CHAPTER 3

METHODOLOGY

Introduction

This Chapter discusses on the methodology of identifying the characteristics of UE that escape taxes, conceptual framework of estimating its size, growth and correlation coefficient with priori economic variables. It explains how data were gathered, highlights the limitation of data availability and elaborates the importance of assumptions to qualify estimates. Malaysian UE is hypothesized to associate with priori economic variables of; federal expenditure, economic crises, inflation rate, cash in circulation and electricity consumption, GDP per capita, taxes, crime rate and bribery index.

3.1 Scope of underground economy in this study

This study estimates a portion of the second economy which is the UE that escape taxes. It is based on income of legal activities that escape taxes (irregular economy) and illicit activities that naturally escape taxes (illegal economy). The former constitutes the permissible production of goods and services (legal activities) but the distributions of goods are non compliance with tax rules. While the latter deals with goods and services, which are prohibited by criminal law that naturally, escape tax law.

A schematic presentation of the components of the "potential economy" is illustrated in Figure 2.1. The "potential economy" consists of GDP as the official economy and the "second economy". The "second economy" often known as hidden economy or shadow economy comprises informal economy and UE. UE is divided into components of irregular economy and illegal economy with a portion of income not reported to tax authority.

Based on the concept that participants of the second economy would conceal their activities, the income of UE would not be reported to tax authority. Assuming that the unreported taxable incomes represent unreported value added economy, tax non compliant could be used to measure the unreported economy.

3.2 Approaches and methods employed in this study

The country's economic incidences were first explored to obtain some initial views on the characteristics of proxy indicators of UE in Malaysia. Descriptive statistics of proxy indicators are discussed in Chapter 4. The size of UE is estimated based on tax data. It adopts the idea of Frey and Pommerehne (1984) and Clotfelter (1983), who measured tax evasion by the difference between taxable income calculated by audit and individuals, as the indicator of UE.

Despite many critics on the assumption of value added income using tax audit data, Slemrod (2007) gave some comfort of employing tax data. He showed that about 80% of tax evasion comes from understating of income rather than overstating deductions. Based on this finding, and considering that UE is a latent variable that gives impact on taxes, he emphasized that it is a practical measure.

The large proportion of understating income reasonably offers some comfort in assuming that UE estimates derived from tax evasion are largely value added income rather than redistribution income. Tax non compliance ratio is assumed to mirror in the country's economic non-compliance (UE).

To reduce perception of unique UE estimates, methodology employed five statistical techniques on four sets of enforcement time series data that escaped taxes for comparison. Estimates were compared whether they vary with samples, data period, statistical technique employed and assumptions made.

The first estimate is based on the coefficient of correlation between reported income and non-reported income using linear regression techniques as discussed in paragraph 5.3.7 of Chapter 5. The second estimate and third estimate is based on the coefficient of correlation matrix and regression between reported and unreported income respectively as discussed in paragraph 5.3.7 of Chapter 5. The fourth estimate is based on the percentage of non submission of direct tax declaration forms as discussed in paragraph 6.1 of Chapter 6. The fifth estimate is the time series estimates over a 30 - year period (1980-2009), based on tax non compliance ratio of enforcement statistic and voluntary tax records as discussed in paragraphs 6.2 and 6.5 of Chapter 6. The time series estimate is in point and range series. The latter is computed based on amplifying techniques, which is the product of enforcement success rate and point estimate.

This study differs from past studies as it employed various statistical methods on different data sets to generate comprehensive insights into UE (established the characteristics, size, growth, economic development and government interaction).

Data set encompasses both direct and indirect tax gap (non compliance) that include income of legal and illicit activities of institutional and individual participants.

As data availability is limited, to qualify estimate of the size of a latent variable (UE), estimates must be interpreted according to assumptions below:

- The economic components of UE are irregular economy and illegal economy.
- The components of irregular economy are direct tax gap and indirect tax gap.
- The components of direct tax gap are tax evasion due to under reporting of income in the existing tax base, unpaid reported tax and tax evasion due to non filing of income outside the captured tax base.
- The participants of tax non compliance are mainly institutional and individual tax payers.
- The unreported income of the tax payers who filed tax form is assumed the unreported income of legal activities.
- The income of indirect tax gap (smuggling activities) and income of commercial crime (illicit activities) are assumed the income of illegal economy, which naturally escape taxes.
- The components of indirect tax gap are tax evasion, mainly on smuggled goods mainly liquor, cigarettes, vehicles etc. and unpaid taxes representing evasion on other duties.
- The components of illicit activities are mainly the income of commercial activities of bribery, drug trafficking and frauds).
- The income characteristics of direct tax non compliance of tax audits and investigation cases mirrors in the characteristics of UE.
- The participation risk as determined by enforcement efficiency represents UE risk.
- The enforcement effectiveness as determined by enforcement ability

represents UE deterrence.

- The correlation between the time series UE estimates and economic variables explains for its economic development.

3.3 Study framework of estimating the size of underground economy

The components of tax gap are tax evasion (unreported taxable income) and non payment of reported taxable income. Assuming that they are mutually exclusive, estimates based on tax gap would uncover more underground activities compared to tax evasion as employed by past studies. They include estimates of the proportion of UE components by income of legal activities (irregular economy) and illicit activities (illegal economy). The former are activities of tax non compliance while the latter are activities of commercial crime – non compliance with criminal law. Figures 3.1 - 3.3 provide conceptual framework on estimation of UE. It illustrates a schematic version of the development of UE based on the concept of circular revenue-expenditure from tax perspective. The main participants are the public sector (government) who imposed taxes on the private sector (institutional and individual tax payers). This framework assumed that; the "actual" tax base mirrors in the "actual" or potential economy; tax compliance mirrors in the official economy; and tax non compliance mirrors in the UE. The income of direct tax non compliance of the tax net (as captured by tax audit on tax payers who file tax form) mirrors in the unreported income of legal activities. The income of illicit activities (commercial crime and smuggling) mirrors in the unreported income of tax payers who do not file tax form (people who escape the tax net). Based on these assumptions, the taxable income exiting the legal

framework of voluntary reporting would resemble the economy exiting the official economy, i.e. the UE.

The private sector has the option to either report or conceal taxable income. Where taxable income is reported to tax office and tax is assessed and paid, it is tax compliance (TC). There will be no tax loss on unreported income that is below tax bracket or tax scope. Tax non compliance (TNC) or tax gap is where the taxable income is not reported (mission of income or tax evasion) including unpaid taxes on reported income.

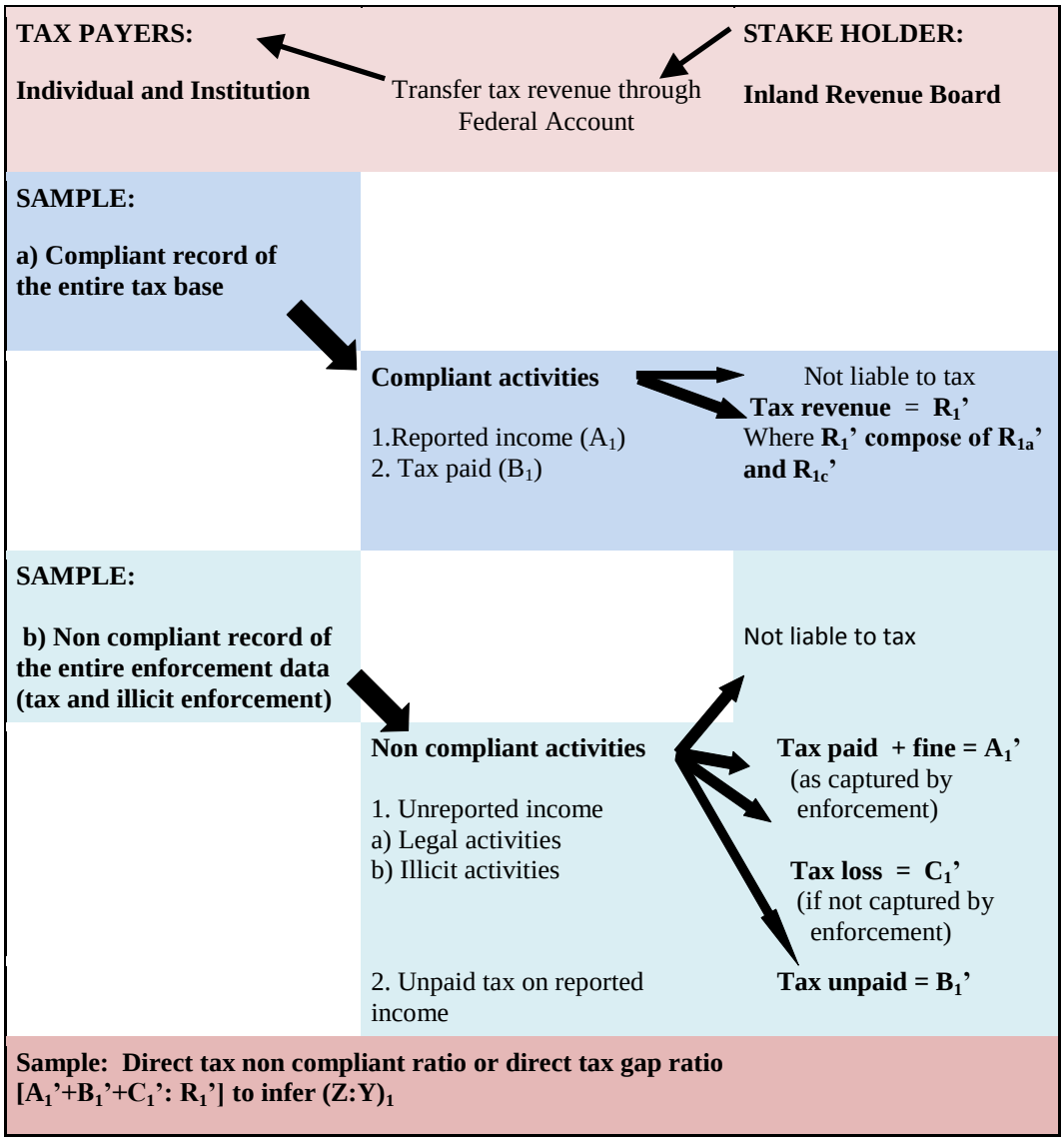

Figure 3.1: Schematic presentation of estimating direct tax gap – evade direct tax law

a) The entire compliant record constitutes tax revenue
b) The entire non compliant record consists of unreported taxable income (legal and illicit activities) and unpaid tax, tax paid plus fine and tax loss.
c) Direct tax non compliant to compliant ratio is **$[A_1'+B_1'+C_1': R_1']$**
d) Where, $R_1' = R_{1a} + R_{1c}$, and R_{1a} is total tax assessed and paid and R_{1c}' is total tax collected (paid)

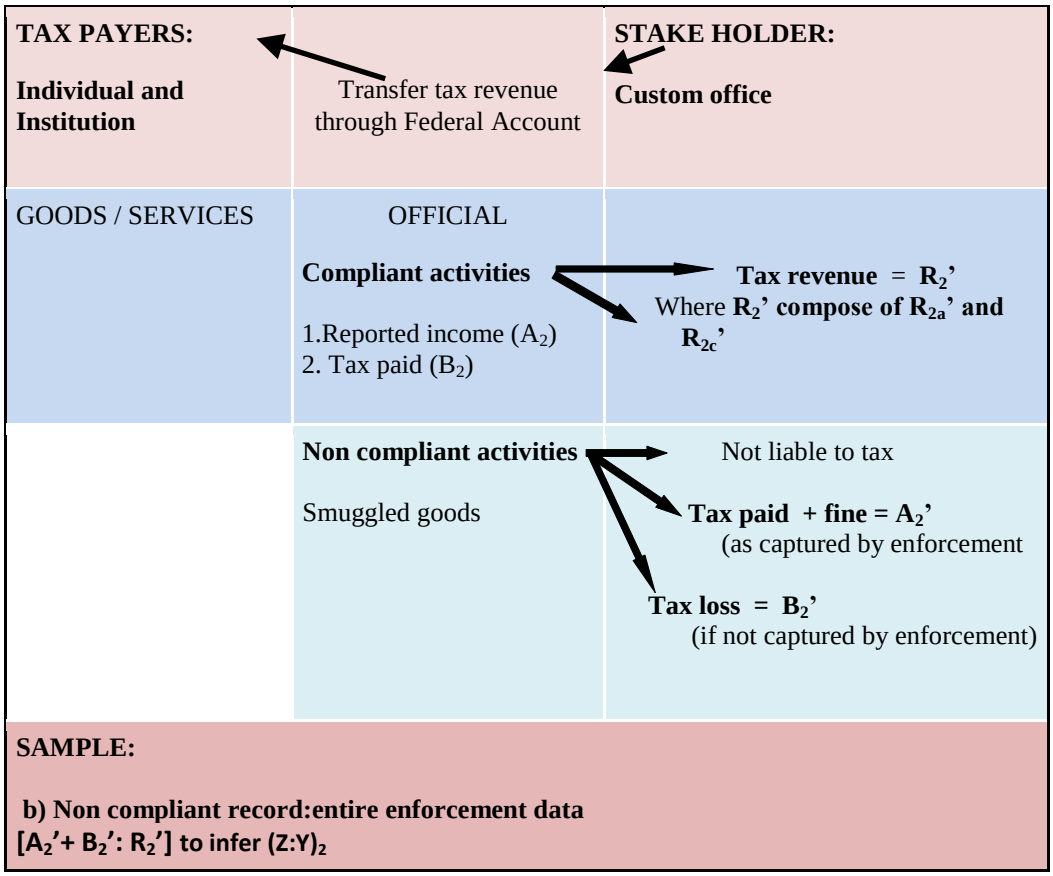

Figure 3.2: Schematic presentation of estimating indirect tax gap - evade indirect tax law

a) The entire compliant record constitutes tax revenue, as part of the federal revenue
b) The entire non compliant record consists of smuggled goods that escape indirect tax.
c) Direct tax non compliant to compliant ratio is **[A_2'+ B_2': R_2']**
d) R_{2a}' is total tax assessed and R_{2c}' is total tax collected (paid)

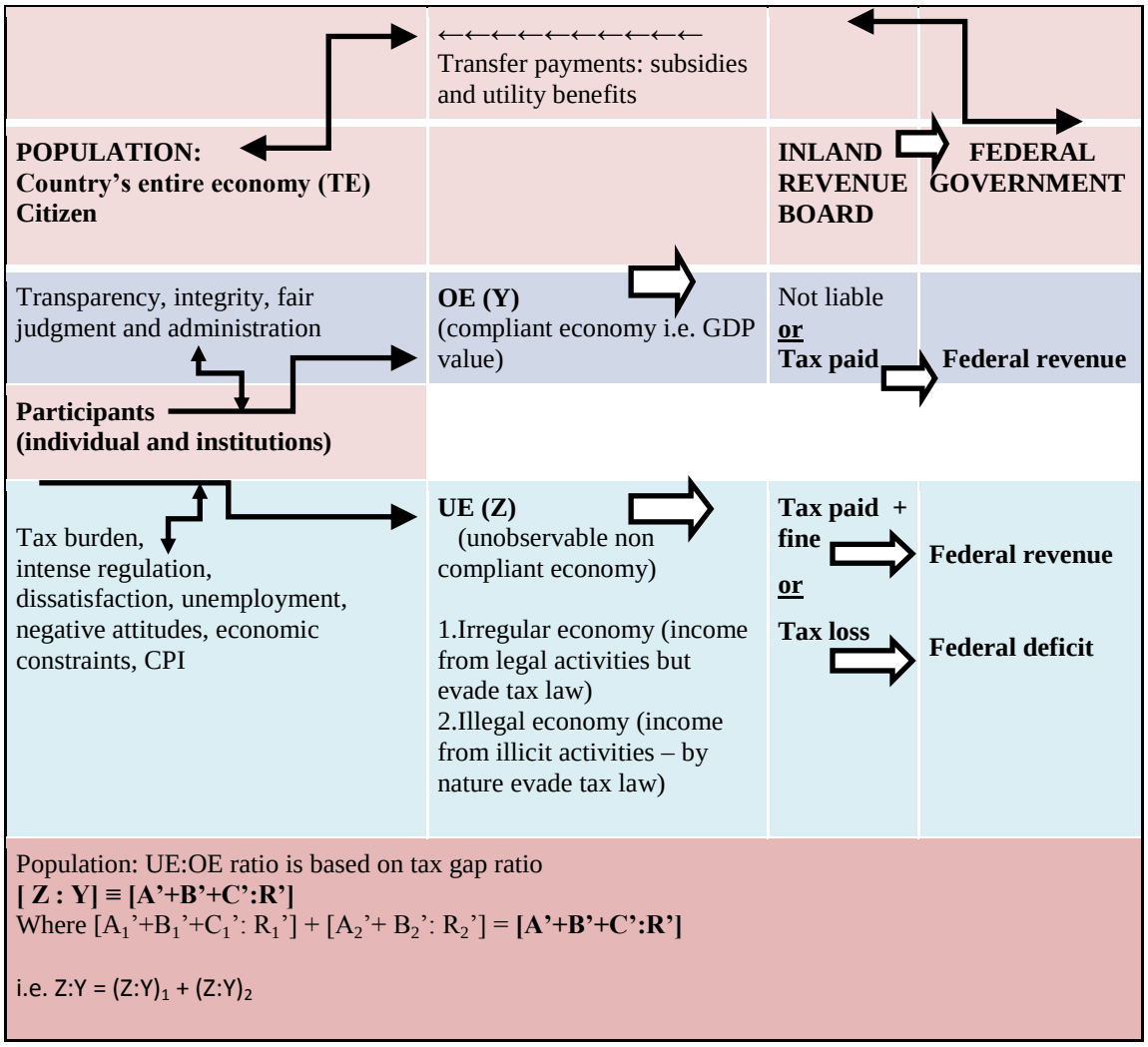

Figure 3.3: Schematic presentation of estimating UE that evade taxes - entire economy (population inferred to).

a) The OE completes with tax paid – contribute to federal revenue for the benefits of citizens and country.
b) The UE evade tax – resulting to tax loss, distort federal account and affect allocation for the citizen.
c) The country's non compliant to compliant economy is based on direct tax gap ratio – [Z: Y] ≡ **[A'+ B'+ C':R'].**

3.4 Data collection

Tax and enforcement reports of different agencies were gathered in three stages.

First, published aggregate data of determinants and proxy indicators and related information (functional relevant laws and management reports) were gathered from National records; Malaysia Economic Report, Department of Statistic Malaysia, Malaysian Central Bank of Malaysia and other relevant articles to gain initial views on activities of non- compliance.

Second, unpublished aggregate data of "voluntary tax reports" and "enforcement reports" were obtained through official application to the relevant agencies. Voluntary tax reports consist of reported income and paid taxes. Enforcement tax reports consist of unreported income and unpaid taxes. Direct tax data were obtained from Inland Revenue board of Malaysia and indirect tax from Royal Malaysian Customs, while enforcement data of illicit activities were obtained from Royal Malaysian Police and Malaysian Anti Corruption Commission.

Third, aggregate and disaggregated direct tax evasion data were obtained through official application to audit and investigation units of Inland Revenue board of Malaysia for assessing information on direct tax enforcement files (upon availability of files).

All aggregate and disaggregated data are in number of cases and amount of unreported income. The periods for the annual time series are of 1980-2009, 1998-2009 and 2000-2009, hence comparisons were made accordingly.

The entire available aggregate annual and panel data of voluntary reports and enforcement reports tax compliance were analysed. Data of different years were

analysed descriptively. Aggregate and disaggregated data of different samples were analysed for comparison, and interpreted and justified accordingly.

The aggregate data constitutes two types of enforcement statistics. First is the enforcement data of legal activities that do not comply with tax rules. It constitutes tax loss of unreported income and tax loss of unpaid reported tax within the tax net. Second is the enforcement data of illicit activities (smuggling, briberies, drug offences, gambling, betting, piracies, cyber and multimedia crimes, ATM crimes, forgeries, frauds and breach of trusts), which naturally evade tax rules to avoid detection. It constitutes unreported income outside the tax net i.e. income of whom supposed to be in the tax net yet to be captured.

To facilitate discussion on this massive collection of data, they are classified and defined as below:

1. Irregular economy: comprise of income of legal activities but irregular as it evades tax law.
a. Direct tax
i. Disaggregated audit panel data of 2005 and 2008 and investigation cases of 2005 and 2008 by characteristics of tax group (institutional and individuals), state's economy, economic industry and business sectors, age group, reported and unreported income, income level. ii. Aggregate of 2003 to 2009 annual time series data of entire tax audit and investigation cases by characteristics of tax group (institutional or individuals), state's economy, economic industry and business sectors, age group, reported and unreported income, income level. iii. Aggregate of 1997 to 2009 annual time series of entire tax audit cases and aggregate of 1980 to 2009 time series of tax evasion data of the

entire investigation cases. iv. Aggregate of 1990-2009 annual tax time series of uncollected tax data by characteristics of entire tax group (institutional or individuals) and state's economy. The unpaid tax constitutes current (not accumulative) annual tax of voluntary income reporting and some of unreported income. v. Tax gap data comprises of tax of unreported direct taxable income within the tax net, unpaid taxes and unreported taxable income outside direct tax net.
b. Indirect tax
i. Aggregate 1980 to 2009 annual time series data of smuggling activities - number of smuggling activities, value of goods smuggled and amount of tax evasion. ii. Aggregate 1990-2009 annual time series data of uncollected tax time series data by state's economy.
2. Illegal economy: Aggregate 1990 to 2009 annual time series data of commercial illicit activities as framed by regulations or criminal law.
3. Secondary 1980-2009 annual time series data of determinants and other proxy indicators of UE; cash in circulation, GDP, federal expenses, tax revenue, federal revenue, GDP per capita income, CPI, electricity consumption, unemployed labor force , crime rate, corruption perception index of for1980-2009.

3.5 Data Analyses

The aggregate and disaggregated data were analysed and discussed in 4 chapters as follows:

1. The first analysis as discussed in Chapter 4 comprises of the descriptive statistics of the proxy indicators and determinants of UE to provide some initial views for further examination and ensure proper further techniques.

2. The second analysis as discussed in Chapter 5 comprises of two stages. The first step was the examination of the characteristic of disaggregated data of unreported income to direct tax authority by states' economy, economic industries, business sectors, profession, gender, age and income background, are used to infer the characteristics of UE.

   The second step was the estimation of the size of UE in point estimate based on direct tax audit aggregated and disaggregated data by group of tax payers (institutional and individuals). Statistical technique employed was single OLS regression and matrix correlation.

3. The third analysis as discussed in Chapter 6 was the estimation of the size of UE based on two sets of data for comparison based on non-compliance ratio.

   First estimate was based on a time series data of aggregate non filing of direct tax forms over a period of 1996-2006. The proportion of non filing of tax forms or non-compliance of filing ratio was used to infer the size of UE as point estimate.

   Second estimate was based on a time series data of enforcement data over a period of 1980-2009. It consists of aggregate enforcement data of direct tax non compliance, indirect tax non compliance and commercial criminal. The income of these activities was the unreported taxable income to infer the size of UE in time series over 30 annual time series (1980-2009). It was analysed in four stages.

- The size of UE was first generated in point estimate. The finalized enforcement cases (i.e. the successful enforcement cases [SC]) were amplified by enforcement success rates ratio (ESR) to generate range estimate in three level series. The ESR was computed based on finalized enforcement cases over total enforcement cases. The annual proportion of

number of finalized prosecuted bribery over total investigated bribery cases was taken as the ESR of illicit activities. ESR for legal activities was based on finalized audit or investigation cases over total tax audits or investigation cases. The midpoint between maximum and minimum ESR over the period of enforcement data was taken as the average ESR to generate the UE size in moderate (SC x maximum ESR) and upper estimates (SC x minimum ESR).

- The estimate in absolute term was calculated based on GDP value and tax non compliant ratio of aggregate data of tax non compliance and commercial illicit activities.
- The UE growth was calculated on a year on year basis and in average of 5 time year period.
- The size of UE components was computed in percentage proportion of GDP. The composition of UE mix was divided into: economic components (irregular and illegal economy); direct tax non compliance or tax gap (tax evasion on income of legal and illicit activities [unreported taxable income], and unpaid tax [reported taxable income but not paid]), indirect tax non compliance (tax evasion [smuggling activities] and unpaid taxes).
- The amount of tax loss was computed by multiplying the estimated UE with a range of probable tax rates.

4. The fourth analysis was the estimation of the correlation coefficient between UE and priori proxy indicators or determinants in double log regression models as hypothesised in paragraph 3.9 of Chapter 3 and paragraph 7.2 of Chapter 7.

3.6 Equations of size estimates

The time series UE estimate was constructed based on the income of direct tax non-compliance, indirect tax non-compliance and income of illicit activities. The equations of estimations are as follows:

1. Sample based on an entire direct tax compliance (DTC) and indirect tax compliance (inDTC) of the existing (captured) tax base

i.	DTC = [voluntary reported direct tax] + [voluntary paid direct tax] = [Dtxrev] + [Dtxp]
ii.	inDTC = [voluntary reported indirect tax] + [voluntary paid indirect tax] = [inDtxrev] + [inDtxp]

2. Sample based on an entire captured direct tax non compliance (DTNC) and indirect tax non compliance (inDTNC)

i.	DTNC = [Direct tax evasion due to under reporting] + [unpaid DT] + [Direct tax evasion due to non filing] = [Dtxev] + [Dtxup] + [(inDtev + inDtxup) + ($UE_{crbribe}$ + $UE_{crdrugs}$ + $UE_{crothers}$)]
ii.	inDTNC = [inDirect tax evasion due to under reporting] + [unpaid inDT] = [inDtev] + [inDtxup)

3. Tax non compliant ratio (TNCR) for each activity is assumed to mirror in the UE ratio i.e. A'+B'+C': R' ≡ (Z:Y) as in Figures 3.3 to 3.5

i. TNCR due to direct tax evasion = [Dtxev : Dtxrev] ≡ UE_{Dtxev} : OE where [Dtxev : Dtxrev] is $[A_1 + C_1] : [R_{1a}]$

ii. TNCR due to unpaid direct tax = [Dtxup : Dtxp] ≡ UE_{Dtxup} : OE where [Dtxup : Dtxp] is $[B_1] : [R_{1c}]$

iii. TNCR due to indirect tax evasion = [inDtxev : inDtxrev] ≡ $UE_{inDtxev}$: OE where [Dtxev : Dtxrev] is $[A_2 + C_2] : [R_{2a}]$

iv. TNCR due to unpaid indirect tax = [inDtxup : inDtxp] ≡ $UE_{inDtxup:}$: OE where [Dtxup : Dtxp] is $[B_2] : [R_{2c}]$

4. Sum of TNCR (sample) mirrors in the non compliance ratio of the legal activities in the entire economy (UE_{TNCR}). The absolute value of legal income in the UE (UE_L) is computed based on the GDP value, and expressed as a percentage proportion of GDP

i. $UE_{TNCR} = UE_{Dtxev} + UE_{Dtxup} + UE_{inDtxev} + UE_{inDtxup}$

ii. $UE_{TNCR} : OE$

iii. $UE_L = \frac{UE_{TNCR}}{OE} \times GDP$

5. Sample based on an entire enforcement on illicit activities (cr)

As it is not possible to compute non compliance ratio of illicit activities, the incomes of these activities are added directly to the UE_L.

$(UE_{cr} = UE_{crbribe} + UE_{crdrugs} + UE_{crothers}]$

$UE_{cr} + UE_L = UE$

6. UE consists of irregular economy (UE_{L} = tax non compliance of legal activities) and illegal economy (UE_{cr} = tax non compliance of illicit activities) are estimated in point estimate for each year.

i. $\{UE_L + UE_{cr}\}_t = \{UE_{Dtxev} + UE_{Dtxup} + UE_{inDtxev} + UE_{inDtxup} + UE_{crbribe.} + UE_{crdrugs} + UE_{crothers}\}_t$

ii. $\{UE\}_t = \{UE_{Dtx(ev + up)} + UE_{inDtx(ev + up)} + UE_{cr(bribe + drugs + others)}\}_t$

7. Assuming that the point estimates represent UE in low series (Ls), amplifying them by ESR, gives UE estimates in moderate series (Ms) and upper series (Us) as follows:

i. TNC activities ≡ irregular economy
   - Ls estimate for DT evasion ≡ Ls UE_{Dtxev}
   - Ls estimate for inDT tax evasion ≡ Ls $UE_{inDtxev}$
   - Ls $_{Dtxev}$ x average ESR = Us estimate for DT evasion ≡ Us UE_{Dtxev}
   - Ls $_{inDtxev}$ x average ESR = Us estimate for inDT evasion ≡ Us $UE_{inDtxev}$

ii. Illicit activities (cr) ≡ illegal economy
   - Ls estimate of illicit activities ≡ Ls UEcr
   - Ls cr x high ESR = Ms estimate for illicit activities ≡ (Ms UEcr)
   - Ls cr x low ESR = Us estimate for illicit activities ≡ (Us UEcr)

iii. Sum of TNC activities and cr activities ≡ UE consisting of the irregular and illegal economy.
   - Ls UE_{Dtxev} + Ls $UE_{inDtxev}$ + UE_{Dtxup} + $UE_{inDtxup}$ + Ls cr = Ls UE

- Us UE $_{Dtxev}$ + Us UE $_{inDtxev}$ + UE $_{Dtxup}$ + UE $_{inDtxup}$ + Ms cr = Ms UE
- Us UE $_{Dtxev}$ + Us UE $_{inDtxev}$ + UE $_{Dtxup}$ + UE $_{inDtxup}$ + Us cr = Us UE

Simplified equation:

- Ls (cr + UE $_{Dtxev}$ + UE $_{inDtxev}$) + UE ($_{Dtxup}$ + $_{inDtxup}$) = Ls UE
- Ms cr + US (UE $_{Dtxev}$ + UE $_{inDtxev}$) + UE ($_{Dtxup}$ + $_{inDtxup}$) = Ms UE
- Us (cr + UE $_{Dtxev}$ + UE $_{inDtxev}$) + UE ($_{Dtxup}$ + $_{inDtxup}$) = Us UE

3.7 Limitations and assumptions to qualify estimates

Many researchers are skeptical about methods used and estimates generated may be arbitrary. The difficulty of measuring a latent variable consisting of large scope of activities, the uncertainty of an appropriate estimation method and the difficulty of data availability must be considered as subject limitation to reduce skeptical views over estimates.

One important advantage of using enforcement data is that they captured real activities. Tanzi (1980) stated that taxes and restrictions are the main determinants of UE. He proposed that tax non compliant data is a major UE component. Based on his statements, tax non-compliant data is an appropriate sample for estimating UE that escape taxes.

Even though tax non compliant data is a reliable sample to estimate UE that escape taxes and enforcement data reflects true non compliant event, estimate may be argued “bias or unique” as enforcement data is a stratified sample (a purposive sampling rather random).

On the low side (“under estimate”), enforcement data is not likely to capture the entire UE activities. Estimates may be far beyond actual occurrence as it excludes

income of capital profit and many other incomes. As enforcement data is far below the actual violation level, it is assumed to represent the minimal estimate. To estimate the potential UE, the enforcement data was amplified by enforcement success rates (ESR) to generate moderate and upper range estimates ("over estimate").

In other words, to account for an "under estimate" (i.e. under captured non compliance), the success or finalized enforcement data (success captured non compliance cases [SCNC]) represents the lower bound estimate (LBE). SCNC was then amplified by the upper and lower bound of ESR.

Assuming that the characteristics and distribution of finalized and non finalized cases are approximately identical, the product of SCNC or LBE and ESR increased proportionately. The products of LBE and two levels of ESR (its lower and upper bound) would generate the moderate and upper UE series. The lower bound range estimate refers to the finalized captured non compliance. The extrapolated estimates (moderate and upper range estimate) represents the actual or potential tax loss.

On the high side ("over estimate"), estimates may include redistribution income, which could have been accounted in the GDP, hence double counting of value added economy.

To reduce further doubts over unique estimates, study employed four statistical techniques on four data sets to see whether estimates approximate or differ with data set and techniques. Assuming that "under estimate" and "over estimate" mix

calculation contra each other, they should be acceptable as conservative estimates of UE that escape taxes.

Besides accounting for the consequence of "under and over estimation", estimates must be interpreted carefully and qualified based on several 'heroic" assumptions as follows:

- The tax base structure is assumed to approximate the country's economic structure so that tax non compliance ratio and characteristics of the unreported taxable income mirrors in the non compliant economy (UE).
- The incomes of indirect tax non compliance and illicit activities are assumed naturally not reported to direct tax authority because people would conceal illicit activities from authority. These incomes are assumed taxable and not reported at all. Thus these incomes are taken as the estimates of incomes of the non reporting group outside the captured tax base.
- UE and official economy have identical income tax structure.
- The unreported income is taxable and escaped taxes.
- The incomes relating to tax evasion and unpaid tax is assumed not overlapping thus no double counting.
- The unreported taxable incomes are assumed unreported value added economy ("over estimate") that is not covered by the GDP.

3.8 Priori economic association

Thomas (1999) stated that there is no clear theoretical model that specifies what variables should be included in the UE models. He criticised the macroeconomic

estimates were not based on consumer theory, thus employing flawed econometric techniques. However, many studies indicated otherwise.

Dilnot and Morris (1981), Pissarides and Weber (1989), and Lyssiotou et.al. (2004) are among researchers who encountered Thomas criticism. They used expenditure-based method on the relationship between food consumption and income (i.e. an engel curve) to estimate the size of UE.

Bhattacharya (1990) emphasised that estimates obtained in the empirical work on UE was based on an established theory. The national product (Y) is a function of various economic inputs (labor (L), capital (C), money in circulation (M), energy (E), other factors (Z)); Y = T (L,C,M,E,Z). Following an optimizing principle, the level of national product would depend on the level of inputs which could be influenced by the "formal or shadow" factor prices. The factor prices would generally rely on the push and pull factors (opportunity and risk) whether to participate in the formal or opt out entering the UE.

There are various indicator and determinant variables of UE observed in Malaysia for over 1980-2009 periods. For instance the declining tax to GDP ratio is believed to reflect a forefront indication of a sizeable UE.

It is ideal to investigate UE association with many priori variables and it is interesting to see which variables correlate most. This study proposed that UE associate with GDP per-capita, federal expenditure, electricity consumption, cash in circulation, tax burden, bribery index, criminal index and unemployment. The hypotheses in this study is based on; priori knowledge as discussed in Chapter 2 on the literature studies of UE; and the extent of UE visibility in Malaysia as

discussed in Chapter 4 on the Malaysian economy. The underlying hypotheses are summarized in paragraphs 3.9.1 to 3.9.8. The null hypothesis of the variable coefficients were tested significant at 95% confidence level or 0.05% significant error level as a function of UE in double log regression models.

Ho: $\beta i = 0$, where βi represents the coefficient of the independent variable

3.9 Hypotheses

The time series UE in range estimates of income of legal and income of illegal activities were tested whether they develop economically. The UE estimates were regressed on priori economic variables in double log multivariate models.

3.9.1 Hypothesis 1: UE associates negatively with federal expenditure

Priori knowledge indicates that federal expenditure could either promote (a positive association) or suppress (a negative association) UE activities.

A positive association implies that federal expenditure facilitates the activities of UE. Conversely, a negative association suggests that the government had intervened the UE activities by employing federal expenditure to remedy the impact of UE and dampen the UE growth. In this case the UE may cause undue strains on the government finances (for more detail see Bawly, 1982).

Malaysia has experience a declining tax to GDP ratio and an increasing growth of federal expenditure in the post 1997-1998 economic crises. Reducing tax to GDP ratio and rising growth of federal expenditure are proxy indicators of UE. This study hypothesized that federal expenditure associate negatively with UE.

3.9.2 Hypothesis 2: UE associates positively with economic crises

The adverse effects of the 1997-98 economic crises are economic slacks in the aggregate demand partly through a collapse spending, that drag on and initiated other economic structural changes. This study hypothesizes that macro post structural changes of a depressed economy are incentives of underground activities. Dummy variables (0,1) between pre and post 1997-98 period were inserted in regression model to see the effect of economic shift on UE.

3.9.3 Hypothesis 3: UE associate positively with inflation

High level of CPI implies an unstable economy and reflects a macroeconomic failure on the part of government. As CPI is a cost to consumers, it is often known as an "illegal tax". Generally, high inflation rates associate with increase of cash in circulation, low purchasing power and insufficient goods. Shortage in supply of goods and services caused increase in demand and price that leads to increases in cost of living which is an economic constrain.

This study hypothesizes that UE associate positively with CPI, as it is an incentive of underground activities because people would seek any available means of earning extra income to meet increase cost of living.

3.9.4 Hypothesis 4: UE associates positively with cash in circulation

Cash is commonly the preferred mode of payment in underground transactions because it leaves minor financial trails. High level of cash in circulation in large denomination notes has been cited as a further evidence of a flourishing UE (Bawly, 1982; Feige, 2000). In fact many studies use growth of cash traces or currency in circulation relative to demand deposits to infer UE growth. Based on

priori knowledge, this study hypothesizes that UE associate positively with cash in circulation (MYR – Malaysian currency).

3.9.5 Hypothesis 5: UE associates positively with electricity consumption

Physical resource input is essential for production of goods and services in most economic activities. Assuming that electricity usage is a good physical indicator of economic activity such that they develop a one to one and consistent relationship, its growth rate reflects economic growth. Despite Malaysia consumes more energy from petroleum sources to electricity, electricity consumption is the most versatile energy and commonly used in economic production. As UE compose of economic activities, this study hypothesizes that UE associate positively with electricity consumption.

3.9.6 Hypothesis 6: UE associates positively with GDP per capita

UE was shown in several international studies to associate positively with the official economy i.e. the GDP. A probable explanation to this relationship is that low UE level associate positively with tax revenue and federal revenue that provide more facilities to the citizen and more opportunities for public spending simulation.

A contrary view to this positive association is that UE may be more competitive and efficient than the official economy hence UE could be redeployed to boost the official economy. Cash in circulation that increased with UE could facilitate purchasing power constituting a vicious circle of economic stimulation. This study hypothesized that UE associate positively with GDP per capita (GDP controlled for population).

3.9.7 Hypothesis 7: UE associates positively with "negative indexes" - unemployment rate and crime rate and bribery index

A typical business scenario of a weak economy is a reduction of operating cost. For instance low demand for manpower would cause labor force out of jobs that lead to increases of unemployment rate. This socio-economic pressure is an incentive of job creation to substitute for income which could have otherwise earned formally. Informal jobs could extend from a "household", part timer to committing illicit activities (fraud, forge, corruption).

This study hypothesizes that UE rises with unemployment rate; bribery index; and crime index as desperate people would opt out of the official economy and enter the UE that offer informal jobs.

3.9.8 Hypothesis 8: UE associates positively with taxes

As taxes are compulsory "costs" that reduce "take home pay", people would try to reduce or avoid tax liability for "private benefit". Any dissatisfaction against the government for instance against unfair incentives or charges would add more salt to sharing income earned with taxes, because tax payment is thought to be supporting the "unpopular government". In this modern world, refusal of income to tax is believed a silent protest against the ruling government. UE is expected to increase with tax burden as part of living cost. Hence taxes are hypothesized to encourage people to evade or avoid reporting income to tax authority, a major component of UE.

3.10 Diagnostic tests

All data collected were examined in eview-6 software, to identify for any statistical evidence of structural characteristics and variable association according to the hypotheses made above.

3.10.1 Statistical tests on time series data

3.10.1.1 Unit root test

Unit roots are stationary if average is invariant with respect to time (Hanke & Reitsch, 1989), also illustrated in Perron (1989) and Phillips and Perron (1988). All time series variables and estimated UE series were resolved for stationary using the augmented Dickey-Fuller Unit root test. The importance of this test is explained in (Gujarati, 1995).

3.10.1.2 Logarithm form

Log transformation by taking the natural logarithms of all variables has also been shown adequate to remove non stationary characteristic based on the augmented Dickey-Fuller Unit root test.

3.10.1.3 Johansen co-integration test

The basic idea behind co-integration is that if the components of a time series vector have a unit root, the co-integration tests would check whether linear combination among variables are non stationary or stationary. If two non stationary components are co-integrated, they would form a linear combination in a long run relationship. Such a phenomenon is called static equilibrium relationship (Gujarat, 1995).

3.10.1.4 Variance test

The variance tests are used to compare the extent of data volatility as measured by standard deviation.

3.10.1.5 Multi-collinear effect

According to Dougherty (2002), any multi regression models will face multi collinear problems, unless all explanatory variables are uncorrelated, though some may due to sample problem. A typical phenomenon of multi-collinear effect is high coefficient of correlation matrix between variables tested. Variables of high correlation coefficient could be spurious and regressive as opposed to a true relationship between variables, resulting unreliable coefficient estimates (Gujarati, 1995).

To reduce multi collinear problems, closely related variables should be specified in different equation models. In other words a good model should consist of a set of independent variables that are of distant correlated and the variables that are of close proximity are separated and grouped into different set explanatory variables as alternative models. Researchers often quote that model is always best to be parsimonious to avoid problems of multi collinear, so reduced models are preferred to full models.

3.10.1.6 Pairwise Granger causality test

The correlation coefficient between UE and each explanatory variable was further analysed whether they extend to some causal relationship, by the Pair-wise Granger causality test at 95% significant level. This test is based on the hypothesis such that:

Hypothesis one

Ho: UE growth does not influence x_i (null hypothesis of no causal relationship)

Hypothesis two

Ho: x_i does not influence UE growth

However, as the Granger causality test is designed only to handle pairs of variables (examining the extent of causality between one variable and UE), any significant causal relationship does not guarantee a true causality. The limitation of Granger causality test may produce misleading true relationship when there are three or more variables in a multivariate regression.

3.10.1.7 Autoregressive distributed lag estimates

Economic variables may take shorter or longer time to show some positive or negative effects. In other words, a moment of period is needed to make necessary adjustments following a change in other variables due to force of habits or "waiting for optimum opportunity". Considering the possibility of delayed effect instead of immediate effect, dynamic factors were inserted to construct lag models based on Schwartz-Bayesian Criterion to estimate the number of optimal lags in this model.

3.10.2 Statistical tests on coefficients and model

3.10.2.1 Significant test

In many areas of social research, the p-value of 0.05 is customarily treated as a "border line acceptable error level" implying that the estimates obtained can be accepted with 95% confidence level. On this basis, the estimated coefficients in the OLS were all tested at the significant level of 5% probability error. The models were then simplified and evaluated for validity, adequacy and robustness fit.

3.10.2.2 Best fit test

All regression models generated are evaluated statistically based on the significance of an explanatory variable by its dynamic t value and the standardized beta coefficient. The statistical diagnostic tests used are the goodness of fit tests that include relative R^2 value for its variation explained, F-statistic for valid model; and t-value at 5% significant error level for effective coefficient.
The estimated regression coefficient was used to infer the marginal effect estimate of the explanatory variables on the dependent variable, the UE. A larger beta coefficient implies a more important explanatory variable, compared to other variables in a multivariate model equation.

Briefly beta coefficient can be defined as follows:

If βi is the regression coefficient for Xi and $\beta^* i$ is the corresponding beta coefficient, S_{UE} is the sample standard deviation of the dependent variable UE, and Sxi is the sample standard deviation of the explanatory variable xi, then,
$\beta^* i = \beta i Sxi / S_{UE}$
where β^* is the estimated coefficient for the explanatory variable and
βi is the estimated regression coefficient for the explanatory variable

To ascertain whether models with significant coefficients are valid and robust, the models are diagnosed for serial correlation, heteroskedascity and model error specification (abnormalities), by Breusch-Godfrey – LM test, Breuch-Pagan Godfrey and Ramsey Reset tests, respectively. Models of best fit tests (no evidence of "abnormalities") with priori signs and small residual errors based on AIC and SIC values are recommended for policy measure.

3.11 Double log regression models and equations

The UE estimated series were then examined for economic association in double log regression models instead of in level form to reduce the problem of heteroscedasticity and to generate reliable estimates of practical values. The coefficients in double log equations could be interpreted as economic associations in terms of buoyancy, elasticity and growth ratio. Coefficients constructed in double log models give better stability, less variability and more consistency, thus equation would be more properly specified, (Gujarati, 1995), hence useful for policy measures.

UE models are constructed by regression of the UE estimated series as the dependent variable on independent variables as hypothesized in paragraph 3.9 as a function of a set of explanatory variables as below:

UE = f{GDP per capita, CPI, Cash in circulation, M1 money aggregates, electricity consumption, federal expenditure, taxes, economic crises and negative indexes}

The UE series were first examined in the uni-variate double log models. UE generate significant positive coefficients at 95% confidence level or 5% significant error with the explanatory variables (GDP per capita, tax burden, federal expenditure and consumer price index cash, electricity consumption, crime index, corruption perception index). Some coefficients are only significant after having adjusted for AR insertion (uni-variate regression), an indication of short term buoyancy relationship. The positive relationships between UE and the variables are consistent with priori as evidence of buoyancy relationship.

UE series were then examined in multivariate regression to evaluate its elasticity and collective interactions among the explanatory variables. Although the growth of both GDP nominal and GDP real were in the same direction, they were of different magnitude as the GDP real was deflated by inflationary factors.

To evaluate whether UE was strongly influenced by inflation, the UE estimated series were regressed in two sets; in nominal level model and real level model. The real data were obtained by deflating the nominal data with CPI (1980=100). Assuming CPI is the main GDP deflator, coefficients of real variable models could be interpreted as correlation eliminating for inflation element.

Summary

Past studies have used various direct and indirect methods to estimate the size of the "second economy" and its portion, the UE. The employment of various data to uncover a wide scope of activities and making heroic assumptions have resulted a wide range of estimates. Comparison among estimates remains crude and estimates must be interpreted carefully and qualified.

The aim of this study is to estimate the size of UE (the income of legal activities [irregular economy] and the income of illicit activities [illegal economy]) that escape taxes which is of any country's treasury concern. The conceptual framework of this study is based on the concept that people would conceal income of illicit activities from any authorities to avoid detection, in this case evade taxable income from tax authority. This study employed five statistical techniques on four sets of data samples to reduce skeptical view over unique

estimates and to evaluate whether estimates generated approximate or differ largely.

The estimated of UE size were computed in relative to GDP based on TNC and computed in point and range time series estimates by correlation matrix, regression coefficients, proportion ratio and amplifying techniques. Assuming that the income of UE activities were taxable, the potential tax loss of UE or UE that escape taxes would be the product of UE and tax rates. The characteristics of direct tax evasion were used to infer the characteristics of UE.

The estimated UE series were then tested in multivariate models on priori variables (GDP per capita, CPI, cash in circulation, federal expenditure, dummies of economic crises, electricity consumption, crime rate, unemployment rate and bribery rate). The correlation coefficients were statistically tested and the double log regression models were evaluated for best fit statistical tests. Any significant economic association is used to infer inter-temporal development to suggest that the UE models could be established to formulate right and effective policy measures.

Like other studies, estimates must be acceptable according to assumptions made considering the limitation of a latent variable and data availability. The assumptions made to justify and qualify estimates as inference drawn from TNC data are; approximate country's labor force structure and tax base structure; tax structure resembles country's economy; income of tax evasion is largely due to income omission that gives comfort of assuming it representing the value added economy instead of non redistribution one; income of different non compliant

activities sum up to total income that evade tax; unreported taxable income constitutes a large portion of UE; unreported taxable income is synonymous to unreported income.