CHAPTER 3

METHODOLOGY

3.1 Introduction

This chapter begins by discussing the study's conceptual framework of this research. It further discusses the theoretical framework to attain the four objectives of this study. The flow of data analysis and presents the evaluation techniques for the identified variables and determinants are presented at the end of the chapter.

3.2 Research Paradigm

Research paradigms can be clarified by referring to the philosophical investigation of three distinct components, ontology, epistemology and methodology, (Denzin & Lincoln, 2011; Creswell, 2013). This implies that researchers start a project with some expectations about how to learn and what to learn during their investigation (Creswell, 2013).

Ontology is interested in the nature of reality or what scholars query about knowledge (Creswell, 2013). Ontology is the study of what is in every field of reality, of the kinds and structures of objects, properties, events, processes, and relationships (Smith, 2003). In other phrases, ontology asks how the universe is viewed by us. Hence, in view of economic ontology is the correspondence between expectations and fulfilment in realities.

As such, the economic ontology in this study focused on the objectivism which holds the truth and misrepresentation are aspects of conceptual knowledge. Objectivism is based on the assumptions that social phenomena exist in everyday discussions that is independent or separate from actors (Bahari, 2010). Also, it is believed that genuine empirical

knowledge involves universal logical structures of inferences which results can be tested against theory-neutral 'objective' data."

On the other hand, the general assumption about what constitutes "valid" and "good" research is the concerned in every research conducted. Those philosophical assumptions are related to the underlying epistemology that guides the research. Epistemology is dealing with metaphysical arguments about the way the universe is known or revealed to us (Hughes & Sharrock, 1997). In other words, epistemology is the question of how the universe that points to the science of knowledge can be examined (Babbie, 2012).

Several scholars have come up with different epistemological model paradigms. Tschamler (1996), for example, suggests six philosophical research streams: the phenomenological-hermeneutic approach, the empirical-analytical approach, constructivism, critical rationalism, radical constructivism, and critical theory. In contrast, Guba and Lincoln (1994) mention four epistemological approaches especially applicable in a qualitative research context: positivism, post-positivism, critical theory, and constructivism. Orlikowski and Baroudi (1991), on the other hand, suggest three research epistemology streams: positivist, interpretive and critical.

Nevertheless, these distinctions are not always so straight cut in social research (see, e.g., Lee 1989). It has to be mentioned that the epistemological approaches is beyond the focus of this thesis since the detailed analysis and examination of various epistemological approaches usually related to qualitative study. For instance, there is well-cited qualitative work that has an underlying positivist research approach i.e. the case study research by Yin (2002). However, it also applicable to quantitative study by understanding the definition of positivist.

Positivist scholars generally aim to translate this objective reality into testable hypotheses (typically in the form of statistical or other numerical analyses). In this way they attempt to increase the predictive understanding of phenomena and thus produce knowledge (Myers et al., 2004). Hence, this study stated several testable hypotheses in section 3.3.2.

Meanwhile, methodology is described as data collection procedures, defining, explaining and predicting phenomena using methods or techniques selected (Bryman & Bell, 2003; Bryman 2012; Creswell 2013). In other words, it is the method of investigating it (Creswell 2013) or the science of discovering it (Babbie 2012). It is also an attempt in answering the query related to how we should best collect data. Then, the methodology used in this study is discussed through secondary data collection which detailed out in Table 3.1, .3.2 and 3.4. While the procedures for data cleaning and estimation techniques used in analysing the testable hypotheses are explained in section 3.4 and 3.5 accordingly.

3.3 Research Framework

3.3.1 Conceptual Framework

The conceptual framework as shown in Figure 3.1 is derived from the theoretical underpinning discussed in Chapter 2. It is necessary to provide an understanding of how household debt plays a key role in influencing economic stability in four aspects: i) the impact of household debt on economic growth (see Figure 3.1, RO1), ii) the role of institutional quality in the household debt-growth nexus (see Figure 3.1, RO2), iii) the role of financial development in the debt-growth nexus (see Figure 3.1, RO3), and iv) the role

of household debt as an indicator for predicting systemic banking crises (see Figure 3.1, RO4).

First, this dissertation investigates the role of household debt in the growth model based on the endogenous growth model of Barro and Sala-i-Martin (1997). Second, the growth model is extended to include the role of institutional quality. Third, financial development as a determinant of household debt is discussed based on the frameworks of life cycle hypothesis (LCH) and permanent income hypothesis (PIH). The new factors are postulated as the major contributors to the household debt-growth nexus along with other explanatory factors. Lastly, this study examines the role of household debt in the EWS as a predictor for systemic banking crises. The diagram in Figure 3.1 demonstrates the leading indicators for systemic banking crises. To sum up, the current research is interested in investigating the household debt-growth nexus as well as its cause and impact.

3.3.2 **Main Research Hypotheses**

The testable hypotheses, which are implied in the research questions, are as follows:

H1: Household debt has a negative effect on economic growth

H2: Institutional quality has a positive influence on economic growth

H3: The interaction term between institutional quality and household debt play a significant role in the relationship between household debt and economic growth

H3: Financial development has a significant influence on the growth of household debt

H4: Household debt increases the probability of systemic banking crises in the EWS model

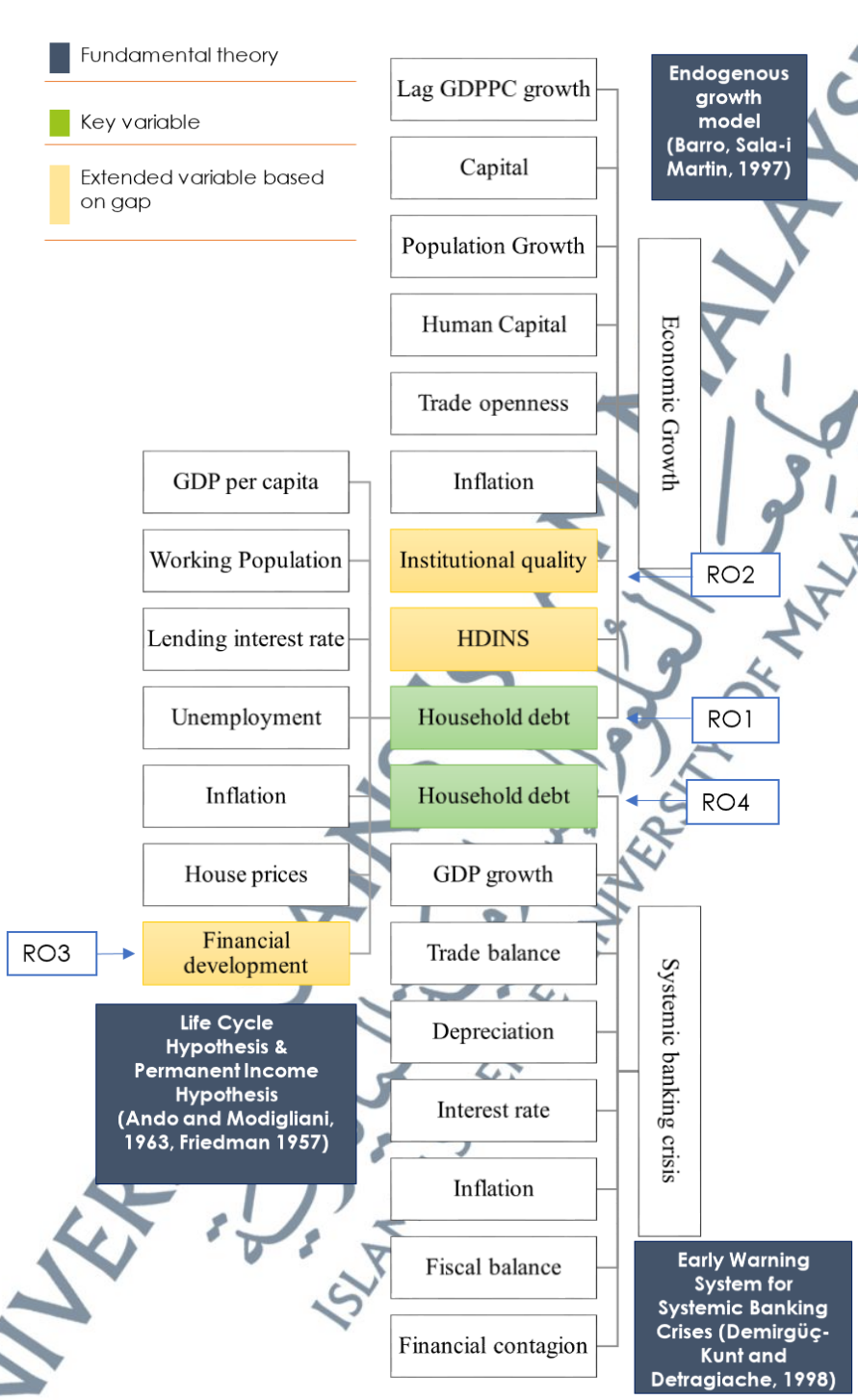

Figure 3.1: Conceptual Framework

3.3.3 Household Debt and Economic Growth

The first model relates to the first and second research objectives (RO1 and RO2), which aim to examine the impact of household debt and institutional quality on economic growth. Following both the relevant economic theory and previous empirical results, this study's strategy is to incorporate the specification and estimation of a growth equation based on the growth literature (e.g., Mankiw et al., 1992, and simplified version by Law and Habibullah, 2006) augmented by technological innovation and human capital as described in Equation 2.2. Consider the following Cobb-Douglas production function:

$$Y_t = K_t^{\alpha} H_t^{\beta} (A_t\, L_t)^{1-\alpha-\beta} \tag{3.1}$$

Where *Y* is real output growth, *K* is the capital stock, *H* is the stock of human capital, *L* is the raw labour, A is a labour-augmenting factor reflecting the level of technology and efficiency in the economy and the subscript *t* indicates time. Assuming that $\alpha + \beta < 1$, which implies that there are decreasing returns to all capital. Raw labour and labour augmenting technology are assumed to grow according to the following functions:

$$L_t = L_0 e^{nt} \tag{3.2}$$

$$A_t = A_0 e^{gt+X\theta} \tag{3.3}$$

where n is the exogenous rate of growth of the labour force, g is the exogenous rate of technological progress, X is a vector of other factors that can affect the level of technology and efficiency in the economy, and θ is a vector of coefficients related to these variables.

In this model, variables *A* depends on exogenous technological improvements, the degree of openness of the economy and level of other variables.

$$\frac{Y_t}{A_t\, L_t} = (k_t\ )^{\alpha}(h_t)^{\beta} \tag{3.4}$$

$$\frac{Y_t}{L_t} = A_t(k_t)^{\alpha}(h_t)^{\beta} \tag{3.5}$$

The neoclassical model emphasizes that the evolution of the economy is determined by the following equations:

$$h_{i,t} = s_h f\left(k_{i,t}, h_{i,t}\right) - \ (n + \mathrm{g} + \delta)\, h_{i,t} \tag{3.6}$$

$$k_{i,t} = s_k f\left(k_{i,t}, h_{i,t}\right) - \ (n + \mathrm{g} + \delta)\, k_{i,t} \tag{3.7}$$

Where *y= A/L*, *k= K/L*, *h = H/L* are quantities per effective unit of labor, stand for the growth per capita, capital labour ratio, average human capital and output per worker, respectively. s_h stands for the fraction of human capital and s_k is the fraction of income investment in physical capital, and δ is the rate of depreciation. The steady state values of h and k are defined as

$$h^* = \left[\frac{(s_k)^{\alpha}\ (s_h)^{1-\alpha}}{(n + \mathrm{g} + \delta)}\right]^{1/1-\alpha-\beta} \tag{3.8}$$

$$k^* = \left[\frac{(s_k)^{1-\beta}\ (s_h)^{\beta}}{(n + \mathrm{g} + \delta)}\right]^{1/1-\alpha-\beta} \tag{3.9}$$

Substituting these two equations into the production function (Equation 3.5) and taking logs yields the expression for the steady state output.

$$\ln\left(\frac{Y}{L}\right) = \ln A + \alpha \ln k + \beta \ln h \tag{3.10}$$

Where $A_t = A_0 e^{gt+X\theta}$

$$\ln\left(\frac{Y}{L}\right) = \ln A_0 + \mathrm{g}t + X\theta - \frac{\alpha+\beta}{1-\alpha-\beta}\ln(n+\mathrm{g}+\delta) + \frac{\alpha}{1-\alpha-\beta}\ln s_k + \frac{\alpha+\beta}{1-\alpha-\beta}\ln s_h \tag{3.11}$$

Equation 3.11 shows how income growth per capita depends on population growth and accumulation of physical and human capital. It indicates steady state output per worker where a vector of household debt and institutional quality proxies exist; s_k and s_h are the shares in output of physical capital K and human capital H; and $(n + \mathrm{g} + \delta)$ is the population growth, rate of technology growth and technological progress rate of depreciation. Rearranging Equation 3.12, it yields an estimating equation for the relationship between household debt and growth as follows:

$$\ln y = \beta_0 + \beta_1 \ln K - \beta_2 \ln(n + \mathrm{g} + \delta) + \beta_3 \ln H + X\theta + T \tag{3.12}$$

Where y is growth of real GDP per capita, K is capital stock per capita, $(n + \mathrm{g} + \delta)$, n is the rate of labour growth using population growth, g is the rate of technology growth and technological progress, δ is the rate of depreciation and T denotes time dummy. The addition of g and δ is assumed to be constant across countries and over time, following the Mankiw et al. (1992), is equal to 0.05. X denotes the household debt and other explanatory variables. The technological improvements are encouraged by developments in financial institutions offer more lending to household, which tend to increase the productive sector's efficiency or increase the productivity of investment (Pagano, 1993). Following the

convention in the literature (Cecchetti et al., 2011; Gómez-Puig & Sosvilla-Rivero, 2017), this study incorporates household debt in the growth model.

$$\ln GDCPPCG = \beta_0 + \beta_1 \ln GCF - \beta_2 \ln(POPG) + \beta_3 \ln HC + \beta_j X + T \qquad (3.13)$$

In this study, where $GDPPCG$ is real GDP per capita growth (y), GCF is capital stock per capita (K), $POPG$ denotes $(n + g + \delta)$ and HC stands for human capital. X stands for other explanatory variables. In addition, T denotes time dummy and this study analysed the regression model with the banking crises dummy ("CRISIS") measured using systemic banking crises. As suggested by Laeven and Valencia (2013), it takes the value of zero if there is no banking crisis in the subsequent five years, and the values of 1/5, 2/5, and so forth if a banking crisis occurs in one year, two years, and so forth in the subsequent five years.

Following the Barro and Sala-I-Martin (1997), there is need for initial income that covers level of convergence. Hence the model includes y_{t-1} and all variables are in logarithm forms.

$$y_{i,t} = \gamma y_{i,t-1} + \beta_j X_{i,t} + \varepsilon_t \qquad (3.14)$$

where y_{t-1} is the one-period lagged real GDP per capita growth for country *i* at time *t* with γ to capture the conditional convergence of the economy to its steady state. The model incorporates a one-period lagged risk measure, in line with Cecchetti et al. (2011), Lombardi et al. (2017), Gómez-Puig and Sosvilla-Rivero (2017), and Bahadir and Valev (2020), among others. β_j are the parameters to be estimated and X_{it} is a set of explanatory regressors of the independent variables, namely capital, population growth, human capital, trade openness, and inflation. HD_{it} is the household debt-to-GDP ratio at time *t* for country

i, following Cecchetti et al.'s (2011) debt-growth model on panel dataset of countries. ε_{it} represents the error term and *i* represents the country-specific effect.

In this study, debt is considered as a critical factor that contributes to sustained economic growth. According to Ando and Modigliani's (1963) LCH, debt is an apparatus that maintains a stable level of consumption over an individual's life cycle. The hypothesis postulates that a household makes a loan when earning is lower than expected except when there is an unexpected increase in income or during its most productive working days. The theory highlights the importance of household debt in encouraging economic growth. The final form of the household debt-growth model in the present study is

$$GDPPCG_{it} = \beta_0 + \beta_1 Y_{i,t-1} + \beta_2 GCF_{it} + \beta_3 POPG_{it} + \beta_4 HC_{it} + \beta_5 TO_{it} + \beta_6 INF_{it} + \beta_7 HD_{it} + \varepsilon_{it} \quad (3.15)$$

where Y_{it} is real GDP per capita growth, Y_{t-1} is one-period lagged real GDP per capita growth for country *i* with β_1 to capture the conditional convergence of the economy to its steady state, and $\beta_2 \ldots . \beta_7$ represent the coefficients of independent variables including gross capital formation, population growth, human capital, trade openness, inflation, and household debt-to-GDP ratio at time *t* for country *i*. ε_{it} represents the error term and *i* represents the country-specific effect.

Crucial to the analysis is an extension of the framework of economic growth to cover institutional quality as an explanatory variable in order to empirically analyse the impact of household debt on economic growth. According to Acemoglu et al. (2003), many banking crises were caused by weak political and economic institutions, leading to macro-economic turmoil (Demirgüç-Kunt & Detragiache, 1998). An immediate implication is that

poor institutional quality can reduce the welfare of a country (Huang & Wei, 2006). Klein (2005) indicated that better institutional quality exhibits an economically meaningful and statistically significant impact of capital account openness on economic growth. High institutional qualities provide secure environments and policies leading to economic development (Kim & Loayza, 2017). In addition, the A_t in Equation 3.3 contains control variables that affect growth, such as human capital, policies, and institutions, as identified in the literature (Cooray et al., 2014). This thus accentuates the importance of institutional quality in supporting economic growth. Hence, the final form of the household debt-growth model in the present study with the incorporation of institutional quality is

$$GDPPCG_{it} = \beta_0 + \beta_1 Y_{i,t-1} + \beta_2 GCF_{it} + \beta_3 POPG_{it} + \beta_4 HC_{it} + \beta_5 TO_{it} + \beta_6 INF_{it} + \beta_7 HD_{it} + \beta_8 INS_{it} + \varepsilon_{it} \quad (3.16)$$

where Y_{it} is real GDP per capita growth, Y_{t-1} is the one-period lagged real GDP per capita growth for country *i* with β_1 to capture the conditional convergence of the economy to its steady state, and $\beta_2 \ldots . \beta_8$ represent the coefficients of independent variables including gross capital formation, population growth, human capital, trade openness, inflation, household debt-to-GDP ratio, and institutional quality at time *t* for country *i*. ε_{it} represents the error term and *i* represents the country-specific effect.

In addition, the study investigates the interaction term between household debt and institutional quality. Lombardi et al. (2017) found that creditors' level of legal protection served as an indicator for institutional quality which caused lower levels of economic growth to correspond to higher levels of household indebtedness. Their finding implies that households borrow as they possess a strong faith and confidence in the institutions.

Furthermore, households' optimism about future income is possible as they feel secure about the country's stability. Bahadir and Valev (2020) extended the household debt-growth model by incorporating institutional quality. The study showed that institutional quality was significant in mediating the negative effect of household debt on growth. Thus, this study is interested to examine the interaction term between household debt and institutional quality in the model.

$$GDPPCG_{it} = \beta_0 + \beta_1 Y_{i,t-1} + \beta_2 GCF_{it} + \beta_3 POPG_{it} + \beta_4 HC_{it} + \beta_5 OPEN_{it} + \beta_6 INF_{it} + \beta_7 HD_{it} + \beta_8 INS_{it} + \beta_9 HD * INS_{it} + \varepsilon_{it} \quad (3.17)$$

where Y_{it} is real GDP per capita growth, Y_{t-1} is the one-period lagged real GDP per capita growth for country *i* with β_1 to capture the conditional convergence of the economy to its steady state, $\beta_2 \ldots . \beta_9$ represent the coefficients of independent variables including gross capital formation, population growth, human capital, trade openness, inflation, household debt-to-GDP ratio, institutional quality, and the interaction term between household debt and institutional quality at time *t* for country *i*. ε_{it} represents the error term and *i* represents the country-specific effect.

As explained above, the aim of the study is to examine empirically the hypothesis that institutional quality and household debt are complementary with respect to enhancing economic growth. Therefore, Equation 3.17 provides the basis for the empirical model by interacting institutional quality and household debt at minimum and maximum values of institutional quality as a supporting function for household debt in accelerating growth.

According to Brambor et al. (2006), the interpretation of the coefficients for β_7 and β_8 of Equation 3.17 might be misleading in a model with an interaction term. The

coefficient β_7 captures the effect of household debt only when institutional quality is zero. Similarly, β_8 captures the effect of institutions on growth only in the absence of household debt. Thus, the results should not directly interpret the direct signs of β_7 and β_8 and their significance on the relationship between household debt, the interaction term, and economic growth. Hence, Brambor et al. (2006) posited that the marginal effect, which captures the role of institutional quality in supporting the effect of household debt on economic growth can be calculated by examining the partial derivative of economic growth, as follows:

$$\frac{\partial GDPPCG_{it}}{\partial HD} = \beta_7 + \beta_9 INS_{it} \tag{3.18}$$

Equation 3.18 states that a change in economic growth due to a change in household debt depends on household debt (HD) itself and institutional quality (INS). If the interaction term $HD \times INS$ is positive and significantly related to economic growth, and *HD* alone is significantly negative, then this supports the view that the effect of *HD* on economic growth could be reduced only if the institutional quality achieved a certain minimum level. At the margin, the total effect of increasing household debt due to institutional quality can be calculated by examining the partial derivative of economic growth with respect to the household debt variable in Equation 3.18. The study follows Brambor et al. (2006) where the significance of the marginal effect of household debt on economic growth is calculated based on the standard error.

3.3.3.1 Data and the Variables of the Economic Growth Model

The panel dataset used in this study possesses two main dimensions, namely a cross-sectional dimension consisting of 43 countries and a time series dimension spanning from 1980 to 2018. For estimation purposes, data were collected from several sources such as BIS, DataStream, and the World Bank database. The selected variables are based on the theoretical framework of the research objectives and the novelty of this study.

The factors that affect the changes in economic growth are gross capital formation, population growth, human capital, trade openness, inflation, household debt, and institutional quality. The dependent variable, economic growth is proxied by real GDP per capita growth. In previous studies on economic growth, real GDP per capita growth was found to be a suitable indicator (see Barro, 2013; Cecchetti et al., 2011; Gómez-Puig & Sosvilla-Rivero, 2017; Puente-Ajovín & Sanso-Navarro, 2015; Woo & Kumar, 2015). GDP per capita growth is defined as the growth of gross domestic product per capita in constant prices, acquired from World Bank Statistics.

3.3.3.2 Capital

Gross capital formation as a percentage of GDP is used as a proxy for investment, which affects economic growth. Data for this variable were also acquired from the World Bank database. According to Solow's neoclassical growth model, the growth in per capita income is dependent on the initial level of human and physical capital, and it slowly converges over time to its steady state rate, which in turn positively depends on savings rate and negatively depends on labour force's growth rate combined with country preferences and other technological parameters. On the link between capital investment and

economic growth, Harrod (1939), Domar (1946), and Solow (1956) proposed that gross capital investment positively affects economic growth. According to Keynes (1936), the increased investment in amplified savings encourages financial progression. Some studies used gross capital formation as proxy for investment (Park et al., 2019). Thus, economic growth is expected to be positively influenced by investment, which is this study is proxied by gross capital formation (GCF).

3.3.3.3 Population growth

Several studies used population growth as a proxy for the level of human capital (see Cecchetti et al., 2009; Gómez-Puig & Sosvilla-Rivero, 2017). Data on this variable were mainly retrieved from the World Bank database. Researchers have suggested a positive link between economic growth and population growth. As such, population growth (POPG) is expected to negatively affect the growth.

3.3.3.4 Human capital

The study employs life expectancy at birth as a proxy for human capital following some studies suc as Gómez-Puig and Sosvilla-Rivero (2017) and Sachs and Warner (1997). Data for this variable were retrieved from the World Bank database. As shown in Jayachandran and Lleras-Muney (2009), a longer life expectancy encourages human capital accumulation, since a longer time horizon increases the value of investments that pay out over time. Moreover, better health and education are complementary with a longer life expectancy (Becker, 2007). Indeed, life expectancy at birth correlates strongly with the index of human capital per person provided by the Penn World Table version 8.0 (Feenstra

et al., 2013), years of schooling (Barro & Lee, 2013), and returns on education (Psacharopoulos, 1994). While some studies found a positive relationship between human capital and economic growth (Radelet, Sachs, & Lee, 2001), others discovered a negative relationship (Barro, 2003). Consequently, the effect of human capital on economic growth is expected to be either positive or negative.

3.3.3.5 Trade openness

The sum of export and import as a percentage of GDP proxies for trade openness based on many studies for exxample Cecchetti et al. (2009), Gómez-Puig & Sosvilla-Rivero (2017) and Woo and Kumar (2015). Data were gathered from the World Bank database. Past studies, Edwards (1998) and Yanikkaya (2003) found evidence that trade openness attracted more capital investments. Thus, it is expected that the relationship between trade openness and growth is positive since most previous studies reported a positive link between these variables (Barro & Sala-i-Martin, 1997; Frankel & Romer, 1992).

3.3.3.6 Inflation

The inflation rate is proxied by consumer price index following the Barro (2013). Inflation measures the uncertainty of a country's economy. It has been found to have a negative relationship with economic growth (Cecchetti et al., 2009; Gómez-Puig & Sosvilla-Rivero, 2017). Barro (2013) found that while growth may be adversely influenced by inflation, the standard of living may be affected substantially in the long run. Hence, it is expected that inflation has negative relationship with growth.

3.3.3.7 Household debt

Cecchetti et al. (2011) and Lombardi et al. (2017) used household debt ratio to GDP as an indicator for housheold debt. Household debt is used a tool to smoothen consumption and engineer the economic activities of a country. Even though it is expected to have a positive effect on growth, eventually it poses a negative effect. Past studies concurred on the negative relationship between economic growth and debt long run (Cecchetti et al., 2011; Lombardi et al., 2017). Thus, household debt is expected to have a negative sign. Data for household debt to GDP were retrieved from the BIS database.

3.3.3.8 Institutional quality

Institutional quality is represented by five indices which closely interpret the quality of institutional factors that may strengthen economic growth along with rising household debt. The indices used are bureaucracy quality (BQ)[4], corruption (COR)[5], government stability (GS)[6], law and order (LO)[7], and democratic accountability (DA)[8]. These indices are sum up and represent the institutional quality (INS) which used in the literature to measure institutional quality and have previously been shown to affect favourably on

[4] Bureaucratic quality refers to an assessment of the strength and expertise of the bureaucracy to govern without drastic changes to policy or interruptions in government services. In countries that score well on this index, the bureaucracy tends to be autonomous from political pressure (maximum score: 4 points).
[5] Corruption refers to an assessment of the level of corruption within the political system (maximum score: 6 points).
[6] Government stability refers to an assessment of the government's ability to execute its declared programmes, as well as its ability to stay in office (maximum score: 12 points).
[7] Law and order refers to the law subcomponent that measures the strength and impartiality of the legal system, while the order subcomponent is an assessment of the popular observance of the law (maximum score: 6 points).
[8] Democratic accountability refers to an assessment of how well the government responds to its people (maximum score: 6 points).

economic growth (Khan et al., 2019; Law et al., 2017). The indices were obtained from The International Country Risk Guide database (ICRG).

Table 3.1: List of Variables for RO1 and RO2

RO1: To examine the effect of household debt on economic growth in the investigated countries. **RO2: To examine the role of institutional quality in the relationship between household debt and economic growth in the investigated countries.**				
Variables	**Abbreviation**	**Description**	**Expected Sign**	**Source**
Dependent Variable				
Economic Growth	GDPPCG	Growth of real Gross Domestic Product per capita of the country.		The World Bank
Independent Variables				
Gross capital formation	GCF	Gross capital formation as a % of GDP of the country.	+	The World Bank
Population	POPG	Population growth.	-	The World Bank
Human capital	HC	Life expectancy of birth	+/-	The World Bank
Trade openness	TO	Sum of export and import % of GDP.	+	The World Bank
Inflation rate	INF	The consumer price indexes.	-	The World Bank
Household debt	HD	Credit to the household of the country.	-	BIS database/ DataStream
Institutional quality	IQ	Bureaucracy Quality (BQ), Corruption (COR), Government Stability (GS), Law and Order (LO), and Democratic Accountability (DA)	+	ICRG database

3.3.4 **Determinants of Household Debt**

A starting point for examining trends in household debt borrowing is presented by the life cycle and permanent income model of Modigliani (1986) and Friedman (1957). It refers to household choice of consumption path to maximise utility over its lifetime, subjected to intertemporal budget constraints. Tudela and Young (2005) used Overlapping Generation (OLG) model developed by Barnes and Young (2003) they proposed a framework for understanding aggregate indebtedness regarding individual optimising decisions and further adopted the model to explain the rise in borrowing.

The model introduced consumption-income and housing-finance motives for borrowing adapted life cycle of household consumption behaviour with standard constant relative risk aversion (CRRA) preference in a partial equilibrium overlapping generation model framework. The model consists of motives; household borrowing constraint, household behaviour and aggregation (Barnes & Young, 2003).

First, the equation explained the household consumption motive. The path of desired consumption over the life cycle is derived from CRRA utility function. The utility at date t of a household with N future period to live is:

$$U_t^N = \sum_{i=0}^{N} \beta^t \frac{1}{1-\gamma} (h_{t+i}^{\alpha} c_{t+i}^{1-\alpha})^{1-\gamma} \tag{3.19}$$

Utility U is derived from two non-durable goods, housing services h and non-housing consumption goods c, α is a households have Cobb-Douglas preferences with taste parameter. Utility is time-separable δ for each period, which could vary over time in response to changes in household composition; the discount factor is $\beta = \frac{1}{1-\delta}$. It is

assumed that households derive no utility from past consumption. γ is the coefficient of relative risk aversions.

Secondly, the housing-finance motive for borrowing arises from need to fund for home ownership. Hence, in this model, households may hold their wealth in the form of financial assets or housing wealth. The financial asset a is a net asset that either accumulated or borrowed at a single fixed nominal interest r. The value of housing wealth at the end of each period is equal to the number of units of housing h multiple by their price is p. Household disposable income is y. The flow balance sheet constraint at date $t+i$ is given by:

$$a_{t+i} = y_{t+i} + (1 + r_{t+i-1})a_{t+i-1} + q_{t+i}(1-d)h_{t+i-1} - p_{t+i}c_{t+i} - q_{t+i}h_{t+i} \quad (3.20)$$

The third motive is consumption-income, household adjust their stock of net wealth by saving and dissaving to smooth their consumption. The difference between current income and desired expenditure are from both the path of desired consumption over the life cycle and the pattern of income. Household labour income consists of two components; an age-related premium and a component related to aggregate income.

Fourth motive, old age borrowing constraint indicates households with no uncertainty or a bequest motive would commonly choose to be indebted but solvent at death, thus, allowing their estates to repay the debt from selling the house. Although borrowing constraints are more often associated with the young, low and declining net worth, as well as low presence credit market frictions. Old age borrowing constraint which precludes households from borrowing in the final period of their lives:

$$a_{t+N} \geq 0 \tag{3.21}$$

Household behaviour: Households' optimal non-housing consumption is derived by maximising Equation 3.19 subject to (3.20) and (3.21). Relative consumption of goods and housing services in each period is then given by:

$$\frac{h_{t+i}}{c_{t+i}} = \frac{\alpha}{1-\alpha}\frac{p_{t+i}}{R_{t+i}} \tag{3.22}$$

where R_{t+i} is the (nominal) user cost of housing. In every period apart from the last period of life, it is given by:

$$R_{t+i} = \frac{q_{t+i}}{(1+\rho^{q}_{t+i})}(\rho^{q}_{t+i} + d) \tag{3.23}$$

where $1 + \rho^{q}_{t+i} = (1 + r_{t+i})q_{t+i}/q_{t+i+1}$, is (one plus) the real interest rate where inflation is measured by expected house price inflation. In the last period of life, the nominal user cost of housing is simply given by the stock price of housing $R_{t+N} = q_{t+N}$. As households have no motive to hold assets to yield utility in the future, the no rental housing constraint imposes a real cost from having to hold housing assets to enjoy housing services. This change in costs implies that households will modify their behaviour.

Aggregation: The aggregate pattern of household behaviour can be derived by aggregation across generations n_i as we take a partial equilibrium approach. Each cohort's weighting in the aggregate is determined by the variation in the formation of new young households m:

$$n_{1,t} = m_t n_{1,t-1} \tag{3.24}$$

Aggregate debt holdings are then given by:

$$D_t = \sum_{i=1}^{i=10} n_{i,t}\, d_{i,t} \tag{3.25}$$

where aggregate debt D_t depends on the structure of the population and the debt d of each cohort, where d is the absolute value of the negative asset holdings of each cohort.

Following both the relevant economic theory and existing empirical evidence, this study incorporates the specification and estimation of a household debt equation based on the LCH and PIH (Ando & Modigliani, 1963; Friedman, 1957), stating that household debt (HD) is a function of GDP per capita growth (GDPPC), unemployment (UN), working population (WPOP), interest rate (LIR), inflation rate (INF), household consumption (CON), and housing price index (HPI).

$$HD = f(GDPPC, UN, WPOP, LIR, INF, CON, HPI) \tag{3.26}$$

To capture the effects of income, unemployment, working population, household consumption, and house price, the study proposes the following household debt model:

$$HD_{i,t} = \alpha_0 + \alpha_1 HD_{i,t-1} + \alpha_2 GDPPC_{it} + \alpha_3 UN_{it} + \alpha_4 WPOP_{it} + \alpha_5 INF_{it} + \alpha_6 LIR_{it} + \alpha_7 CON_{it} + \alpha_8 HPI_{it} + \varepsilon_{it} \tag{3.27}$$

where $HD_{i,t}$ is the household debt of country *i* at time *t*, and HD_{t-1} is the one-period lagged household debt percentage to GDP with α_1 to capture the persistent[9] within the stock of debt following the literature Hartropp (1992). α_0 is a constant term, α_j (j=2…8) are coefficients for the explanatory variables, namely, GDP per capita growth (GDPPC),

[9] Household debt is accumulated over time and hence potentially persistent. Hence, HD_{t-1} is incorporated in the model.

unemployment (UN), working population (WPOP), interest rate (LIR), inflation rate (INF), household consumption (CON), and housing price index (HPI).

Several studies have identified the pivotal role of financial development in influencing household debt. Financial development acts as an intermediary in directing households' funds from surplus units to deficit units (Antzoulatos & Tsoumas, 2010). Levine (1997) asserted the importance of financial services in offering financial support to spur investment activities. The intuition to postulate a relationship between financial development and household debt is simple. Financial innovation underpins the growth of consumer credit when house prices are higher (Crawford & Faruqui, 2012). Financial institutions are tempted to engage in mortgage financing in the financial deregulation and increasingly competitive environment (Gerlach & Peng, 2005). In the model of Justiniano et al. (2016) and Debelle (2004), household debt expansion occurs when lending constraints are relaxed, leading to higher levels of household debt. In fact, contemporary findings have shown the possible association between credit growth and house price amplified by broad money in financially deregulated markets, creating financial excess (Goodhart & Hofmann, 2008; Mian & Sufi, 2018). Important evidence was provided by Dellas and Hess (2005) and Antzoulatos and Tsoumas (2010), confirming the significant effect of financial development on the household balance sheet.

The testable econometric modelling captures the role of the supply side, namely financial development, as follows:

$$HD_{i,t} = \alpha_0 + \alpha_1 HD_{i,t-1} + \alpha_2 GDPPC_{it} + \alpha_3 UN_{it} + \alpha_4 WPOP_{it} + \alpha_5 INF_{it} + \alpha_6 LIR_{it} + \alpha_7 CON_{it} + \alpha_8 HPI_{it} + \alpha_9 FD_{it} + \varepsilon_{it} \quad (3.28)$$

where $HD_{i,t}$ is the household debt of country *i* at time *t*, and HD_{t-1} is the one-period lagged household debt as a percentage of GDP with α_1 to capture the persistent of household debt. α_0 is a constant term, α_j (j=2…9) are coefficients for the explanatory variables, namely GDP per capita income (GDPPC), unemployment (UN), working population (WPOP), interest rate (LIR), inflation rate (INF), household consumption (CON), housing price index (HPI), and financial development (FD). All variables are transformed into logarithmic form.

The study extended the analysis by conducting a robustness test using alternative proxies such as liquid liability and private credit, which were retrieved from the Financial Structure Development database. In addition, the study analysed the regression model with the banking crises dummy ("CRISIS"), measured using systemic banking crises. As suggested by Laeven and Valencia (2013), it takes the value of zero if there is no banking crisis in the subsequent five years and the values of 1/5, 2/5, and so forth if a banking crisis occurs in one year, two years, and so forth in the subsequent five years.

3.3.4.1 Data and List of Variables in the Household Debt Model

The panel dataset used in this study possesses two main dimensions, namely a cross-sectional dimension consisting of 41 countries and a time series dimension spanning from 1980 to 2018. Data availability on house price and interest rate are incomplete for Argentina and Saudi Arabia respectively, hence excluded from data analysis. For estimation purposes, data were collected from several sources, such as BIS, DataStream, and the World Bank

database. The selected variables were based on the theoretical framework of the research objectives and the novelty of this study.

3.3.4.2 Household debt

The dependent variable (i.e., household debt) as a percentage of GDP is based on the studies of Jappelli et al. (2013), Rubaszek and Serwa (2014), and Mian et al. (2017). Data for the 41 countries spanning the period of 1980–2018 were obtained from the BIS database.

3.3.4.3 Income

Real GDP is the best proxy for income (Ma'in et al., 2016; Meng et al., 2013; Meniago et al., 2013). Meanwhile, some studies used real GDP per capita as a proxy for the income of households as it is the most appropriate to represent household income (Jappelli et al., 2013). Rubaszek and Serwa (2014) used gross national disposable income per capita to represent income in investigating the OECD countries and found a strong positive relationship between the variable and household debt. Meniago et al. (2013) discovered a positive and significant link between real GDP and household debt. The finding is supported by the evidence provided by Ma'in et al. (2016). Data for real GDP per capita were obtained from the World Bank database. Thus, the current household debt model is expanded to include GDP per capita as a proxy for income and is expected to be a positive or negative influence on household debt.

3.3.4.4 Unemployment rate

Unemployment rate signals development in the labour market. The effect of unemployment on household debt has been found to be mixed. Rashid et al. (2017) asserted that household debt will significantly decrease if the unemployment rate increases. Further, Meng et al. (2013) and Ma'in et al. (2016) found a negative effect of unemployment on household debt. This finding indicates that households lose their source of income with high unemployment rates. However, Jappelli et al. (2013) found a positive and statistically significant coefficient between unemployment rate and household debt. In this study, data for unemployment proxied using unemployment rate were obtained from the World Bank database following the Rashid et al. (2017), Park and Lee (2019) and Kusairi et al. (2019).

3.3.4.5 Working population

Working population represents young households who have a role in household debt during their working life. Krishnan et al. (2015) argued that young individuals are expected to have higher debts during their working age. Betti et al. (2001) reached a similar conclusion that being indebted is an unavoidable and normal occurrence for young individuals or households. Yilmazer and Devaney (2005) found the likelihood of holding each type of debt and the amount of each type of debt compared to total assets decrease with age. Hence, working population is expected to have a positive relationship with household debt, measured using the percentage of the population aged 15 to 64 over the total population following study by Kusairi et al (2019), based on the data retrieved from the World Bank database.

3.3.4.6 Inflation rate

Inflation rates were obtained from the World Bank database and used in recent many studies Kusairi et al. (2019) and Rashid et al. (2017). Tudela and Young (2005) and Debelle (2004) highlighted that theoretically, high inflation leads to high nominal interest rates, which could lower mortgage demand as the part of household income that has to be spent on housing finance increases. Hence, inflation is expected to have a negative and significant effect on household debt because higher inflation rates reduce the ability of households to borrow. Similar results were reported in several studies such as Philbrick and Gustafsson (2010), which found a negative relationship between inflation and household debt in Australia. In contrast, Meniago et al. (2013) found a positive link between these two variables in their study on South Africa.

3.3.4.7 Lending interest rate

Interest rate plays a crucial role as borrowers significantly monitor the cost of borrowing. Thus, the lending interest rate is the main cause that explains households' debt behaviour and used mainly in the literature as a proxy (Ma'in et al., 2016; Meng et al., 2013; Meniago et al., 2013; Rahman & Masih, 2014). However, Philbrick and Gustafsson (2010) and Rashid et al. (2017) found an insignificant link between household debt and interest rate and justified that house price data might have already captured the effects. The data of lending interest rate were retrieved from IMF, OECD database, and The World Bank statistics.

3.3.4.8 House price

The housing price index is available on the BIS database and has been used to represent the variable of house price in most studies (Ma'in et al., 2016; Meng et al., 2013; Rubaszek & Serwa, 2014; Philbrick & Gustafsson, 2010; Wildauer & Stockhammer, 2018). Meng et al. (2013) showed that house price had a positive effect on household borrowing. Meanwhile, Ma'in et al. (2016) found a positive but statistically insignificant dependency between house price and debt. Wildauer and Stockhammer (2018) provided evidence of a positive link between real property prices and household debt. Rubaszek and Serwa (2014) found that household debt was positively affected by house price in the long term. Interestingly, Turk (2015) reported a bidirectional relationship between the two variables. Hence, the relationship between house price and household debt is expected to be positive.

3.3.4.9 Financial development

The financial development index as the broadest available indicator of financial development was obtained from the Global Financial Development Database of IMF. Financial development is defined as a combination of depth (size and liquidity of markets), access (ability of individuals and companies to access financial services), and efficiency (ability of institutions to provide financial services at low cost and with sustainable revenues, and the level of activity of capital markets). This broad multi-dimensional approach to defining financial development follows the matrix of financial system characteristics developed by Svirydzenka (2016). Interestingly, financial development has been found to facilitate the growth in household debt (Rajan & Zingales, 1996; Kim et al., 2014). In this study, the relationship is expected to have a positive sign.

Table 3.2: List of Variables for RO3

RO3: To investigate the influence of financial development on household debt in the investigated countries.				
Variables	**Abbreviation**	**Description**	**Expected Sign**	**Source**
Dependent Variable				
Household debt	HD	The credit to the household of the country.		BIS
Independent Variables				
1. Income	GDPPC	real GDP per capita	-	The World Bank
2. Working population	WPOP	Population ages 15-64 (% of total population)	+	The World Bank
3. Unemployment	UN	The unemployment rate in the country.	-	The World Bank
4. Inflation rate	INF	The inflation rate of the country	-	The World Bank
5. Lending Interest rate	LIR	Lending rate of the country.	-	IMF/The World Bank/OED
6. Household consumption	CON	Households and NPISHs final consumption expenditure (% of GDP)	+	The World Bank
7. House price	HPI	The house price index of the country.	+	BIS
8. Financial development	FD	The Financial development index.	+	IMF, Svirydzenka (2016)

3.3.5 The Role of Household Debt in a Crisis

The fourth objective is to investigate household debt as a predictor of the probability of systemic banking crises, following Laeven and Valencia (2013). The choice of explanatory variables follows Demirgüç-Kunt and Detragiache (1998, 1999) and modernised by Davis and Karim (2008), which propose that too many financial variables may lead to biased results. Thus, the study excludes domestic credit as a substitute for household debt. Furthermore, the study focuses on household debt and the change in household debt growth at $t-1$ based on the studies of Angeles (2015), Mian and Sufi (2018), and Park et al. (2018), which found that the change in household debt growth hampered future growth. The indicators of systemic banking crises are functions of macroeconomic variables and financial variables. The function form of systematic banking crises is as follows:

$$CRISIS = f(GDPG, TB, DEP, LIR, INF, FISCAL, FC, HD, HDTG_{t-1}) \quad (3.29)$$

where $CRISIS$ is systemic banking crisis, and the macroeconomic variables are growth in GDP (GDPG), trade balance (TB), depreciation (DEP), lending interest rate (LIR), inflation rate (INF), and fiscal balance (FISCAL). Meanwhile, the financial variables consist of financial contagion (FC), household debt (HD), and growth in household debt (HDTG_{t-1}).

From the macroeconomics factors, real GDP growth portrays a good economic condition as opposed to the event of crisis. Demirgüç-Kunt and Detragiache (1998) claimed that the expansion of the real side of the economy has been a major source of systemic banking sector problems. GDP measures economic performance, as high GDP growth signifies economic expansion, low unemployment rate, high household income, and

increased aggregate consumption that stimulates economic growth. In contrast, low GDP growth indicates a higher unemployment rate, negative income, reduced consumer expenditure, and in turn, economic slowdown. Simply put, lower GDP growth is due to adverse economic shocks during a crisis or economic bust. Hence, GDP growth and crisis is expected to have a negative relationship.

Terms of trade are another signal for crises when it is related to a country's background as an open economy that may be exposed to macroeconomic changes (Davis & Karim, 2008). When banking institutions are actively participating in lending and investment during an economic boom, local and foreign investors are attracted to invest substantial funds, and this decision involves terms of trade and also foreign exchange. While there are market risks such as high borrowers' default and higher non-performing loans, financial institutions tend to neglect the reality of systemic risk[10] catastrophe (Borio et al., 2001). Following the event, asymmetric information is established due to banking failure. However, investors are late to realise the systematic risk exposure on the balance sheet (Kaufman & Scott, 2003). Thus, a sudden deterioration in an economy's terms of trade could precipitate a banking crisis and a negative sign is expected in explaining the relationship.

Currency depreciations arise due to adverse movements in the trade openness. During an economic boom, local currency experiences appreciation and then depreciates during a

[10] Occhino (2016, p. 6) stated that the existence of systemic risk may also be signalled by "high default probability and leverage, low expected values of real activity, low equity values, low risk-free rates and high bond spreads". Davis and Karim (2008, p. 91) argued that the banking systemic risk reflects "a correlation of performance between institutions". Liquidity failures of solvent banks that are contagious trigger the crises. In these cases, asymmetric information and associated bank runs precipitate the crises. Illing and Liu (2003, 2006) postulated that the financial system's exposure generally stems from deteriorating macroeconomic conditions and, more precisely, from innovative developments in the real economic and financial sectors, shocks within the financial system, banks' idiosyncratic risks, and contagion among institutions.

bust. Thus, currency depreciation reflected in the exchange market risk plays a pivotal role as one of the early warning signals for banking crises (Demirgüç-Kunt & Detragiache, 1998). In addition, Babecký et al. (2012) confirmed that banking crises cause currency crises and eventually lead to twin crises. Currency depreciation and crisis are expected to have a positive relationship.

Honohan (2000) highlighted that interest rate fluctuation corresponding to financial liberalisation may trigger crises. During financial liberalisation, intense competition in the banking industry might cause interest rate volatility and an accumulation of systemic interest rate risk (Davis & Karim, 2008). Therefore, real interest rate serves as a potential indicator of banking crises and is expected to have a positive sign.

When the cost of borrowing is lower, consumption will increase and inflate commodity prices. High inflation signals policy mismanagement, which causes higher nominal interest rates at the expense of lenders. Higher inflation may also reflect the market risk of asset price booms (Davis & Karim, 2008). The inflated house prices due to financial liberalisation incorporate market risk. Thus, inflation has been suggested as a significant indicator of banking crises in many studies (e.g. Joy et al., 2017; Kaminsky & Reinhart, 1999). Higher inflation is expected to increase the probability of a crisis.

Davis and Karim (2008) found that good fiscal surplus mirrors government efficiency and an ability to bail out the banking system and reduce crisis. In contrast, low fiscal surplus reflects policy mismanagement. Some countries fell into the global credit crunch due to disastrous fiscal deficits (Bonin et al., 1998). Thus, higher fiscal surplus is expected to reduce the probability of a crisis.

For the financial variables, financial contagion is usually measured using bank liquidity or broad money, indicating the strong position of financial development in surviving a crisis. Dawood et al. (2017) suggested that the vulnerability of currency crises could also be depicted using broad money as lower ratios indicate weak ability to defend monetary volatility. However, Laeven and Valencia (2013) categorised the banking crises whereby there are significant policy interventions in the banking sector, such as extensive liquidity support. Diamond and Dybvig (1983) hypothesised that peculiar bank runs are the result of increased liquidity risks. Thus, a systemic banking crisis is predicted by financial contagion, either with a negative or a positive sign.

Another financial variable used by Demirgüç-Kunt and Detragiache (1998) is private credit. Since household debt is credit offered by financial institutions, it can also represent the financial variable. In addition, the previous work by Babecký et al. (2012) shows that household debt helps predict banking crises. This explains the household debt ratio to GDP HD as one important variable in predicting crises. In addition, empirical literature such as Büyükkarabacak and Valev (2010), Jordà et al. (2013), Mian et al. (2017), and Schularick and Taylor (2012) highlighted the role of credit growth $HDTG_{t-1}$ as a powerful predictor of economic misery. Looking at the indicators typically used in EWS models, in the light of the discussion above, Davis and Karim (2008) highlighted that the need to include both private credit ratio to GDP and change in real credit growth indicating credit risk accumulation. So, this study replaced private credit ratio to GDP and change in real credit growth with HD and $HDTG_{t-1}$ as discussed above (first paragraph of this section) to avoid biased results if too many financial variables are included.

Thus, a higher HD and $HDTG_{t-1}$ is expected to increase the probability of banking crises. Hence, the equation in this study considers household debt as a financial indicator of systemic banking crises.

3.3.5.1 Data and List of Variables in the Systemic Banking Crisis Model

A dummy variable is used in this study to denote the impact of a crisis on economic growth. It takes the value of 1 for each observation, otherwise taking a zero value. The dummy variable is denoted by systemic banking crises (CRISIS).

Table 3.3: List of Systemic Banking Crises

Country	Systemic Banking Crises (Laeven & Valencia, 2013)	Country	Systemic Banking Crises (Laeven & Valencia, 2013)
Australia*		Japan*	1997
Austria*	2008	Korea**	1997
Belgium*	2008	Luxembourg*	2008
Brazil**	1990, 1994	Malaysia**	1997
Canada*		Mexico**	1981, 1994
Chile**	1981	Netherlands*	2008
China**	1998	New Zealand*	
Colombia**	1982, 1998	Norway*	1991
Czech Republic**	1996	Poland**	1992
Denmark*	2008	Portugal*	2008
Finland*	1991	Russia**	1998, 2008
France*	2008	Singapore**	
Germany*	2008	South Africa**	
Greece*	2008	Spain*	2008
Hungary**	1991, 2008	Sweden*	1991, 2008
India**	1993	Switzerland*	2008
Indonesia**	1997	Thailand**	1983, 1997
Ireland*	2008	Turkey**	1982, 2000
Israel**		United Kingdom*	2007
Italy*	2008	United States*	1988, 2007

* Advanced economies
** Emerging economies

Data for CRISIS were retrieved from the comprehensive database of Laeven and Valencia (2013), as shown in Table 3.3. They defined a systemic banking crisis based on two conditions, namely (i) significant signs of financial distress in the banking system (as indicated by significant bank runs, losses in the banking system, and/or bank liquidations) and (ii) significant banking policy intervention measures in response to significant losses in the banking system.

The present research is interested to investigate the role of household debt as an early indicator of a systemic banking crisis. Data of household debt were obtained from the BIS database which refers to credit to household and non-financial sectors. As for the fundamental explanatory variables, data were sourced from the annual data of the World Bank database and Financial Structure Database (FSD) (see Table 3.4).

Table 3.4: List of Variables for RO4

RO4: To examine the role of household debt as an indicator in the EWS in predicting the probability of systemic banking crises.				
Variables	**Abbreviation**	**Description**	**Expected Sign**	**Source**
Dependent Variables				
Systemic banking crises dummy	CRISIS	Crisis =1, otherwise zero		Laeven and Valencia (2013)
Independent Variables				
Household debt	HD	Credit to the household percentage of GDP of the country	+	BIS
Household debt growth	HDTG	The total of real credit to the household in US Dollar of the country	+	BIS
GDP growth	GDPG	real GDP per capita	-	The World Bank
Trade balance	TB	Current account balance	-	The World Bank
Nominal depreciation	DEP	Exchange rate based on market risk of the country.	+	The World Bank
Interest rate	RIR	Real interest rate	+	The World Bank/OECD
Inflation	INF	Consumer price index	+	The World Bank
Fiscal surplus/GDP	FISCAL	Fiscal balance of the country	-	The World Bank
Financial contagion	FC	The measure for the bank liquid reserves of total asset of the country.	+	The World Bank/FSD

Note: Measurement for all variables is based on yearly frequency.

3.4 Data Cleaning

3.4.1 Data Transformation into the Logarithmic Form

Box and Cox (1964) argued that the log-linear model is better in statistically interpreting the dependent and non-dependent variables because transforming data into the logarithmic form can adjust the scale of values across all variables. However, transformation into the logarithmic form is not possible for data with negative values. Negative values in the logarithmic form are considered as missing values, which will affect the accuracy in estimating the regression. Thus, Johnson (1949) formulated a method for negative values in transforming data into the logarithmic form in statistics. Johnson (1949) used the inverse hyperbolic sine (IHS) method, which has been described by Burbidge et al. (1988), to transform data from negative to positive values, as follows:

$$x^* = \ln\left((x + \sqrt{(x^2 + 1)}\right) \tag{3.30}$$

where x^* is transformed data from a negative value x. In this study, there are several series of explanatory variables in household debt data that have negative values such as inflation and lending interest rate. Since these negative values are considered as missing values in the logarithmic form, the data series were transformed to positive values first using the formulas.

3.4.2 Removing Outliers

The study detected the influential outliers with respect to *x* and *y* using Cook's Distance Outlier Test (Cook's D). Cook's D D_i of observation *i* (for *i*= 1, …,*n*) is defined

as the sum of all the changes in the regression model when observation *i* is deleted. A formula for Cook's D for identifying influential cases are:

$$D_i = \sum_{j=1}^{n} \frac{(\hat{y}_{j(i)} - \hat{y})^2}{p\hat{\sigma}^2} \tag{3.31}$$

where $\hat{y}_{j(i)}$ is the estimated mean of *y* at observation *j*, based on the reduced data set with observation *i* deleted, *p* is the number of regression coefficients and $\hat{\sigma}^2$ is the mean squared error of the regression model. Equivalently, it can be expressed using the leverage h_i

$$D_i = \frac{e_i^2}{k+1} * \frac{h_i}{1+h_i} \tag{3.32}$$

where h_i is the hat-value for each observation and is the standardised residual. The first fraction measures discrepancy, and the second measures leverage. There is no significance test for *Di* (i.e., the *F* value here measures only distance) but a cut-off rule of thumb is:

$$D_i > \frac{4}{n-k-1} \tag{3.33}$$

When the value of D_i is more than value of $\frac{4}{n-k-1}$, hence the ith data point is removed when the regression is recalculated. In other meaning, the analysis is regressed on the data which the ith data point is removed.

3.5 Estimation Procedure

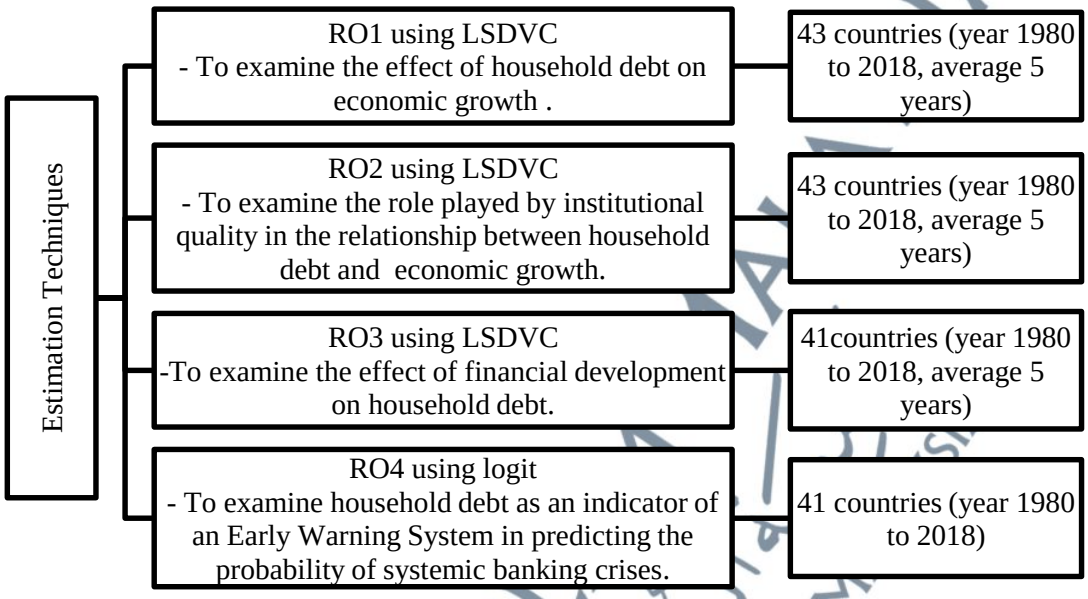

Figure 3.2: Estimation Technique for Each Research Objective

3.5.1 Data average

The panel time series are first examined for stationary results. Overall, the panel unit root test results support the persence of a unit root at the level for some variables (see Appendix 7.6). Hence, this study use 5-year averages of the variables to reduce the volatility and to avoid the drawback of having strong cyclical factors in the data following several papers such as Cecchetti et al. (2011) and Law and Habibullah (2009). This transformation entails that for the 43 countries mostly 5-year periods exist (1980-1984, 1985-1989, 1990-1994, 1995-1999, 2000-2004, 2005-2009, 2010-2014, 2015-2019).

3.5.2 Static panel analysis

Table in Appendix 7.7 and 7.8 report findings from the panel ordinary least square (POLS), random and fixed effect regression method. The static panel analysis shows that

the static panel analysis is not suitable for the model in RO1, RO2 and RO3 since there is presence of serial correlation and hence the dynamic panel is applied respected to this condition. The study also reports the endogeneity results based on the Durbin–Wu–Hausman (DWH) test and the models free from endogeneity problem.

3.5.3 **LSDVC Estimation Method**

This study employed a bias-correlated LSDV; LSDVC (AH), LSDVC (AB), and LSDVC (BB) estimators mainly for empirical analysis in evaluating RO1, RO2, and RO3. The LSDVC was proposed by several scholars including Bruno (2005a,b), Bun and Kiviet (2003), Judson and Owen (1999), and Kiviet (1995). To ascertain robustness, this study applied LSDVC (AH), LSDVC (AB), and LSDVC (BB) estimators, which stand for dynamic bias-corrected estimations developed by Anderson and Hsiao (1982), Arellano and Bond (1991), and Blundell and Bond (1998), respectively. The study employed the system-GMM approach to check for robustness.

In comparing LSDVC with other panel estimators through the Monte Carlo simulation, the researchers found that LSDVC provided higher performance than LSDV, Anderson-Hsiao Instrumental variable, first-difference GMM and system GMM estimators in terms of bias and root mean square errors for a balanced panel (Kiviet, 1995; Judson & Owen, 1999; Bun & Kiviet, 2003). In their proposition, the system GMM estimator, IV and GMM estimators may display nice properties when *N* is large. However, they become biased with high instrumental variables when there is a small number of cross-sectional units in the panel.

Bruno (2005b) demonstrated consistency in the quality of results as opposed to other estimators in his application of the method to an unbalanced panel. It was further confirmed that LSDVC worked best for a small cross-sectional (*N*) panel sample in comparison to the widely used GMM estimator and standard fixed-effect panel estimator (see Flannery & Hankins, 2013). Moreover, the LSDVC estimator offers efficient and robust estimates compared to other dynamic estimators; GMM estimators are used because this approach is characterised by higher root mean square errors (RMSE) and less bias. Also, it is robust as it provides consistent and efficient estimates since it resolves the persistence of the financial variables, which is cited as the major cause of serial correlation (Lemmon et al., 2008). In addition, it evades unobserved heterogeneity, autocorrelation, and fractional dependent variables particularly related to the finance field (Dang et al., 2015).

Some studies employed LSDVC since it reduces the bias arising from small samples and outperforms other dynamic panel analyses such as the GMM (Perić, 2019). For instance, in a recent study, Dahir et al. (2019) analysed the capital and funding liquidity on bank lending by employing LSDVC on 57 banks in BRICS (Brazil, Russia, India, China, and South Africa) countries. In another study, Ibrahim et al. (2019) applied LSDVC to a sample of 37 banks in Malaysia to examine competition and bank stability. Meanwhile, Chaudron (2018) employed LSDVC in analysing bank interest rate and the profitability of 41 Dutch banks. In another recent study, Van Eyden et al. (2019) used LSDVC to estimate 17 OECD countries in examining the effect of oil price volatility on economic growth[11].

Standard dynamic panel data model

[11] These studies reported the number of observations and cross-sections. Nevertheless, no further diagnostic test was done for the LSDVC (i.e. unit root test and etc.). Hence, the results obtained were further analysed for robustness with the different GMM and system GMM dynamic estimators.

$$y_{it} = \gamma y_{i,t-1} + x'_{it}\beta + \eta_i + \epsilon_{it};\ |\gamma| < 1; i = 1, \dots, N \text{and } t = 1, \dots, T, \qquad (3.34)$$

Where y_{it} is the dependent variable; x_{it} is the $((k-1) \times 1)$ vector of strictly exogenous explanatory variables; η_i is an unobserved individual effect; and ϵ_{it} is an unobserved white noise disturbance.

Collecting observations over time and across individuals gives

$$y_{it} = D\eta + W\delta + \epsilon \qquad (3.35)$$

where y and $W = [y_{-1}: X]$ are the $(NT \times 1)$ and $(NT \times k)$ matrices of stacked observations; D the matrix of individual dummies and vector of all unity elements); η is the vector of individual effects; ϵ is the vector of disturbances; and δ is the vector of coefficients.

Bruno (2005) extend Bun and Kiviet (2003) formulae to a more general version which allows missing observations in the interval [0, T] for some individuals. Define a selection indicator r_{it} such that $r_{it} = 1$ if (y_{it}, x_{it}) is observed and $r_{it} = 0$ otherwise. From this define the dynamic selection rule $s(r_{it}, r_{it-1})$ selecting only the observations that are usable for the dynamic model, namely those for which both current values and one-time lagged values are observable:

$$s_{it} \begin{cases} 1 \ if\ (r_{it}, r_{it-1}) = (1,1) \\ 0 \quad otherwise \end{cases} \quad i = 1, \dots, N \text{and } t = 1, \dots, T, \qquad (3.36)$$

Thus, for any i the number of usable observations is given by $T_i = \sum_{t=1}^{T} s_{it}$. The total number of usable observations is given by $n = \sum_{t=1}^{N} T_i$; and $\bar{T} = N/n$ denotes the average group size.

$$s_{it}y_{it} = s_{it}\left(\gamma y_{i,t-1} + x'_{it}\beta + \eta_i + \epsilon_{it}\right),\ i = 1, \dots, N \text{and } t = 1, \dots, T, \quad (3.37)$$

The (possibly) unbalanced dynamic model can then be written as Equation 3.37 and formulated in matrix form. For each i define the $T \times T$ diagonal matrix S_i =diag (s_{it}). Define also the ($NT \times NT$) block-diagonal matrix S_i =diag (s_i):

$$Sy \quad = SD\eta + SW\delta + S\epsilon \quad (3.38)$$

The LSDV estimator is given by $\delta_{LSDV} = (W'A_SW)^{-1}W'A_Sy$, where $A_S = S(I - D(D'SD)^{-1}D')S$ is the symmetric and idempotent ($NT \times NT$) matrix wiping out individual means and selecting usable observations.

Let y_{t0} denote the $(N \times 1)$ - vector of start-up values and assume $\epsilon_{it}|X, S, \eta, y_{t0} \sim i.i.d.N(0, \sigma_\epsilon^2) \forall i, t.$ Then, considering all expectations below as conditional on (X, S, η, y_{t0}), the LSDV bias is given by

$$E(\delta_{LSDV} - \delta) = E[(W'A_SW)^{-1}W'A_S\epsilon)] \quad (3.39)$$

Under our assumption all the properties of normally distributed variables can be used as in Kiviet (1999) to derive the terms of the bias approximation. These generalize the formulae of Bun and Kiviet (2003) by replacing the standard within operator with A_S (A_S s also matters for the order of the approximation terms):

$$c_1(\bar{T}^{-1}) = \sigma_\epsilon^2 tr(\Pi)q_1 \quad (3.40)$$

$$c_2(N^{-1}\bar{T}^{-1}) = -\sigma_\epsilon^2[Q\overline{W'}\Pi A_S\bar{W} + tr(Q\overline{W'}\Pi A_S\bar{W})I_{k+1} + 2\sigma_\epsilon^2 q_{11}tr(\Pi'\Pi\Pi)I_{k+1}]q_1 \quad (3.41)$$

$$c_3(N^{-1}\bar{T}^{-2}) = \sigma_\epsilon^4(\Pi)\{2q_{11}Q\overline{W'}\Pi\Pi'\bar{W}q_1 + [(q_1'\overline{W'}\Pi\Pi'\bar{W}q_1) + q_{11}tr(Q\overline{W'}\Pi\Pi'\bar{W}) + 2tr(\Pi'\Pi\Pi'\Pi)q_{11}^2]q_1\} \quad (3.42)$$

where $Q = E[(W'A_SW)^{-1}= [\overline{W'}A_S\bar{W} + \sigma_\epsilon^2 tr(\Pi'\Pi)e_1e_1']^{-1}; \bar{W} = E(W); e_1 = (1,0,\dots,0)'$ is a $(k \times 1)$ vector; $q_1 = Qe_1;\ q_{11} = e_1'q_1; L_T$ is the $(T \times T)$ matrix with unit first lower subdiagonal and all other elements equal to zero. With an increasing level of accuracy, the following three possible bias approximations emerge

$$B_1 = c_1(\bar{T}^{-1});\ = B_2 = B_1 + c_2(N^{-1}\bar{T}^{-1});\ B_3 = B_2 + c_3(N^{-1}\bar{T}^{-2}) \quad (3.43)$$

So

$$LSDVC_i^j = LSDV - \hat{B}_i^j i = 1,2,3, j = ah, ab, bb \quad (3.44)$$

$LSDVC_i^j$ is implemented by Stata code -xtlsdvc-, for the three levels of approximation accuracy and with three alternative initial estimators: Anderson-Hsiao (option: initial (ah)); Arellano-Bond (option: initial (ab)); Blundell-Bond (option: initial (bb)).

3.5.4 **Logit Model**

The fourth objective (RO4) is aimed at determining the link between crisis and household debt, motivated by evidence from the literature that rising household debt coupled with rising house price may pose a great risk to economic growth. The multivariate logit approach is enthused by Demirgüç-Kunt and Detragiache (1998) to relate the likelihood of occurrence or non-occurrence of a banking crisis to a vector of explanatory variables, depicted in Equations 3.29. The probability that the banking dummy takes a value

of one (crisis occurs) at a point in time is given by the value of the logistic cumulative distribution evaluated for the data and parameters at that point in time. Thus,

$$Prob(CRISIS_{it} = 1) = F(\beta X_{it}) = \frac{e^{\beta' X_{it}}}{1 + e^{\beta' X_{it}}} \tag{3.45}$$

where $CRISIS_{it}$ is the banking crisis dummy for country i at time t, β is the vector of coefficients, and X_{it} is the vector of explanatory variables, namely growth of gross domestic product (GDPG), trade balance (TB), depreciation (DEP), lending interest rate (LIR), and inflation rate (INF). The financial variables include broad money, fiscal balance (FISCAL), financial contagion (FC), household debt (HD), and growth in household debt ($\mathrm{HDTG_{t-1}}$), while $F(\beta X_{it})$ is the cumulative logistic distribution. The parameters are obtained by maximum likelihood estimation where each possible value of $CRISIS_{it}$ contributes to the joint likelihood function so that the log likelihood becomes

$$log_e \mathrm{L} = \sum_{i=1}^{n} \sum_{t=1}^{T} [(CRISIS_{it} log_e F(\beta' X_{it})) + (1 - CRISIS_{it}) log_e (1 - F(\beta' X_{it}))] \tag{3.46}$$

The parameters obtained by maximising this function are not constant marginal effects of X_i on the crisis probability since the underlying relationship is non-linear. Rather, the marginal effect of X_{it} on $CRISIS_{it}$ is given by the probability of crisis times the probability of no crisis times the coefficient β_i. Since the probabilities depend on the values of X_{it}, for a given coefficient, a single explanatory variable can have changing marginal contributions to the crisis probability depending on its starting level. The logistic

cumulative distribution shows that an explanatory variable will make marginally little difference to crisis if the crisis probability is already at the extreme (low or high), but if the crisis probability is around the 0.5 range then a change in the same variable is more likely to tip the balance and trigger a crisis. The sign of the coefficient still indicates the direction of change of the crisis probability. While the sign of the coefficients can easily be interpreted as representing an increasing or decreasing effect on crisis probability, the values are not as intuitive to interpret. Equation 3.46 shows the coefficients of X_{it} are not constant marginal effects of the variable on banking crisis probability since the variable's effect is conditional upon the values of all other explanatory variables at time *t*. Rather, the coefficient β_i represents the effect of X_{it} when all other variables are held at their sample mean values. The logistic model has the benefit of being easily replicable by the policymakers concerned with potential systemic risk in their countries through the detection of the logit linear function.

The next step is to assess the quality of model specification. This study employed three different types of model evaluators: Akaike's information criterion (AIC), in-sample classification accuracy of crisis episodes, and Receiver Operating Characteristic (ROC) statistics. The AIC and non-parametric approach of crisis episodes model evaluators are widely used in comparing the best EWS models following Kaminsky et al. (1999). AIC is joint significance of the regressors by comparing the likelihood of the model with that of a model with only the intercept. A smaller AIC indicates a better model. The prediction accuracy of crisis episodes reports the percentage of crises that are correctly classified, the percentage of non-crises that are correctly classified, and the total percentage of observations that are correctly classified. The model appears to perform fairly well, as the

overall classification accuracy varies between 67% and 84%, while up to 70% of the banking crises are accurately classified. The third model assessor is ROC statistic, following Minoiu et al. (2015) and Comelli (2014), and is used to measure improvements of the predictive power of the benchmark model. The ROC depicts the relationship between true and false positives for a range of probability thresholds. A higher value of the ROC statistic depicts a better model. These three model evaluators are very useful for policy making purposes in choosing the most consistent and the best parsimonious model.

3.6 Chapter Summary

This chapter has provided information on the methodology employed to answer the research objectives outlined in Chapter 1. Discussions include the study's framework, source of data along with the variables and determinants. Besides, this chapter discussed the data cleaning process as well as the estimation techniques used to derive the results.