  SC(n) FPE(n)
##      2      1      1      2
```

```
endog_alt <- endog_df %>% mutate(OPENY = log(OPENY))
VARselect(endog_alt[,c(2:8)], type = "both", lag.max = 8, season = 4)$selection
```

```
## AIC(n)  HQ(n)  SC(n) FPE(n)
##      2      1      1      2
```

```
VARselect(endog_alt[,c(2:8)], type = "both", lag.max = 8, season = 4,
          exogen = exog_df[,5:8])$selection
```

```
## AIC(n)  HQ(n)  SC(n) FPE(n)
##      2      1      1      2
```

The model selection based on the Akaike Information Criteria (AIC) suggests that two lags be included in the model. The AIC and Schwartz Criterion (SC) tends to select over-paramterised models however, but the Schwartz Criterion tends to yield asymptotically consistent results compared to the AIC. For the purpose of the replication the AIC will be used. It is worth noting that this might yield more inconsistent results. Based on the all of information criterion displayed above, there seems to be no reason to choose a model with higher lag order than 2. Therefore, using the suggested lag length from MacDonald & Ricci (2004), 4, is in disagreement with the replication above. Many of the variables have already been scaled to span over similar smaller ranges. The only variable that remains relatively large is the openness indicator (`OPENY`). This paper thus continues with an additional variable specification compared to that of MacDonald & Ricci (2004). This alternative includes a natural log standardised value for `OPENY` (i.e. `LOPENY`) instead, and compares this specification with the model obtained from data used in the original paper. The suggested number of lags with this alternative specification results in a choice of two lags to be included in the VECM, according to the AIC. The same problem still exists in terms of the conflict between the AIC and SC.

Adding additional lags is usually an alternative option for when the residuals display serial correlation (Johansen, 1995). However, MacDonald & Ricci (2004) do not explicitly mention this concession and its reasoning. It is important to note that adding additional lags when it is not explicitly suggested so by the information criterion is not recommended in the multivariate case (Johansen, 1995). The tests for serial correlation that follows in the next section provide a likely reason for the decision to include four lags.

Using the results as they are displayed above leads to a series of VAR model with serial correlation exisiting for each of these mentioned above.

Table 7.1: VAR Estimation Results

	Dependent variable:	
	y	
	(1)	(2)
LREERS.l1	1.111***	1.163***
RIRR.l1	−0.003	−0.004
LRGDPPCR.l1	−0.134	0.245
OPENY.l1	0.083	0.001
FBYA.l1	0.003**	0.001
NFAOFPY.l1	−0.004*	−0.008***
LPR2COMM5.l1	0.022	0.075
LREERS.l2	−0.142	−0.262
RIRR.l2	0.001	−0.002
LRGDPPCR.l2	−0.234	0.045
OPENY.l2	−0.046	−0.005*
FBYA.l2	−0.001	−0.001
NFAOFPY.l2	0.004*	0.006
LPR2COMM5.l2	0.007	0.020
LREERS.l4		−0.174
RIRR.l4		0.002
LRGDPPCR.l4		0.249
OPENY.l4		0.003
FBYA.l4		−0.0004
NFAOFPY.l4		−0.004
LPR2COMM5.l4		−0.138**
const	0.273	0.133
trend	−0.002**	
sd1	0.015	0.033*
sd2	0.001	0.018
sd3	0.007	0.002
DUMRER1	0.177***	
DUMRER2	0.197***	
DUMFBYA	0.003	
DUMNFAOFPY	−0.020	
Observations	127	125
R^2	0.905	0.894
Adjusted R^2	0.885	0.859
Residual Std. Error	0.044 (df = 104)	0.049 (df = 93)
F Statistic	45.002*** (df = 22; 104)	25.281*** (df = 31; 93)
Note:		*p<0.1; **p<0.05; ***p<0.01

Tests to perform: 1. Structural break 2. White noise residuals 3. autocorr resid's 4. time-varying params 5. heteroskedasticity 6. test (or check?) for Weak Exog

8. 2.B: Tests on VAR

8.1. I. White Noise Residuals

8.1.1. Plot Residuals

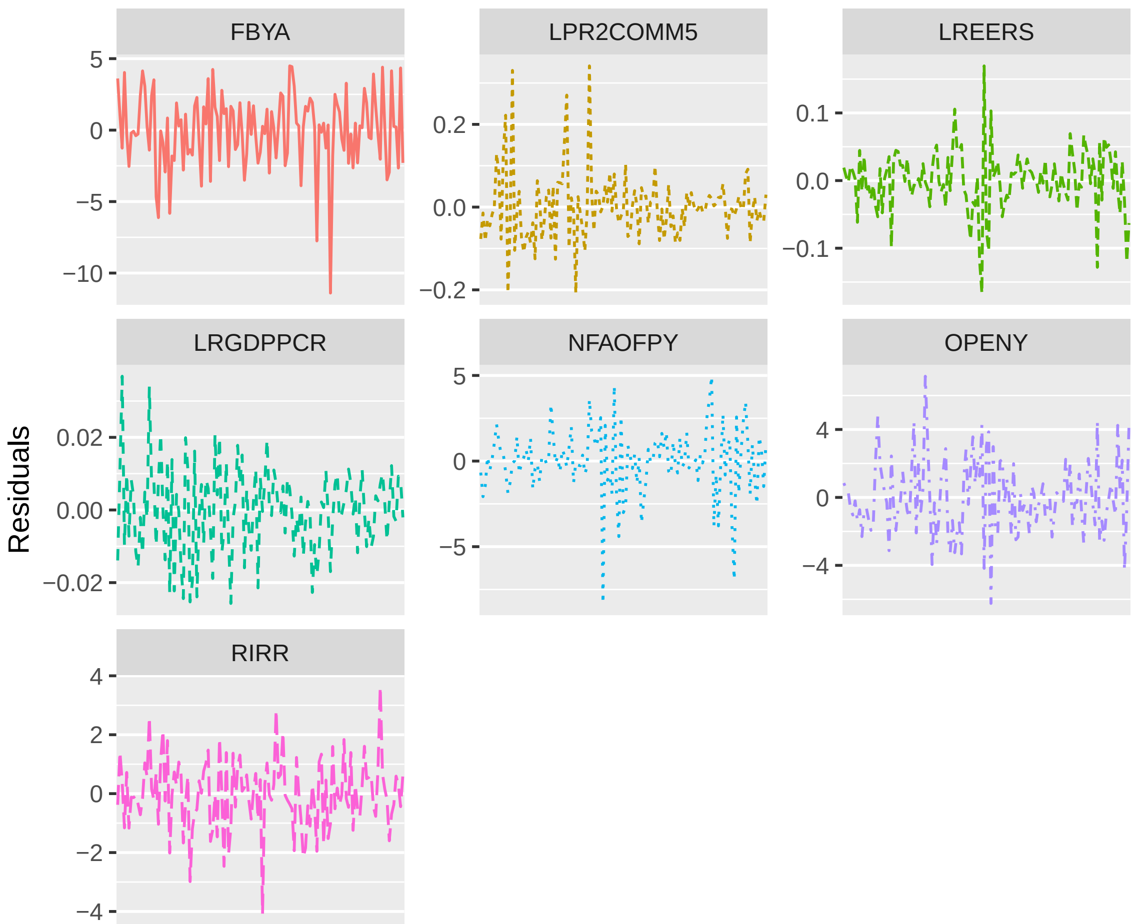

Figure 8.1: VAR Residuals plot - Model 1

The plot of the residuals from the first VAR model in figure 8.1 seems very close to a white noise process, however, there do appear to be deviations from the zero mean that could imply otherwise. The clusters of larger residuals appear to be concentrated around the timeline that corresponds with the outliers identified by MacDonald & Ricci (2004). Alternatively, this could add an additional illustration of the underlying effect of a structural change in the data. The nature of the Johansen (1995) method, however, allows for some non-normality in the errors as MacDonald & Ricci (2004) point out. The asymptotic nature of their method allows for this, however, the assumption of serial correlation is not allowed to be violated (Johansen, 1995).

8.1.2. Serial Correlation Tests:

```
stargazer(rbind(pmt_test, es_tests), header = FALSE, summary = F, title = "Serial Correlation Te
    notes = c("Null-hypotheses: The Residuals Display No Serial Corrrelation", "Note: Both tests
```

Table 8.1: Serial Correlation Tests on VAR Models' Residuals

	Model	DoF1	DoF2	p_value	Statistic	Test
1	VAR_Residuals_Alt	686	N_A	0.0177	766.2365	Portmanteau_(adj)
2	VAR_Residuals_Alt_Outliers	686	N_A	0.0584	745.0479	Portmanteau_(adj)
3	VAR_Residuals_Auto	686	N_A	0.0177	766.2365	Portmanteau_(adj)
4	VAR_Residuals_Auto_Outliers	686	N_A	0.0584	745.0479	Portmanteau_(adj)
5	VAR_Residuals_Forced	588	N_A	0.0149	665.0041	Portmanteau_(adj)
6	VAR_Residuals_Forced_Outliers	588	N_A	0.0149	665.0041	Portmanteau_(adj)
7	VAR_Residuals_Alt	245	445	0.1761	1.1085	Edgerton-Shukur_F
8	VAR_Residuals_Alt_Outliers	245	445	0.1761	1.1085	Edgerton-Shukur_F
9	VAR_Residuals_Auto	245	473	0.0307	1.2274	Edgerton-Shukur_F
10	VAR_Residuals_Auto_Outliers	245	445	0.1795	1.1069	Edgerton-Shukur_F
11	VAR_Residuals_Forced	245	370	0.0986	1.1603	Edgerton-Shukur_F
12	VAR_Residuals_Forced_Outliers	245	370	0.0986	1.1603	Edgerton-Shukur_F

Null-hypotheses: The Residuals Display No Serial Corrrelation
Note: Both tests adjust for small sample bias

Null: No Serial corr $\therefore$ low p-val => reject Null => There exists serial corr

Recall the hypothesis from section 7, the inclusion of additional lags up to 4, is most likely towards solving the problem of serial correlation. From table 8.1, the Models with the suffix `Forced` or `Forced_Outliers` are the VAR models with four lags, the replication of the model from MacDonald & Ricci (2004). The latter includes the added dummies for the outlier observations. The `Alt` models are those with `LOPENY`, i.e. the openness indicator has been scaled. The `Auto` models are those selected by automating the VAR selection process based on the AIC. The `alt` model was also selected in this way. The meaning of `Outliers` remains consistent. Table 8.1 above displays the results from the 'Adjusted Portmanteau Test' and the Edgerton & Shukur (1999) F-test. Both of these tests incorporate small sample corrections that are necessary in the case of this dataset (Lütkepohl, 2005). The results seem relatively inconsistent in this case. The Portmanteau Test results in p-values less than 0.05 for the `alt`, and `auto` specifications with outliers. Therefore, the only models that are not rejected for having serial correlated errors is the alternative specification with outlier dummies, and the model from the original paper with 2 lags instead of 4, and outlier dummies. There exists some concerns about whether this test is accurate for these types of models Lütkepohl (2005). The p-values from the Edgerton & Shukur (1999) F-test imply that only the `Auto` specification without outliers is a suboptimal specification with serial correlation. The `Forced` specifications, however, have p-values equal to 0.0986 such that their serial correlation cannot be rejected at the 10% confidence level. This marginal conclusion of uncorrelated residuals is concerning when the model is already not the ideal specification according to the SC and AIC. The `Alt` specification has a more confident rejection of serial correlation with p-values equal to 0.1761 with and without dummies for outliers.

8.1.3. ACF of Residuals Plot

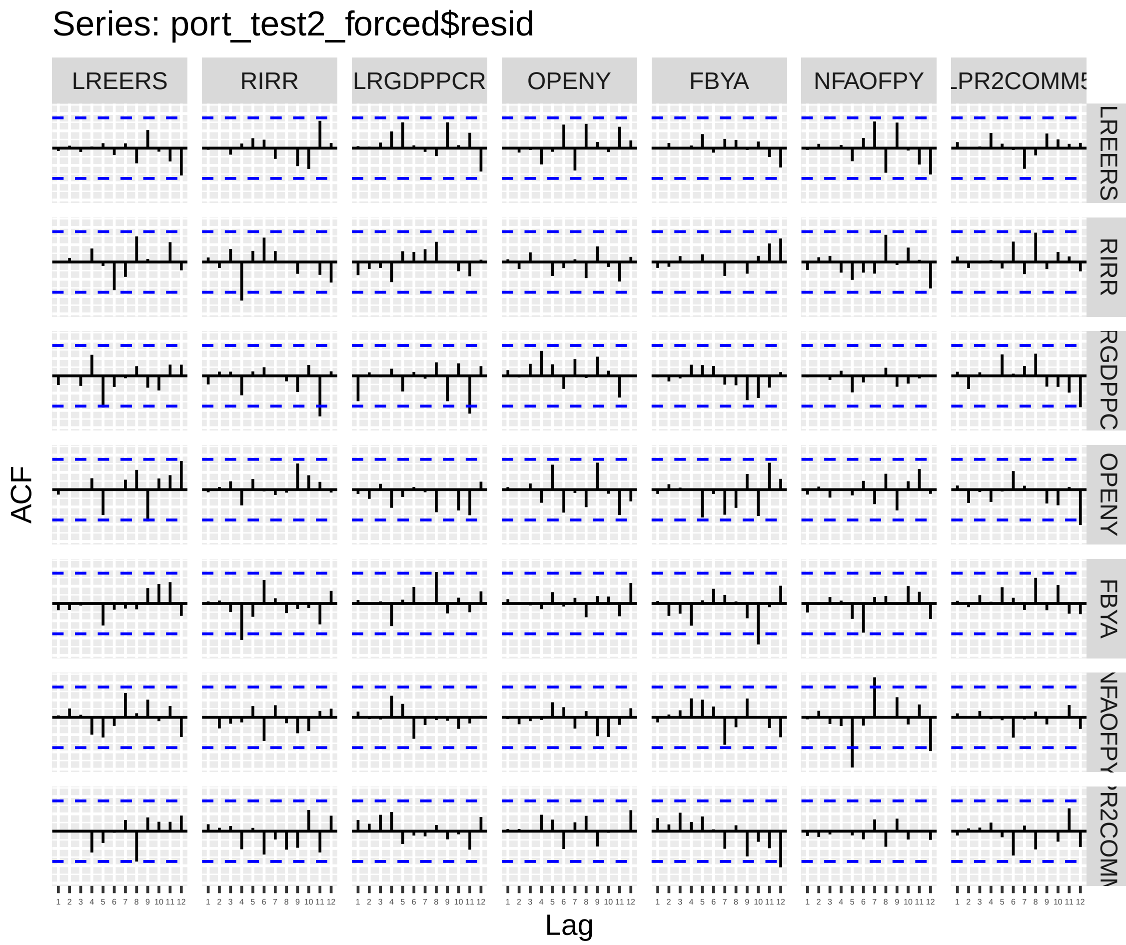

Figure 8.2: Plot of the ACF of the `Forced` Model with OutlierDummies

Figure 8.2 above visually illustrates the deviation from serial correlated errors. `LREERS`, the log of the REER, and the net foreign assets variable seems to display the largest deviations.

8.2. II Check for the presence of Structural Break

9. 3. Determining the order of CI

	test	10pct	5pct	1pct
r <= 6 |	6.34	7.52	9.24	12.97
r <= 5 |	15.57	17.85	19.96	24.60
r <= 4 |	32.34	32.00	34.91	41.07
r <= 3 |	50.67	49.65	53.12	60.16
r <= 2 |	79.25	71.86	76.07	84.45
r <= 1 |	124.66	97.18	102.14	111.01
r = 0 |	174.12	126.58	131.70	143.09

Figure 9.1: Johansen Trace test on alternative model (no outlier dummy)

	test	10pct	5pct	1pct
r <= 6 |	4.87	7.52	9.24	12.97
r <= 5 |	11.56	17.85	19.96	24.60
r <= 4 |	26.61	32.00	34.91	41.07
r <= 3 |	42.91	49.65	53.12	60.16
r <= 2 |	68.68	71.86	76.07	84.45
r <= 1 |	115.27	97.18	102.14	111.01
r = 0 |	170.63	126.58	131.70	143.09

Figure 9.2: Johansen Trace test on alternative model (with outlier dummy)

	test	10pct	5pct	1pct
r <= 6 |	6.07	7.52	9.24	12.97
r <= 5 |	14.23	17.85	19.96	24.60
r <= 4 |	28.95	32.00	34.91	41.07
r <= 3 |	44.97	49.65	53.12	60.16
r <= 2 |	72.83	71.86	76.07	84.45
r <= 1 |	101.60	97.18	102.14	111.01
r = 0 |	141.65	126.58	131.70	143.09

Figure 9.3: Johansen Trace test on Forced model (no outlier dummy)

	test	10pct	5pct	1pct
r <= 6 |	4.93	7.52	9.24	12.97
r <= 5 |	11.51	17.85	19.96	24.60
r <= 4 |	25.02	32.00	34.91	41.07
r <= 3 |	43.44	49.65	53.12	60.16
r <= 2 |	66.41	71.86	76.07	84.45
r <= 1 |	94.47	97.18	102.14	111.01
r = 0 |	143.54	126.58	131.70	143.09

Figure 9.4: Johansen Trace test on Forced model (with outlier dummy)

	test	10pct	5pct	1pct
r <= 6	4.80	7.52	9.24	12.97
r <= 5	12.06	17.85	19.96	24.60
r <= 4	26.61	32.00	34.91	41.07
r <= 3	43.97	49.65	53.12	60.16
r <= 2	69.41	71.86	76.07	84.45
r <= 1	115.68	97.18	102.14	111.01
r = 0	171.65	126.58	131.70	143.09

Figure 9.5: Johansen Trace test on auto model (with outlier dummy)

According to the above tables 9.5, 9.2, 9.3, 9.4, and 9.5 the results from the Johansen (1995) test suggests, at the 5% significance, 3, 3, 1, 1, and 2, cointegration vectors existing for the Alternative without outliers, with outliers, the forced with, and without, and the automated model respectively. Therefore, the replication of the result from MacDonald & Ricci (2004) is unsuccessful, however, this should still be a reasonable representation of their model. At the 1% significance, the cointegration relations were 2, 2, 0, 1, and 2, in the same order.

10. 4. Estimating VECM with Restriction

Table 10.1: Normalised Cointegration Relation Vectors

Statistic	N	Mean	St. Dev.	Min	Max
Alt	8	−0.621	2.168	−5.890	1.000
Alt_Outliers	8	−0.693	2.357	−6.420	1.000
Forced	8	−0.576	1.794	−4.890	1.000
Forced_Outliers	8	−0.571	1.767	−4.810	1.000
Auto_Outliers	8	−0.566	1.823	−4.960	1.000

The cointegration vector for the forced models is very similar to that demonstrated in the original paper with the exception of the coefficient on the log of the relative real GDP per capita.

```
rmod1beta <- rmod_alt$
rmod2beta <- rmod2_alt$beta
rmod3beta <- rmod_forced$beta
rmod4beta <- rmod2_forced$beta
rmod5beta <- rmod2_auto$beta

rtest_2alt <- endog_alt %>% dplyr::select(2:8) %>%
    ca.jo(x = ., type = "eigen",
                     ecdet = "const",
                     spec = "transitory", K = 2,
                     dumvar = exog_df[,2:8])
# alrtest(rtest_alt, r = 2, A = A)
```

11. 5. alrtest() for test of Weak exogeneity

```
   A <-  matrix(byrow = TRUE,
                c(1, rep(0, 6)), 1,8)
A
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
## [1,]    1    0    0    0    0    0    0    1
```

12. Conclusion

The next crucial step is to apply the `alrtest()` to the forced model (and every alternative specification) to test for the presence of weak exogeneity in the model. Du Plessis (2005) mentions some concerns regarding the weak exogeneity assumption that MacDonald & Ricci (2004) fails to address. An alternative model is likely better suited for this approach. The next step might be to try some non-paramteric model, or by referring to the economic theory again. The problem with extending to additional lags when a simpler model is not well suited is a loss of bias. The best alternatives suggests obtaining better data. To this end, the solution most likely lies in improving or expanding the search within economic reasoning. A model is after all, not very useful if it does not make sense.

13. Directly from (MacDonald & Ricci, 2004)

13.1. Section 3 Data and Methodology

- Plotting the exp(LREER) and the rest of the variables
- Showcase the vector of interest
- Investigate LR CI Rel's amongst var's in vector
  - Method: MLE of Johansen (1995)
  - Why? Corrects for Autocorr and ednog parametrically using VECM specif.
  - Key Advantage: the estimated coefficient - the β vector - can be used to provide a measure of the equilibrium real exchange rate and therefore a quantification of the gap between the prevailing real exchange rate and its equilibrium level. The methodology also derives estimates of the speed at which the real exchange rates converges to the equilibrium level.

14. References

10 Aron, J., Elbadawi, I. & Kahn, B. 1998. Determinants of the real exchange rate in South Africa. (1997-16). [Online], Available: https://ideas.repec.org/p/csa/wpaper/1997-16.html [2022, June 28].

Du Plessis, S.A. 2005. Exogeneity in a recent exchange rate model: A response to MacDonald and ricci. *South African Journal of Economics*. 73(4):741–746.

Edgerton, D. & Shukur, G. 1999. Testing autocorrelation in a system perspective testing autocorrelation. *Econometric Reviews*. 18(4):343–386.

Johansen, S. 1995. *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford University Press. [Online], Available: https://EconPapers.repec.org/RePEc:oxp:obooks:9780198774501.

Lütkepohl, H. 2005. *New introduction to multiple time series analysis*. New York :, Springer.

MacDonald, R. & Ricci, L.A. 2004. Estimation of the equilibrium real exchange rate for south africa 1. *South African Journal of Economics*. 72(2):282–304.

Sarno, L. & Taylor, M.P. 2002. Purchasing Power Parity and the Real Exchange Rate. *IMF Staff Papers*. 49(1):65–105. [Online], Available: https://www.jstor.org/stable/3872492 [2022, June 27].

Uyanto, S.S. 2020. Power comparisons of five most commonly used autocorrelation tests. *Pakistan Journal of Statistics and Operation Research*. 119–130.

Appendix

```
full_df <- fetch_full() %>% full_clean()
# p1 <- plot_endog1(full_df)
# p2 <- plot_endog2(df = full_df)
plot_endog1(full_df)
# grid.arrange(p1, p2, nrow=2)
plot_endog2(df = full_df)
plot_endog_joint(df = full_df)
diff_df <- full_df %>% mutate(across(names(.)[2:13], function(x) x-lag(x))) %>%
    filter(date > first(date))
plot_endog1_log(diff_df)
plot_endog2_log(df = diff_df)
# adf_temp <- urca::ur.df(full_df[,"LRGDPPCR"], type = "trend", lags = 8, selectlags = "BIC")
# Better format this later
joint_adf_test <- joint_adf(full_df)
# knitr::kable(joint_adf_test)
joint_adf_test
# Better format this later
joint_adf_diff_test <- joint_adf_diff(full_df)
# knitr::kable(joint_adf_test)
joint_adf_diff_test
stargazer(
    joint_adf_test,
    joint_adf_diff_test,
          title = "ADF Test on the Model Time Series Variables",
          align = T,
          # column.sep.width = "1pt",
          # font.size = "tiny",
          header = F, summary = F)

endog_df <- split_data_endog(full_df)
exog_df <- split_data_exog(full_df)
endog_df; exog_df
options(scipen=999)

var_sel <- endog_df %>% dplyr::select(2:8) %>%
```

```
    VARselect(., type = "both", lag.max = 8, season = 4)
var_sel$selection

var_sel2 <- endog_df %>% dplyr::select(2:8) %>%
    VARselect(., type = "both", lag.max = 8, season = 4,
          exogen = exog_df[,5:8])
var_sel2$selection

endog_alt <- endog_df %>% mutate(OPENY = log(OPENY))
VARselect(endog_alt[,c(2:8)], type = "both", lag.max = 8, season = 4)$selection
VARselect(endog_alt[,c(2:8)], type = "both", lag.max = 8, season = 4,
          exogen = exog_df[,5:8])$selection



var_aic_alt <- endog_alt %>% dplyr::select(2:8) %>%
    VAR(., type = "both", lag.max = 8,
              ic = "AIC", season = 4)

var_aic2_alt <- endog_alt %>% dplyr::select(2:8) %>%
    VAR(., type = "both", season = 4,
             exogen = exog_df[,5:8],
             lag.max = 8, ic = "AIC")

var_aic <- full_df %>% dplyr::select(2:8) %>%
    VAR(., type = "both", lag.max = 8,
              ic = "AIC", season = 4)

var_aic2 <- full_df %>% dplyr::select(2:8) %>% VAR(., type = "both", season = 4,
                                             exogen = exog_df[,5:8],
                                             lag.max = 8, ic = "AIC")

var_aic_forced <- VAR(endog_df[,c(2:8)], type = "const", p = 4,
              ic = "AIC", season = 4)

var_aic2_forced <- VAR(endog_df[,c(2:8)], type = "const", p = 4,
              ic = "AIC", season = 4)

# var_aic3 <- endog_df %>% dplyr::select(2:7 | 9) %>% VAR(., type = "const", season = 4,
```

```
#                                                     exogen = exog_df[,5:8],
#                                                     lag.max = 8, ic = "AIC")
#
# var_aic4 <- endog_df %>% dplyr::select(2:7 | 10) %>% VAR(., type = "const", season = 4,
#                                                     exogen = exog_df[,5:8],
#                                                     lag.max = 8, ic = "AIC")
#
# var_aic5 <- endog_df %>% dplyr::select(2:7 | 11) %>% VAR(., type = "const", season = 4,
#                                                     exogen = exog_df[,5:8],
#                                                     lag.max = 8, ic = "AIC")
#
# var_aic6 <- endog_df %>% dplyr::select(2:7 | 12) %>% VAR(., type = "const", season = 4,
#                                                     exogen = exog_df[,5:8],
#                                                     lag.max = 8, ic = "AIC")
#
# var_aic7 <- endog_df %>% dplyr::select(2:7 | 13) %>% VAR(., type = "const", season = 4,
#                                                     exogen = exog_df[,5:8],
#                                                     lag.max = 8, ic = "AIC")

summary(var_aic_forced) #var_aic2_forced$varresult
summary(var_aic2, digits = max(3, getOption("digits") - 4))

options(scipen=0)
stargazer(#var_aic2$varresult$LREERS,
          var_aic2_alt$varresult$LREERS,
          var_aic2_forced$varresult$LREERS,
          title = "VAR Estimation Results",
          align = T,
          column.sep.width = "1pt",
          font.size = "tiny",
          report = "vc*", header = F, summary = F,
          omit = ".*[.][1][3]")
as_tibble(resid(var_aic)) %>% gather(vary, res) %>%
    dplyr::summarise(index = rep(c(1:127), 7), vary, res) %>%
    group_by(vary) %>%
    ggplot() +
    geom_line(aes(x = index, y = res, color = vary, linetype = vary)) +
    facet_wrap(~vary, ncol = 3, scales = "free_y") +
```

```
    guides(color="none", linetype="none") +
    scale_y_continuous("Residuals") +
    scale_x_discrete(element_blank())
    # element_text()

# lubridate::
# a1 <- checkresiduals(var_aic$varresult$LREERS, test = "LB", plot = F)
# a2 <- checkresiduals(var_aic$varresult$RIRR, test = "LB", plot = F)
# a3 <- checkresiduals(var_aic$varresult$LRGDPPCR, test = "LB", plot = F)
# a4 <- checkresiduals(var_aic$varresult$OPENY, test = "LB", plot = F)
# a5 <- checkresiduals(var_aic$varresult$FBYA, test = "LB", plot = F)
# a6 <- checkresiduals(var_aic$varresult$NFAOFPY, test = "LB", plot = F)
# a7 <- checkresiduals(var_aic$varresult$LPR2COMM5, test = "LB", plot = F)
#
# lb_test <- cbind(a1, a2, a3, a4, a5 , a6, a7)
#
# lb_test <- data.frame(matrix(unlist(lb_test), nrow=7, byrow=T),
#                        row.names = c( "LREERS_Residual",
#                                       "RIRR_Residual",
#                                       "LRGDPPCR_Residual",
#                                       "OPENY_Residual",
#                                       "FBYA_Residual",
#                                       "NFAOFPY_Residual",
#                                       "LPR2COMM5_Residual")) %>%
#     dplyr::select(-c(X2, X4, X5)) %>%
#     rename("Statistic" = "X1", "p-value" = "X3") %>%
#     mutate(across(names(.), function(x) as.double(x))) %>%
#     mutate(across(names(.), function(x) round(x, 9))) %>%
#     mutate(Statistic = round(Statistic, 2))
# null hypothesis of Portmanteau Test: No serial correlation
es_test <- serial.test(var_aic, type = "ES")
es_test2 <- serial.test(var_aic2, type = "ES")
es_test_forced <- serial.test(var_aic2_forced, type = "ES")
es_test2_forced <- serial.test(var_aic2_forced, type = "ES")
es_test_alt <- serial.test(var_aic2_alt, type = "ES")
es_test2_alt <- serial.test(var_aic2_alt, type = "ES")
tests <- cbind(es_test$serial, es_test2$serial,
              es_test_forced$serial,
```

```
              es_test2_forced$serial,
              es_test_alt$serial,
              es_test2_alt$serial)
unlist(es_test$serial)

es_tests <- data.frame(matrix(unlist(tests), nrow=6, byrow=F),
                      row.names = c("Statistic",
                                    "DoF1",
                                    "DoF2",
                                    "p_value",
                                    "Test",
                                    "Model"))
es_tests <- es_tests %>%
    # Sensitive to change!!!!!
    filter(c(rep(T, 5), F)) %>%
    mutate(across(names(.), function(x) as.double(x))) %>%
    mutate(across(names(.), function(x) round(x, 4)))
es_tests["Test",] = "Edgerton-Shukur_F"
names(es_tests)[1] = "VAR_Residuals_Auto"
names(es_tests)[2] = "VAR_Residuals_Auto_Outliers"
names(es_tests)[3] = "VAR_Residuals_Forced"
names(es_tests)[4] = "VAR_Residuals_Forced_Outliers"
names(es_tests)[5] = "VAR_Residuals_Alt"
names(es_tests)[6] = "VAR_Residuals_Alt_Outliers"

es_tests <- es_tests %>%
    rownames_to_column() %>%
    gather(Model, value, -rowname) %>%
    spread(rowname, value)
# null hypothesis of Portmanteau Test: No serial correlation
port_test <- serial.test(var_aic, type = "PT.adjusted")
port_test2 <- serial.test(var_aic2, type = "PT.adjusted")
port_test_forced <- serial.test(var_aic2_forced, type = "PT.adjusted")
port_test2_forced <- serial.test(var_aic2_forced, type = "PT.adjusted")
port_test_alt <- serial.test(var_aic, type = "PT.adjusted")
port_test2_alt <- serial.test(var_aic2, type = "PT.adjusted")

test <- cbind(port_test$serial, port_test2$serial,
```

```
              port_test_forced$serial,
              port_test2_forced$serial,
              port_test_alt$serial,
              port_test2_alt$serial)

unlist(port_test$serial)

pmt_test <- data.frame(matrix(unlist(test), nrow=5, byrow=F),
                      row.names = c("Statistic",
                                    "DoF1",
                                    "p_value",
                                    "Test",
                                    "Model"))

pmt_test <- pmt_test %>%
    # Sensitive to change!!!!!
    filter(c(rep(T, 4), F)) %>%
    mutate(across(names(.), function(x) as.double(x))) %>%
    mutate(across(names(.), function(x) round(x, 4)))
pmt_test["DoF2",] = "N_A"
names(pmt_test)[1] = "VAR_Residuals_Auto"
names(pmt_test)[2] = "VAR_Residuals_Auto_Outliers"
names(pmt_test)[3] = "VAR_Residuals_Forced"
names(pmt_test)[4] = "VAR_Residuals_Forced_Outliers"
names(pmt_test)[5] = "VAR_Residuals_Alt"
names(pmt_test)[6] = "VAR_Residuals_Alt_Outliers"

pmt_test["Test",] = "Portmanteau_(adj)"
pmt_test <- pmt_test %>%
    rownames_to_column() %>%
    gather(Model, value, -rowname) %>%
    spread(rowname, value)
stargazer(rbind(pmt_test, es_tests),
          header = FALSE, summary = F,
          title = "Serial Correlation Tests on VAR Models' Residuals \\label{serial}",
          notes = c("Null-hypotheses: The Residuals Display No Serial Corrrelation",
                    "Note: Both tests adjust for small sample bias"))
ggAcf(port_test2_forced$resid) +
```

```
    theme(axis.text.y = element_blank(),
          axis.ticks.y = element_blank(),
          axis.text.x = element_text(size = 3))


jo_alt <- endog_alt %>% dplyr::select(2:8) %>%
    ca.jo(., type = "trace",
             ecdet = "const",
             spec = "transitory",
             K = 2, season = 4)
jo2_alt <- endog_alt %>% dplyr::select(2:8) %>%
    ca.jo(., type = "trace",
             ecdet = "const",
             spec = "transitory", K = 2,
             season = 4,
             dumvar = exog_df[,5:8])


jo2 <- endog_df %>% dplyr::select(2:8) %>%
    ca.jo(., type = "trace",
             ecdet = "const",
             spec = "transitory", K = 2,
             season = 4,
             dumvar = exog_df[,5:8])


jo_forced <- full_df %>% dplyr::select(2:8) %>%
    ca.jo(., type = "trace", ecdet = "const",
                   spec = "transitory",
                    K = 4, season = 4)
jo2_forced <- full_df %>% dplyr::select(2:8) %>%
    ca.jo(., type = "trace",
             ecdet = "const",
             spec = "transitory",
             K = 4,
             season = 4,
             dumvar = exog_df[,5:8])
```

```
summary(jo_alt)
summary(jo2_alt)

summary(jo_forced)
summary(jo2_forced)

summary(jo2)



# Values of teststatistic and critical values of test (jo2_alt):
#



knitr::include_graphics('./images/jo_alt.png')
knitr::include_graphics('./images/jo2_alt.png')
knitr::include_graphics('./images/jo_forced.png')
knitr::include_graphics('./images/jo2_forced.png')
knitr::include_graphics('./images/jo2.png')
options(scipen = 999)

rmod_alt <- endog_alt %>% dplyr::select(2:8) %>%
    ca.jo(x = ., type = "trace",
                     ecdet = "const",
                     spec = "transitory", K = 2,
                     dumvar = exog_df[,2:4]) %>%
    urca::cajorls(.)

rmod2_alt <- endog_alt %>% dplyr::select(2:8) %>%
    ca.jo(x = ., type = "trace",
                     ecdet = "const",
                     spec = "transitory", K = 2,
                     dumvar = exog_df[,2:8]) %>%
    urca::cajorls(.)

rmod_forced <- full_df %>% dplyr::select(2:8) %>%
    ca.jo(x = ., type = "trace",
                     ecdet = "const",
                     spec = "transitory", K = 4,
```

```
                     dumvar = exog_df[,2:4]) %>%
    urca::cajorls(.)

rmod2_forced <- full_df %>% dplyr::select(2:8) %>%
    ca.jo(x = ., type = "trace",
                     ecdet = "const",
                     spec = "transitory", K = 4,
                     dumvar = exog_df[,2:8]) %>%
    urca::cajorls(.)

rmod2_auto <- full_df %>% dplyr::select(2:8) %>%
    ca.jo(x = ., type = "trace",
                     ecdet = "const",
                     spec = "transitory", K = 2,
                     dumvar = exog_df[,2:8]) %>%
    urca::cajorls(.)

rmod1beta <- rmod_alt$beta
rmod2beta <- rmod2_alt$beta
rmod3beta <- rmod_forced$beta
rmod4beta <- rmod2_forced$beta
rmod5beta <- rmod2_auto$beta

rmod_joint <- cbind(rmod1beta,
                    rmod2beta,
                    rmod3beta,
                    rmod4beta,
                    rmod5beta)

rmod_out <- data.frame(matrix(unlist(rmod_joint), nrow=8, byrow=F),
                      row.names = c("LREERS.l1",
                                    "RIRR.l1",
                                    "LRGDPPCR.l1",
                                    "OPENY.l1",
                                    "FBYA.l1",
                                    "NFAOFPY.l1",
                                    "LPR2COMM5.l1",
                                    "constant"))
```

```
names(rmod_out) = c("Alt", "Alt_Outliers", "Forced", "Forced_Outliers",
                    "Auto_Outliers")
rmod_out <- rmod_out %>%
    mutate(across(names(.), function(x) round(x, 2)))
options(scipen = 999)
stargazer(rmod_out, header = F,
          style = "aer",
          title = "Normalised Cointegration Relation Vectors",
          # column.labels = c,
          add.lines = list(c("Short-term impact of NFAOFPY on REERS ", 0)),
          font.size = "small")

rmod1beta <- rmod_alt$
rmod2beta <- rmod2_alt$beta
rmod3beta <- rmod_forced$beta
rmod4beta <- rmod2_forced$beta
rmod5beta <- rmod2_auto$beta

rtest_2alt <- endog_alt %>% dplyr::select(2:8) %>%
    ca.jo(x = ., type = "eigen",
                     ecdet = "const",
                     spec = "transitory", K = 2,
                     dumvar = exog_df[,2:8])
# alrtest(rtest_alt, r = 2, A = A)

   A <-  matrix(byrow = TRUE,
                c(1, rep(0, 6)), 1,8)
A
```

```
joint_adf_diff
```

```
## function (df)
## {
##     df <- df %>% dplyr::select(1:8)
##     logs <- df %>% dplyr::select("date" | starts_with("L"))
##     non_logs <- df %>% dplyr::select(date | !starts_with("L"))
##     adf_logs <- logs %>% mutate(across(names(.)[2:4], function(x) x -
```

```
##         lag(x))) %>% filter(date > first(date)) %>% dplyr::select(-date) %>%
##         ts(.) %>% sapply(., function(x) adf.test(x))
##     adf_non_logs <- non_logs %>% mutate(across(names(.)[2:5],
##         function(x) x - lag(x))) %>% filter(date > first(date)) %>%
##         dplyr::select(-date) %>% ts(.) %>% sapply(., function(x) adf.test(x))
##     df2 <- data.frame(matrix(unlist(adf_logs), nrow = 3, byrow = T),
##         row.names = c("D_LREERS", "D_LRGDPPCR", "D_LPR2COMM5")) %>%
##         dplyr::select(-c(X5, X6)) %>% rename(Statistic = X1,
##         Lags = X2, Null_Hypothesis = X3, p_value = X4) %>% mutate(across(names(.)[-3],
##         function(x) as.double(x))) %>% mutate(across(names(.)[-3],
##         function(x) round(x, 4)))
##     df3 <- data.frame(matrix(unlist(adf_non_logs), nrow = 4,
##         byrow = T), row.names = c("D_RIRR", "D_OPENY", "D_FBYA",
##         "D_NFAOFPY")) %>% dplyr::select(-c(X5, X6)) %>% rename(Statistic = X1,
##         Lags = X2, Null_Hypothesis = X3, p_value = X4) %>% mutate(across(names(.)[-3],
##         function(x) as.double(x))) %>% mutate(across(names(.)[-3],
##         function(x) round(x, 4)))
##     final <- rbind(df3, df2)
##     return(final)
## }
## <bytecode: 0x00000182589be088>
```

```
joint_adf
```

```
## function (df)
## {
##     df <- df %>% dplyr::select(1:8)
##     logs <- df %>% dplyr::select("date" | starts_with("L"))
##     non_logs <- df %>% dplyr::select(date | !starts_with("L"))
##     adf_logs <- logs %>% dplyr::select(-date) %>% ts(.) %>% sapply(.,
##         function(x) adf.test(x))
##     adf_non_logs <- non_logs %>% dplyr::select(-date) %>% ts(.) %>%
##         sapply(., function(x) adf.test(x))
##     df2 <- data.frame(matrix(unlist(adf_logs), nrow = 3, byrow = T),
##         row.names = c("LREERS", "LRGDPPCR", "LPR2COMM5")) %>%
##         dplyr::select(-c(X5, X6)) %>% rename(Statistic = X1,
##         Lags = X2, Null_Hypothesis = X3, p_value = X4) %>% mutate(across(names(.)[-3],
##         function(x) as.double(x))) %>% mutate(across(names(.)[-3],
```

```
##         function(x) round(x, 4)))
##     df3 <- data.frame(matrix(unlist(adf_non_logs), nrow = 4,
##         byrow = T), row.names = c("RIRR", "OPENY", "FBYA", "NFAOFPY")) %>%
##         dplyr::select(-c(X5, X6)) %>% rename(Statistic = X1,
##         Lags = X2, Null_Hypothesis = X3, p_value = X4) %>% mutate(across(names(.)[-3],
##         function(x) as.double(x))) %>% mutate(across(names(.)[-3],
##         function(x) round(x, 4)))
##     final <- rbind(df3, df2)
##     return(final)
## }
## <bytecode: 0x00000182497e3760>
```

```
plot_endog1
```

```
## function (df)
## {
##     plot <- df %>% dplyr::select(date, LREERS) %>% mutate(REERS = exp(LREERS),
##         .keep = "unused") %>% ggplot() + geom_line(aes(x = date,
##         y = REERS)) + scale_y_continuous("Real Effective Exchange Rate") +
##         geom_vline(xintercept = c(as.Date("1985-12-31"), as.Date("1986-03-31")),
##             color = "red", show.legend = T)
##     return(plot)
## }
```

```
plot_endog2
```

```
## function (df)
## {
##     plot <- df %>% mutate(across(names(.[, grepl(x = names(.),
##         pattern = "^L.*")]), function(x) x = exp(x)), .keep = "unused") %>%
##         gather(Econ_Measure, Value, 2:13) %>% dplyr::select(date,
##         Econ_Measure, Value) %>% filter(!Econ_Measure %in% c("LREERS",
##         "LPR2GOLD", "LPR2COMM3", "LPRCOMM3", "LPRCOMM5", "LPRGOLD")) %>%
##         ggplot() + geom_line(aes(x = date, y = Value, color = Econ_Measure,
##         linetype = Econ_Measure)) + facet_wrap(~Econ_Measure,
##         scales = "free_y", ncol = 2) + scale_color_hue(labels = c("Fiscal Balance",
##         "Real Commodity\nPrices", "Relative Real\nGDP per Capita",
```

```
##         "Net Foreign\nAssets", "Openness Indicator", "Relative Real\nInterest Rate")) +
##         scale_linetype_discrete(labels = c("Fiscal Balance",
##             "Real Commodity\nPrices", "Relative Real\nGDP per Capita",
##             "Net Foreign\nAssets", "Openness Indicator", "Relative Real\nInterest Rate")) +
##         guides(linetype = guide_legend(title = NULL), color = guide_legend(title = NULL))
##     return(plot)
## }
## <bytecode: 0x00000182525aef40>
```

```
plot_endog_joint
```

```
## function (df)
## {
##     plot <- df %>% gather(Econ_Measure, Value, 2:13) %>% dplyr::select(date,
##         Econ_Measure, Value) %>% filter(!Econ_Measure %in% c("LPR2GOLD",
##         "LPR2COMM3", "LPRCOMM3", "LPRCOMM5", "LPRGOLD")) %>%
##         ggplot() + geom_line(aes(x = date, y = Value, color = Econ_Measure,
##         linetype = Econ_Measure)) + facet_wrap(~Econ_Measure,
##         scales = "free_y", ncol = 2) + scale_color_hue(labels = c("Fiscal Balance",
##         "Real Commodity\nPrices", "REER", "Relative Real\nGDP per Capita",
##         "Net Foreign\nAssets", "Openness Indicator", "Relative Real\nInterest Rate")) +
##         scale_linetype_discrete(labels = c("Fiscal Balance",
##             "Real Commodity\nPrices", "REER", "Relative Real\nGDP per Capita",
##             "Net Foreign\nAssets", "Openness Indicator", "Relative Real\nInterest Rate")) +
##         guides(linetype = guide_legend(title = NULL), color = guide_legend(title = NULL))
##     return(plot)
## }
## <bytecode: 0x000001825921c840>
```

```
plot_endog1_log
```

```
## function (df)
## {
##     plot <- df %>% dplyr::select(date, LREERS) %>% ggplot() +
##         geom_line(aes(x = date, y = LREERS)) + scale_y_continuous("Real Effective Exchange Ra
##         geom_vline(xintercept = c(as.Date("1985-12-31"), as.Date("1986-03-31")),
##             color = "red", show.legend = T)
```

```
##     return(plot)
## }
```

```
plot_endog2_log
```

```
## function (df)
## {
##     plot <- df %>% gather(Econ_Measure, Value, 2:13) %>% dplyr::select(date,
##         Econ_Measure, Value) %>% filter(!Econ_Measure %in% c("LREERS",
##         "LPR2GOLD", "LPR2COMM3", "LPRCOMM3", "LPRCOMM5", "LPRGOLD")) %>%
##         ggplot() + geom_line(aes(x = date, y = Value, color = Econ_Measure,
##         linetype = Econ_Measure)) + facet_wrap(~Econ_Measure,
##         scales = "free_y", ncol = 2) + scale_color_hue(labels = c("Fiscal Balance",
##         "Real Commodity\nPrices", "Relative Real\nGDP per Capita",
##         "Net Foreign\nAssets", "Openness Indicator", "Relative Real\nInterest Rate")) +
##         scale_linetype_discrete(labels = c("Fiscal Balance",
##             "Real Commodity\nPrices", "Relative Real\nGDP per Capita",
##             "Net Foreign\nAssets", "Openness Indicator", "Relative Real\nInterest Rate")) +
##         guides(linetype = guide_legend(title = NULL), color = guide_legend(title = NULL))
##     return(plot)
## }
## <bytecode: 0x0000018259430d68>
```

14.1. VECM Model Estimation in Theory:

paper table 1 col's:

1. Var's + Seasonal(4) + lags(4) 2. ^{Above} + outlier_dummies -> Trace test $\implies$ two CI vectors

Multiple ways to estimate VEC models:

First approach: ordinary least squares (yields accurate result) but does not allow to estimate the cointegrating relations among the variables. The estimated generalised least squares (EGLS) approach would be an alternative.

Most popular estimator: MLE of Johansen (1995) [In R: ca.jo function of the urca package of Pfaff (2008a)] Alternatively, `VECM` of tsDyn package of Di Narzo et al. (2020)

Before VECM: 1. Determine lag order p -> Det. rank of CI matrix r 2. deterministic terms have to *be specified. 3. Choose lag order: est. the VAR in levels 4. Choose lag specification that minimises an Information criterion*