Chapter 4 Data and empirical analysis

4.1 Methodology

This paper employs cointegration techniques to detect the existence of long run relationship between TB and other macroeconomic variables. The choice of cointegration is based on the following considerations.

First and foremost, the cointegration approach allows us to test for the existence of the equilibrium relationship postulated by the economic theory. (Perman, 1991)

Secondly, regression model for time series data often assumes that the series is stationary. But what happen if the series is found to be non-stationary? In this case, the statistical inference will be invalid. As a result, models incorporating non-stationary series are mis-specified, at least in a statistical sense. To solve this, one could difference the series until stationary is achieved and then performs regression using these differenced variables. However, we will lose valuable inference. (Perman, 1991)

The cointegration approach allows us to specify an equation in which all terms are stationary. As such, the statistical inference is still valid. Besides, it enables us to retain information about the long run relationship between variables at levels. (Perman, 1991)

In addition, Gujarati (1995) pointed out that regressing a time series on another time series often obtains a very high R^2 although there is no meaningful relationship between the two. This is known as spurious regression as the standard t and F testing procedure are not valid in this case. This problem arises as both series are trended or in another words not stationary.

In order to test for the cointegration, we use Engle-Granger Methodology as outlined by Enders (1995) and Bahmani-Oskooee & Pourheydarian (1992). We choose Engle-Granger methodology as it is often used in practise for bivariate cointegration analysis. Maddala and Kim (1998) suggest that the parameter estimated by using Engle-Granger method is unique and superconsistent for 2-variable model.

The first step of the Engle-Granger method is to pre-test the variables for their orders of integration[6]. As cointegration necessitates the variables to be integrated of the same order, we need to pre-test each of the variables to determine their orders of integration (d). In this case, we employ the Augmented Dickey-Fuller(ADF) tests.

If both TB and X_i (one of the other macroeconomic variables) are stationary, that is I (0), we will use standard time-series method rather than the

[6] A non-stationary time series is said to be integrated of order d, or is denoted as I (d) if it achieves stationary after being differenced d times.

cointegration method. If they are integrated of different orders, we can conclude that they are not cointegrated.

The next step is to estimate the long run equilibrium relationship. If both TB and X_i are integrated of order d, which is denoted as I (d), we will estimate the following equations by using the ordinary least square method:

$$TB = a + b X_i + u_1$$

$$X_i = c + d TB + u_2$$

Subsequently, we will determine the order of cointegration of the residuals, u_1 and u_2 by using the ADF test. If the order of cointegration of u_1 and u_2 are less than d, we conclude that there is a long run relationship between TB and X_i.

4.2 Data definition and sources

Before we progress into the empirical tests, we should discuss on the data definition and sources. The data used in this study were obtained from various issues of the International Financial Statistics of IMF, Bank Negara Quarterly Bulletin, Bank Negara Monthly Statistical Bulletin and Economic Report of the Ministry of Finance.

We use quarterly data from 1973 1Q to 1998 4Q for this research. 1973 1Q is chosen as 1973 is the year where most data required are available. In addition, March 1973 marks the beginning of the flexible exchange rate system after the collapse of the Bretton Woods system. 1998 4Q is chosen as the annual data for 1999 are not yet available. In addition, the calculation of NEX and REX (which based on trade-weighted basket of currencies) for 1999 are greatly hindered by the problem of inconsistency as European Union members (except United Kingdom) ceased to use their original currencies since 1 January 1999.

All the variables are expressed in real terms (at 1980 prices). For the purpose of deflating the nominal data into the real terms, Consumer Price Index (CPI) was used, as it is the only measurement of price, which is available throughout the observed period. For example, in deflating the trade balance, export price index and import price index are better alternatives. However, the data for these indexes are only available up to the year 1987. The base year is 1980 as it is a relatively stable year for Malaysia (with relatively moderate inflation and economic growth).

As the quarterly data for Malaysian GDP are not available prior to 1993, we employ the interpolation method. In this case, we use the rate of change of quarterly IPI figures as the proxy. This can be shown mathematically as follows.

$$GDP_{Qi} = AnnualGDP \times \frac{IPI_{Qi}}{\sum_{i=1}^{4} IPI_{Qi}} \text{ where Q = quarter and i = 1, 2, 3, 4.}$$

In addition, as some Wholesale Price Index (WPI) figures for Malaysian trading partners are either not available or not consistent; CPI figures are being used as the proxy for the computation of the REX.

The definitions and interpretations of the variables are as follows.

- Trade balance (TB) in real terms is defined as the excess of export of goods (f.o.b.) over imports of goods (c.i.f.) after deflated by CPI (1980=100).
- M1 for Malaysia in real terms is defined as the sum of coin, currency notes and demand deposits held by the private sector after deflated by CPI (1980=100).
- The definition of M2 for Malaysia has changed over the years to reflect the introduction of new financial instruments, which serve as temporary store of value. M2 is currently defined by Bank Negara Malaysia as M1 plus the following financial assets of the private sector at or by the commercial bank: (i) savings deposits; (ii) fixed deposits; (iii) net issues of negotiable certificates of deposits (NCDs); and (iv) repurchase agreements (Repos). M2 in real terms is deflated by CPI (1980=100).
- Fiscal balance (FB) in real terms is defined as the federal government's overall budget surplus after deflated by CPI (1980=100).

  A budget surplus reflects either an increase in tax revenue or a cut in government expenditure, that is, it represents contractionary fiscal policy

- Terms of Trade (TOT) in real terms are defined as the ratio of export price index over the import price index after deflated by CPI (1980=100).
- Gross Domestic Product (GDP) is defined as real GDP in RM' million (1980=100).
- Interest rates (R) is defined in real terms as the average overnight money inter-bank rates after deflated by CPI (1980=100).
- Nominal Exchange Rates (NEX) is defined as the index of total trade-weighted nominal effective exchange rate using 1980 as the base year.
- Real Exchange Rate (REX) is defined as the index of total trade-weighted real effective exchange rate using 1980 as the base year. The computation method of the REX is largely adapted from the one used by Bahmani-Oskooee and Pourheydarian (1992). The REX of each trading partners (REX_i) was derived by using the following equation:

$$REX_i = E_i \times WPI_i \div CPI_{Malaysia}$$

where E = Nominal exchange rate (price of foreign currency in terms of ringgit)

WPI_i = Wholesale price index for trading partner i.[7]

CPI = Consumer price index

Subsequently, the REX_i will be converted into index form with 1980 as the base year by using the following equation:

$$IREX_i = REX_i \div REX_i^{1980} \times 100$$

[7] For some countries CPI instead of WPI was used as the WPI data were either not consistent or unavailable throughout the period of observations.

As the REX is total trade-weighted, the $IREX_i$ will then be multiplied with the share of Malaysian total trade by trading partner i (which is denoted as S_i) where the $\sum s_i = 100\%$. Fifteen major trading partners (i's) were selected. Their names and respective shares of total trade are listed in Table 2. The average shares throughout the observation periods were used to construct the REX. The equation of the REX is as follows:

$$REX = \sum s_i REX_i$$

Table 2: The direction of trade
(annual average of 1973-1998)

Country	Total trade*	Export	Import
Japan	20.6%	17.7%	23.6%
Singapore	16.6%	20.2%	12.5%
US	16.0%	16.7%	15.4%
UK	4.5%	4.0%	5.1%
Germany	4.0%	3.5%	4.7%
Taiwan	3.3%	2.6%	4.1%
Australia	3.0%	1.8%	4.5%
Thailand	2.9%	2.6%	3.4%
S. Korea	2.9%	3.4%	2.4%
Netherlands	2.8%	4.4%	0.9%
France	1.6%	1.4%	1.8%
India	1.4%	1.8%	1.0%
Italy	1.2%	1.2%	1.2%
Philippines	1.1%	1.3%	0.9%
Indonesia	1.1%	0.8%	1.4%
Total	83.2%	83.4%	82.6%

* China and Hong Kong have 2.4% and 2.3% shares of Malaysian total trade on average. However, as the quarterly data for WPI/CPI for China and WPI/CPI and exchange rate data are unavailable throughout the period, they are excluded from the list.
Source: Direction of Trade Statistics of IMF.

4.3 Empirical results

After the discussion of the data definitions and sources, we will proceed into the empirical results. Before we started with the pre-test of the order of integration, it is a good practice to inspect the plots of the time series to identify whether the series is stationary.

From the plots of all the series against time (t), we observe no tendency of these variables to revert to their long run mean. This suggests that they are non-stationary. In addition, no transformation such as logarithm is needed, as there is no sign of increasing variance over time. However, the plots of the first differenced of all these series appear to be stationary. This suggests that the series are not I (0) and could be I (1). Please refer to Appendix 1 for the plots.

Subsequently, we will examine whether any series has trend component. The plots of M1, M2, NEX, REX and Y exhibit increasing trend. This is further confirmed by the plots of their first differenced (Y_t-Y_{t-1}) which are horizontal with a non-zero mean.

In addition, as shown in Appendix 2, the correlograms of these variables show slowly declining autocorrelation functions (ACFs). This provides further proof that M1, M2, NEX, REX and Y are trended.

The next consideration is whether these variables are seasonal. As indicated by the correlogram, FB is the only series that appears to have seasonal pattern as its ACF peaks at lag 4, 8, 12, 16 and so on. In order to confirm this, we apply the test of seasonality on all variables. In this case, we used Kruskal-Wallis tests to test for seasonality.

The null hypothesis of the test is no seasonality or $S_1=S_2=S_3=S_4=0$ and the test statistic is H, which can be expressed mathematically as:

$$H = \frac{12}{n(n+1)}\left[\sum \frac{R_i^2}{n_i}\right] - 3(n+1)$$

where n_i = number of observation in *i*th season

n = total number of specific seasonals = $\sum n_i$

$y_t^{'}$ = specific seasonal for time t

R_i = $\sum Rank(y_t^{'})$

The critical value is $\chi^2_{0.05,4-1}$ which is equal to 7.81473. The decision rule is that we will reject the null hypothesis (H_o) if H is greater than the critical value. If H_o is rejected, we conclude that the series has seasonality. If the H_o is not rejected, we conclude that the series has no seasonality. The test results are summarized in the following table.

Table 3: Results for the test of seasonality

	Critical value	H	Decision	Conclusion (at 95% confidence level)
TB	7.81473	0.180642	accept Ho	No seasonality
Y	7.81473	0.255255	accept Ho	No seasonality
R	7.81473	2.792222	accept Ho	No seasonality
M1	7.81473	0.229687	accept Ho	No seasonality
M2	7.81473	0.149085	accept Ho	No seasonality
FB	7.81473	54.36741	reject Ho	Has seasonality
REX	7.81473	0.276166	accept Ho	No seasonality
NEX	7.81473	0.052705	accept Ho	No seasonality
TOT	7.81473	0.101504	accept Ho	No seasonality

As FB is the only variable that was found to have seasonality, we shall remove its seasonality before it could be tested along with other variables that are not seasonal. According to Enders (1995), the Dickey-Fuller procedure must be modified to enable test for seasonal unit root test. In the existence of seasonality, Enders suggests that we can remove the seasonality through the estimation of the following equation by assuming the seasonal pattern to be purely deterministic.

$Y_t = a_0 + a_1D_2+a_2D_3+a_3D_4+e_t$ where D_2, D_3 and D_4 represent seasonal dummy variables of which $D_i=1$ in season (i+1) and zero otherwise.

In this case, e_t, the residuals, is the deseasonalised Y_t. For the case of FB, we will use FBDS (the deseasonalized FB) rather than FB for the unit root test.

Subsequently, we need to consider whether structural change has occurred. According to Enders (1995), when there are structural breaks the Dickey-Fuller test statistics are biased toward the non-rejection of a unit root[8]. From the inspection of the plots in Appendix 1, there is no sign of structural change. Although the quarterly data in the 1997 and 1998 seem to show structural breaks, they should be treated as outliers rather than structural breaks. This is because they only appear at the end of the observed period (1973 to 1998).

The next consideration is the power of the test. Power of a test refers to the probability of rejecting a false null hypothesis. A unit root test of low power indicates that it does not have the power to distinguish between a unit root and near unit root process. The test will often indicate that a series contains a unit root. In this case, the Augmented Dickey-Fuller tests are known to have low power. However, this may not pose serious problem for this study as the sample size is considerably large, that is, 104 for all the series (except TOT which has sample size of 60). In another words, the degree of freedom is relatively high.

After taking into account the above-mentioned considerations, we will now proceed into formal test for stationarity. In order to confirm the non-stationarity of the variables, the Augmented Dickey-Fuller (ADF) unit root tests are conducted. The equation used for the test is shown below:

$$\Delta X_t = \alpha + \beta t + \rho X_{t-1} + \sum_{i=2}^{k} \delta_i \Delta X_{t-i+1} + e_t \qquad \text{............(1)}$$

where t is the time or trend variable, e is the error term and k is the number of lags.

The test is whether the estimate of ρ is equal to zero, that is, a unit root exists in X. Time or trend variables will only be included if the series is identified as trended. As this research is based on quarterly data, we will start with 2 years lag (k=8) while the choice of the number of lags (for each variable) included for the lagged difference term is based on the Akaike Information Criterion (AIC)[9].

[8] For further information on structural change, please refer to Enders (1995), p. 243-251.

[9] AIC=nlog(s^2)+2k where s^2 is the estimate of the residuals variance while k is the total number of estimated parameters. AIC gives non-linear trade-off between the residuals variance and the value of k, since a model with a higher k value will only be preferred if there is a proportionately large fall in s^2 (Holden, Peel and Thompson, 1990).

The AIC suggests that the choice of the lag length shall be based on the lowest AIC value. Table 4 and 5 report the AIC value of different lag length.

Table 4: Akaike Information Criterion (level)

	lag 0	lag 1	lag 2	lag 3	lag 4	lag 5	Lag 6	lag 7	lag 8
FBDS(c)	14.0357	13.9914	13.9651	**13.3883**	13.4175	13.5615	13.5857	13.6171	13.6506
M1(t)	13.4853	13.4344	13.4591	13.4807	13.4954	13.5062	13.4356	13.4629	**13.3393**
M2(t)	15.3264	**15.3170**	15.3295	15.3597	15.3911	15.4127	15.4417	15.4431	15.3849
NEX(t)	3.3553	3.2842	3.2882	3.3034	3.3252	3.3548	3.3525	3.2403	**3.2339**
R(c)	-0.6635	-0.7019	-0.6944	-0.6734	-0.6921	**-0.7110**	-0.7074	-0.7006	-0.6672
REX(t)	2.6391	2.5370	2.5340	2.5498	2.5577	2.5868	2.6142	**2.5334**	2.5502
TB(c)	13.7495	13.7541	13.7007	**13.6847**	13.7099	13.7359	13.7511	13.7781	13.7301
TOT(c)	3.3362	3.1569	3.1438	3.1894	3.0744	2.8488	**2.8173**	2.8535	2.9084
Y(t)	13.9715	13.9992	13.6877	13.6631	13.2810	13.3080	13.3262	13.3586	**13.1800**

Table 5: Akaike Information Criterion (first differenced)

	lag 0	lag 1	lag 2	lag 3	lag 4	lag 5	lag 6	lag 7	lag 8
FBDS(c)	14.1696	14.0991	**13.5754**	13.5991	**13.5894**	13.6144	13.6432	13.6744	13.6873
M1(t)	13.4498	13.4779	13.5055	13.5302	13.5181	13.4222	13.4475	**13.3464**	13.3804
M2(t)	**15.2982**	15.3125	15.3416	15.3724	15.3992	15.4313	15.4224	15.3994	15.3882
NEX(t)	3.3138	3.2973	3.3216	3.3498	3.3811	3.3896	**3.2377**	3.2404	3.2509
R(c)	-0.5751	-0.6183	-0.6209	-0.6079	**-0.6672**	-0.6534	-0.6265	-0.6018	-0.5864
REX(t)	2.6116	**2.5714**	2.5998	2.6114	2.6425	2.6651	2.5820	2.5830	2.6121
TB(c)	13.7407	**13.6864**	13.6926	13.7234	13.7463	13.7524	13.7836	13.7654	13.7895
TOT(c)	3.2296	3.2642	3.3033	3.1594	2.8926	**2.8656**	2.9102	2.9415	3.0074
Y(t)	14.0196	13.6835	13.6545	13.3440	13.3650	13.3951	13.4170	13.3419	**13.3310**

The null hypothesis of the ADF test is that the series is not stationary or has unit root. We will reject H_0 if the absolute value of the test statistics is greater than the critical value. If H_0 is rejected, we conclude that the series is stationary. Otherwise, we conclude that the series is non-stationary. The test results could be summarized in Table 6.

Table 6: The calculated ADF test statistics for level and first differenced data

	ADF Statistics			ADF Statistics		
Series	Level	lag		First Differenced	lag	
FBDS(c)	-1.553949	3		-15.753100	2	***
M1(t)	-1.547591	8		-2.607734	7	
M2(t)	0.270022	1		-8.092680	0	***
NEX(t)	-1.532798	8		-4.254254	6	***
R(c)	-2.458144	5		-6.031678	4	***
REX(t)	-2.486917	7		-7.597572	1	***
TB(c)	-1.641451	3		-5.187218	1	***
TOT(c)	-2.010240	6		-3.827884	5	***
Y(t)	-4.105626	8	***	-0.523022	8	

The sample size of all variables except TOT (which is 60) is 104.
Time trend is included for those variables confirmed to have trend by plots and correlograms.
(c) means with constant and without time trend
(t) means with constant and time trend
*** reject at 99% level of confidence
The MacKinnon t-critical values (obtained from Eviews):

Level of significance		1%	5%	10%		1%	5%	10%
N=50	(c)	-3.5653	-2.9202	-2.5977	(t)	-4.1498	-3.5005	-3.3294
N=100	(c)	-3.4965	-2.8903	-2.5819	(t)	-4.0521	-3.4548	-3.2283

As indicated by Table 6, Y is the only variable that is stationary at level. In another words, Y is an I (0) series. As, all the other variables are not stationary, we have to rely on the cointegration techniques to ascertain the long run equilibrium relationship.

For the ADF test of the first differenced, Y and M1 are found to be non-stationary. Subsequently, M1 is tested for the second differenced and is found to be integrated of order two, I (2). In short, all variables except Y and M1 become stationary after being differenced once, that is, they are I (1) series.

As discussed earlier, cointegration necessitates the variables to be integrated of the same order. In the case where they are integrated of different orders, we can conclude that they are not cointegrated. As such, we can conclude that Y and M1 are not cointegrated with TB.

Subsequently, we perform cointegration tests on all the I (1) series, namely, TB, FBDS, R, REX, NEX, TOT and M2, to further investigate whether FBDS, R, REX, NEX, TOT and M2 are cointegrated with TB.

At first, we regress the two series TB and X_i (macroeconomic variables other than TB in turn) and estimate the following equations by using the ordinary least square method:

$$TB = a + b\,X_i + u_1 \qquad \text{............(2)}$$

$$X_i = c + d\,TB + u_2 \qquad \text{............(3)}$$

As these variables are all I (1) series, the residuals, that is, u_1 and u_2 must be I (0) so that TB and X_i are cointegrated. In this case, the ADF test is again applied to identify the order of integration of u_1 and u_2. However, the intercept (c) or time trend (t), will not be included as u_1 and u_2 are residuals from the regression equation (Enders, 1995). The test results are summarized in Table 7.

Table 7: The ADF test of the residuals of cointegration equation

Pair	Cointegration Equation	D-W	Adjusted R^2	ADF	lag		Conclusion
1	TB=f(FBDS)	0.307370	0.010375	-1.42977	3		] Not cointegrated
	FBDS=f(TB)	1.322410	0.010375	-1.62890	3	*	]
2	TB=f(R)	0.280943	-0.002901	-1.89009	3	*	] Cointegrated
	R=f(TB)	0.394219	-0.002901	-2.49319	5	**	]
3	TB=f(REX)	0.291926	0.041232	-2.02210	3	**	] Not cointegrated
	REX=f(TB)	0.067605	0.041232	-1.50856	1		]
4	TB=f(NEX)	0.294507	0.058420	-2.08289	3	**	] Not cointegrated
	NEX=f(TB)	0.052815	0.058420	0.02682	7		]
5	TB=f(TOT)	0.273153	0.034977	-1.20418	4		] Not cointegrated
	TOT=f(TB)	0.183128	0.034977	-1.59474	6		]
6	TB=f(M2)	0.293509	0.024070	-2.22168	3	**	] Not cointegrated
	M2=f(TB)	0.015706	0.024070	-0.64658	2		]

The sample size of all variables except TOT (which is 60) is 104.

* and ** mean reject at 90% and 95% level of confidence respectively

The MacKinnon t-critical values (without trend and constant) (obtained from Eviews):

Level of significance	1%	5%	10%
N=50	-2.6090	-1.9473	-1.6192
N=100	-2.5864	-1.9433	-1.6174

The results from Table 7 had to be analyzed in pairs. For the first, third, fourth and sixth pairs, only one of the u (that is, u_1 or u_2) is found to be I(0). As such, we can not conclude that TB is cointegrated with FBDS, REX, NEX and M2. In other words, they are no long run relationship between TB and the above mentioned macroeconomic variables. The case for the fifth pair, u_1 and u_2 are not I(0). As such, we can conclude that TB and TOT do not have long run relationship.

On the other hand, the second pair, that is, TB and R, is found to be I(0) at 10% level of significance. This shows that even though both of them are not stationary at level, the linear combination of the two are stationary. As such, we can conclude that they are cointegrated. In another words, we can conclude that TB and R has long run equilibrium relationship.